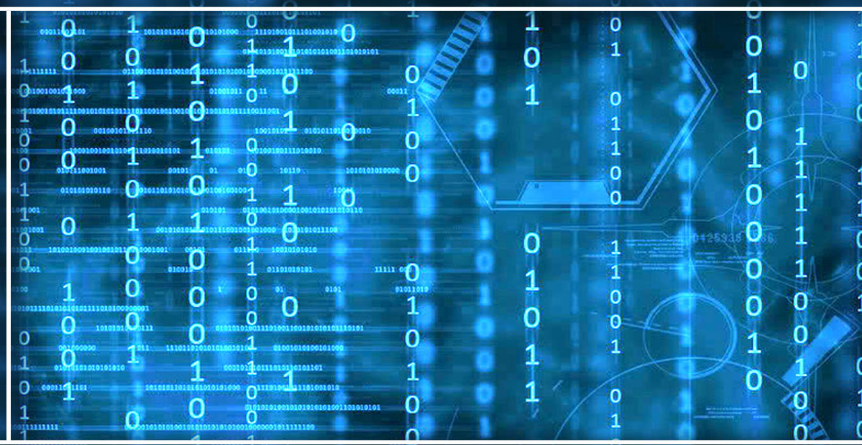
IJACSA

WHERE WISDOM SHARES

International Journal of Advanced Computer Science and Applications

Volume 13 Issue 4

April 2022



ISSN 2156-5570(Online)
ISSN 2158-107X(Print)

SAI

www.ijacsa.thesai.org

Editorial Preface

From the Desk of Managing Editor…

It may be difficult to imagine that almost half a century ago we used computers far less sophisticated than current home desktop computers to put a man on the moon. In that 50 year span, the field of computer science has exploded.

Computer science has opened new avenues for thought and experimentation. What began as a way to simplify the calculation process has given birth to technology once only imagined by the human mind. The ability to communicate and share ideas even though collaborators are half a world away and exploration of not just the stars above but the internal workings of the human genome are some of the ways that this field has moved at an exponential pace.

At the International Journal of Advanced Computer Science and Applications it is our mission to provide an outlet for quality research. We want to promote universal access and opportunities for the international scientific community to share and disseminate scientific and technical information.

We believe in spreading knowledge of computer science and its applications to all classes of audiences. That is why we deliver up-to-date, authoritative coverage and offer open access of all our articles. Our archives have served as a place to provoke philosophical, theoretical, and empirical ideas from some of the finest minds in the field.

We utilize the talents and experience of editor and reviewers working at Universities and Institutions from around the world. We would like to express our gratitude to all authors, whose research results have been published in our journal, as well as our referees for their in-depth evaluations. Our high standards are maintained through a double blind review process.

We hope that this edition of IJACSA inspires and entices you to submit your own contributions in upcoming issues. Thank you for sharing wisdom.

Thank you for Sharing Wisdom!

Kohei Arai
Editor-in-Chief
IJACSA
Volume 13 Issue 4 April 2022
ISSN 2156-5570 (Online)
ISSN 2158-107X (Print)

Editorial Board

CONTENTS

CISE: Community Engagement of CEB Cloud Ecosystem in Box

Dr. Benjamin Garlington
Ph.D. Candidate, University of Arkansas at Little Rock, USA

***Abstract*—The explosion of digital and observational data is having a profound effect on the nature of scientific inquiry, requiring new approaches to manipulating and analyzing large and complex data and increasing the need for collaborative solid research teams to address these challenges. These data, along with the availability of computational resources and recent advances in artificial intelligence, machine learning software tools, and methods, can enable unprecedented science and innovation. Unfortunately, these software tools and techniques are not uniformly accessible to all communities, mainly scientists and engineers at Minority Serving Institutions (MSI). Cloud computing resources are natural channels to enhance these institutions' research productivity. However, utilizing cloud computing resources for research effectively requires a significant investment in time and effort, awkward manipulation of data sets, and deployment of cloud-based applications workflows that support analysis and visualization tools.**

Keywords*—Collaboration; outreach; engagement; narrowcasting; conceptual; methodological; storage-as-a-service; software-as-a-service; data-as-a-service; infrastructure-as-a-service; platform-as-a-service; cloud-ecosystem; minority serving institutions (MSI)

I. Introduction

The current structure of the STEM education research enterprise highly favors individuals who can spend nearly full time on research and so take an idea from inception to publication at a competitive pace. This model is great for individuals at R1 institutions, while new scholars at smaller or less connected institutions cannot leverage communal connection to information supported by new cyberinfrastructure. The last two decades' research data has become a major defining force for America of its economic, social, and national importance. Over the last two decades, scientists within all sectors of STEM have witnessed an explosive rise of data ecosystems as a research tool, and universities are utilizing these infrastructures and their associated applications as an integral part of scientific discovery. The most commonly use tool is cloud computing environments; these ecosystems have emerged as a central tool in the scientific evolution of U.S. STEM rebirth. Additionally, the U.S. has become aware of the urgent need to educate a workforce in all sectors of STEM to take full advantage of these shared resource to ensure that the U.S. remains competitive in STEM research and scientific innovation [1]-[5]. Complicating this task is the diversity of research data that is being placed in such environments. This data is heterogeneous, and the volumes are unprecedented in scale and complexity. It represents the next frontier for innovation, competition, and productivity in research [19].

Therefore, primary intellectual thrust of our research will come from a vision of providing a holistic pathway integrating multiple stakeholders, data sources, and training models to address both immediate and long-term data needs for researchers at MSIs utilizing cloud environments as their primary medium for research. Our project develops a service-oriented resource that assistances diverse communicate, navigate, connect, and inform the research activities of these communities. As a result, the Cloud Ecosystem in a Box (CEB) hub will have a transformational effect for these institutions to by enhancing and propelling research among minority communities and groups currently underserved.

The focus of our particular aspect of CEB will be around engagement that creates a theoretical framework of a network of practice[8], in which individuals can leverage a knowledge portal or hub to facilitates the diffusion of knowledge among a rather loosely coupled and often disconnected community [9]-[11], in certain disciplines i.e. (1) Cloud computing, (2) Nanoscience, (3) Bioinformatic, and (4) Cybersecurity.

Engaging these institutions with multi-disciplines, puts emphasis on data-enabled areas of research that can be achieved within the hub, which is a pivotal area of research in information science that examines how engagement within the context of educational activities can broaden and enhance community engagement with information sources or data.

II. Problem

Across the US, research has been broad and only been characterized through individual contributions in each field as separate entities with no single solution [1]. Tier1 research institutions and Universities greatly compensated, have the advantage in high caliber research, so primary data and secondary data mitigate and escalate to and from these campuses. Meanwhile low-income Minority Serving Institutions (MSI) that only make up 10% [1] of US institutions do not have the knowledge, capacity nor the funds to deliver top-tier data to research on, nor do they have the resources to properly engage training. Around half of MSI students have no inclination of cloud resources or what cloud resources can provide according to the results of the initial survey. Majority of Minority Serving Institutions (MSI) have no forefront knowledge of future endeavors on transitioning to cloud services provided by CEB.

Therefore, a grant has been funded through the National Science Foundation's (NSF) Directorate for Computer and Information Science and Engineering (CISE) to create an on-ramp to allocate resources for specific community domains. A funded EAGER Cloud-Ecosystem in a Box (CEB)(award:1842679) project serving the principal interface for providers in the CISE research community, the education community, and the public cloud community to contribute support from corporate partners by leveraging the Cloud Ecosystem in a Box (CEB). This grant extends cloud services to complement existing resources centered around completely different computational faucets of research. These recommendations were made at the NSF Cloud for Everyone (NSFC4E) workshop (award:1663794) also recommendations made at Enabling Computer and Information Science and Engineering Research and Education in the Cloud (ECISE-REC) workshop(award:818650).

Researchers all over the United States are over-exerted due to the volume of data being computed and stored over the vast cloud movement of processed data and crowd-sourced learning. This data has increased the number of activities to be hosted in particular with, Minority Serving Institutions (MSI), in which, these facilities has resulted in a great number of domain specific publications. In addition to the volume of this unorganized data, new findings are growing at an exponential rate and it has become more time consuming for researchers to draw conclusions within this area of collection of big data due to the lack of addressing engagement.

A. *Dissertation Statement*

Minority Serving Institutions (MSI) should have access to services such as building a cloud-based community to enhance research and infrastructure connected together. By using these engagement techniques, we can increase cloud-driven resources by 20%. Measuring analytic collaborations by building state of the art research and educational cloud ecosystems that leverage advance practices and training within a computational sphere.

B. *Research Area Problem*

How a Hub (CEB) that is an on-ramp onto cloud resources (via AWS, Google, Microsoft, Oracle) could serve Minority Serving Institutions (MSI) community needs and user needs.

With the focus of my research exerting towards Community Engagement by (1)building state of the art research and educational cloud ecosystem (CEB) that leverage advance practices and training within a computational sphere with Cloud providers, and (2)systematically engaging in response time and success of the CEB, plans are put into effect of a more precise nature to establish a mentoring network between Minority Serving Institutions (MSI), through contributing community/user-based needs respectively.

Also maintaining the CEB by being a beacon/liaison for community toolsets to train Minority Serving Institutions (MSI) for cloud-based learning in Cloud *Collaboration* through the life cycles (Phases) of the CEB. This enables superior cloud resources from corporate partners to be at the forefront of separate cloud research tools for a broader research area instead of local centralized research, to accelerate innovation through user-defined computational cloud research while also addressing retention and scalable marketing.

III. Literature Review

This literature review is designed to share similar work corresponding to cloud services provided through an ecosystem like the CEB. It will also compare complexities and pedagogies tied into the various levels of expanding cloud services to areas never incorporated. It will serve as a benchmark for new advancements within cloud services by aiding in filling gaps within the discipline and administering techniques for measuring success through the CEB expounding on the measurement of outreach, engagement, cloud ecosystems**.** I would like to start this literature review with XSEDE (Extreme Science and Engineering Discovery Environment) cloud services, because it is the closest infrastructure related to the CEB that I am most familiar with, also I have extensive expertise within this particular pedological area of cloud resources. I was reluctant enough to be a part of the services provided by XSEDE at the University of Arkansas at Pine Bluff where I practiced as the Campus Champion. My role expanded across a variety of activities such as helping institution researchers, educators and scholars with their computing needs. As the Campus Champion, I was engaged in high-end data intensive research focusing on large scale, high performance computing, distributed high throughput computing environments, visualization and data analysis systems, large-memory systems, data storage, and cloud systems. In addition, I attended/chaperoned the Supercomputing convention.

1) In the paper "XSEDE: Accelerating Scientific Discovery", cloud resources are characterized as a co-existence between multiple cyberinfrastructure communities of scholars, scientists, engineers and researchers to form one ecosystem that addresses the most important and challenging problems [2]. XSEDE's realm of computational technologies and resources divulge in a process that is critical to the success of researchers. These digital services are accessed at a fraction of the cost that they ordinarily would, due to the outsourcing of their complexity to XSEDE. These advanced digital services make up different attributes of technology like supercomputers, storage systems, visualization systems, collections of data, networks and software. The paper emphasizes the motivation to integrate these computational services into a more general form as [2] "judiciously distributed but architecturally and functionally integrated from two arguments": First, the movement of scientific progress within multiple disciplines, which enables a combination of services and resources. Second, that top-tier digital resources are more capable by leveraging multiple smaller institutions, and not being centralized fully at a single institution. XSEDE, as an ecosystem of its on, has an ongoing interface for help-desk support, initiating allocations, and services for sharing other computing digital resources. XSEDE implements tools, policies, and methods that exist with top-tier facilities, campuses, and government institutions as the beginning process for a national e-science infrastructure ecosystem. [3].

2) The paper "open-source cloud management platforms and hypervisor technologies: A Review and Comparison", suggest further implantation of cloud-based solutions and measuring performance in an infrastructural environment [4]. It introduces cloud infrastructure and hypervisor technologies as a means to architect effective cloud solutions, also the need for continues emerging of new tools and technologies used to fit the needs of infrastructure ecosystems. Something very important within cloud infrastructure is what Cloud Framework should be implemented according to needs of the Cloud Service Provider. Software-as-a-Service (SaaS), Infrastructure-as-a-Service (IaaS) and Platform-as-a-Service (PaaS). It goes into intense discussion about Cloud Management Platforms used to deploy virtual machines and how a major consideration for the success of any cloud platform involves the activity and support of the development community behind the project. Hypervisors are introduced as a special operation system that hosts multiple Virtual Machines on a physical machine that allows simultaneous execution of entire operation system instances at the same time. The paper shows how when you combine several virtual machines onto a single physical server it provides on-demand computing to users in real-time. Discussion took place regarding identifying the best cloud technology through different studies, but in today's infrastructure needs it is enough to sustain several open-source cloud computing platforms that addresses difficulties in cloud environment [4].

3) In the paper "DH Box: A Virtual Computer Lab in the Cloud", it outlines how instruction in Digital Humanities (DH) became more efficient since the creation of the DH Box [3]. Challenges within the classroom before the DH Box consisted hardware and software resources and installment/configure of DH tools running of multiple devices. Run-ins with IT departments played a major part in the attrition of facilitating the classroom and he states that on most campuses, computers were on complete lockdown. The author wanted to develop a portable, reproducible, and approachable way to get people interested in DH, up and running with DH tools with the creation of the DH Box starting out on a Raspberry Pi then graduating to the cloud. From local installs on user hardware to a web application that would be accessible remotely with the infrastructure that consists of a sign-in, front-in website, a back-end that listens for request, takes data and initiates the launch of a script, then the launch script itself, and a simple user menu. The DH Box had cut down on class time devoted to preparing technology, shifting the focus away from maintenance and configuration and toward learning and exploration where students were able to access DH tools from their own devices of their choosing [3].

4) "STEM Outreach: A Literature Review and Definition" is a paper that introduces a pragmatic definition of STEM outreach using a heuristic contextual framework [5]. Without expounding on the ontological results used to complete the definition, this paper concludes the definition as:

"The act of delivering STEM content outside of the traditional student/teacher relationship to STEM stakeholders (students, parents, teachers…) in order to support and increase the understanding, awareness and interest in STEM disciplines."

Outreach delivery methods within this paper include Lecture, Active learning/Inquiry Based learning, Problem Based learning, Workshops and Training Events, and Camps and Events. Examining the College (Undergraduate through Doctoral) target area we see how college students can be both the target audience for STEM outreach programs and also to outreach efforts for younger age groups. Through an extended conference/camp where groups of target criteria can emerge educational support, we see an increasing number of providers and more program support. College students get enhanced networking and more exposure to a particular field of study.

5) In the journal "CARES: Mentoring through University Outreach", characterizes a study which utilizes surveys, campus visits, interviews, and observations for a minority serving middle school [6]. These attributes are at the core of outreach, and within this study proves results to show reluctant change and improvement. The participants were students of the forementioned middle school was paired with a mentor selected and matched by race/ethnicity. The findings in this article articulate three major themes. First, the Cares act increased student's aspirations to attend college by providing access to information that expanded the realms of knowledge and exposed new thoughts and the idea that learning is fun. There are 4 categories that the article demonstrates within the first theme that constitute them for increasing the willingness for higher education: Exposure to new options and ideas, increased sense of purpose, feelings of acceptance and connection with mentors, and recognition of encouragement from mentors. Second, mentors experienced personal growth and benefits from their involvement in the Cares act. This second theme has influenced testimonies from the mentors themselves. All acknowledged the benefits, ranging from rejuvenation and increased motivation to growth in their skills and a sense of emotional satisfaction from seeing their influence on students' futures. Thirdly, the outreach from mentors to the students had decreased the negative socialization that characterized the educational experiences. School policy dictated that education was punitive and boring in which it discouraged them form considering higher education as an option. This outreach shows researchable mitigations in the inequities between the differences of white middleclass students and those of poor African American and Latino students. The policies and procedures within this middle school was evaluated because of the CARES act and staff and faculty came to see how the discipline policy discouraged students.

6) In the journal "Outreach, Engagement, and the Changing Culture of the University" the author narrates that in order to change the culture of an organization there has to be a noticeable crisis [7]. A recognized need for higher education to keep pace of an ever-developing societal change takes a forward standpoint pertaining to the culture of education. Outreach and Engagement are characterized by a reaching out from the university to the people and organizations a university serves, also the exchange between the university and its constituents. The author states that engagement is both outreach and in reach into the university by way of extending scholarly creativity to enhance cultural changes inside and out.

7) In the journal "*Outreach initiatives in academic libraries, 2009-2011*", the purpose of outreach with academic libraries is reprimanded with budget cuts declining, also outreach is still on the rise [8]. This particular study represents comparing outreach initiatives through librarians to a particular project conducted by the author. The approach taken was sending a survey to academic librarians that influenced higher learning with outreach, marketing, and promotion. There were successful outreach initiatives in all categories and only a small response rate reflects the target audience. The findings of the survey conveyed a wide range of initiatives involving outreach that target funding. The authors project exceeded the cost of all other initiatives. Certain limitations identified survey flaws and the implications was just a small reach to the targeted audience. A modified survey to the academic library system involved in outreach and engagement is therefore included for future research. The practical implications hardly coexist with budget restraints, because many higher education programs across the nation have inserted outreach into public services positions. The value this type of outreach brings to the table produces limited research about successful initiatives within the spectrum of the timeline. Outreach from academic librarians will be inspired in the collected project outreach takes to create projects with limited funding.

8) In the paper "Pre-university Outreach: Encouraging Students to Consider Engineering Careers" outreach is discussed by activities being affected in educating students about the challenges and rewards of engineering careers [9],. It discusses the impact of certain programs motivating students to consider engineering educational pathways and career pathways. Originally the program was for female students partaking in a summer day camp. It now includes hands-on activities directly to the classroom presented as a co-ed program. There were multiple workshops ranging multiple years that combined males and females, then ranged with the classifications of each individual. The surveys portray a ripple effect with the percentages of K-12 knowledge awareness of engineering. It represents a convergence of STEM intelligence over the years that outreach becomes more and more successful. Proof of the before and after evaluations of student's knowledge of engineering. The questionnaires and evaluations give direct awareness for essential tools about the engineering profession at the pre-university level. It is still in its early stages to tell the increases in expansion of engineering enrollment, but the numbers don't lie about having a positive impact on women in the field.

Whitney Bouk general manager of enterprise at Box said quote: "Future success for students and educators will be dependent on how well we integrate technology into our modern learning environments, Box helps schools to teach, conduct research and run their operations more creatively and effectively, while also giving students new ways to connect and learn. With our growing network of education partners, we are enabling students and teachers to take ownership of education [10]".

Matthew Self Box VP for Platform Engineering said quote: "We're transitioning from a world where the enterprise, and all of your enterprise content, are within the confines of your network and with applications you control, to this world where you have the cloud, all these different applications and all these different devices, with their content scattered all over the cloud. What Box is introducing is a solution… that will bring all of that content to a single place. By using Box OneCloud, we can leverage all of the applications that businesses want to use to run their business and store the data in a single place [11]".

Liz Herbert, an analyst at Forrester Research said quote: "An interconnected network of add-on applications vendors, consultants and partners centered on an anchor cloud provider — and the benefits of a cloud computing ecosystem are outlined. It's a way for organizations to balance agility and speed with control [12].

IV. Methodology

A. Initial Survey

1) Stage/Plans of adoption for research topics: The cloud computing technology fields provided in this question include Private/Internal Cloud, Storage-as-a-Services Only, Infrastructure-as-a-Services, Platform-as-a-Services, and Software-as-a-Service. The results show a top-down approach from the majority results to less inclined results based on answers given, also the same given the sub results of each category. The percentage equated for each category is calculated as the mean, and the sub results are the actual percentages articulated in the summary.

a) Have no plans to use any of the Cloud Computing Services- 41.33%: This first category ranks top priority among Minority Serving Institutions (MSI) survey results schematic. 41.33% Say they have no plans to use any cloud services. This tells me that research could be non-existent in almost half of the Minority Serving Institutions (MSI) participating in this survey. Having no plans to utilize cloud resources is a sure-fire reason why outreach should be manifested as a key component to the growth of desired infrastructure to move the cloud services via existing cloud resources to new communities.

- 50% Have no plans to use Infrastructure-as-a-Service.
- 43.33% Have no plans to use Private/Internal Cloud.
- 43.33% Have no plans to use Platform-as-a-Service.
- 40 % Have no plans to use Software-as-a-Service.
- 30% Have no plans to use Storage-as-a-Service.

Minority Serving Institutions (MSI) deemed Storage-as-a-Service more of a priority within this spectrum of down-top results. Which means Storage-as-a-Service is a more important field based on the results.

b) Using in Production- 18.66 %: This next category is the next top priority among Minority Serving Institutions (MSI) survey results schematic. 18.66 % Say that these stages/plans of adoption will be used in Production. Within this stage, the design, simulation, production, test, and maintenance of a product provides the whole lifecycle of the concept called cloud manufacturing. It provides a new paradigm developed from existing models under the support of cloud computing consisting of integrated and inter-connected virtualized resources for intelligent and on-demand use of services to provide solutions during the lifecycle of manufacturing [13].

- 33.33% Storage-as-a-Service will be used in Production.
- 23.33% Private/Internal Cloud ranks will be used in Production.
- 13.33% Infrastructure-as-a-Service will be used in Production.
- 13.33% Platform-as-a-Service will be used in Production.
- 10% Software-as-a-Service will be used in Production.

Again, Minority Serving Institutions (MSI) deemed Storage-as-a-Service more of a priority within this spectrum of results in Production.

c) Using for Development and Testing only- 8.67%: The third top category 8.67 % Say that these stages/plans of adoption are used for Development and Testing Only. It can be very costly and time-consuming to build the infrastructure for testing in a local environment. Scaling up resources when needed without investing in the infrastructure is what is utilized with cloud-based software development [14].

- 13.33% use Storage-as-a-Service.
- 10.00% use Private/Internal Cloud.
- 6.67 % use Infrastructure-as-a-Service.
- 6.67 % use Platform-as-a-Service.
- 6.67 % use Software-as-a-Service.

Once again Minority Serving Institutions (MSI) deemed Storage-as-a-Service more of a priority within this spectrum of results for The Development and Testing only Category.

d) Plan to use but not sure when- 18%: 18% Say that these stages/plans of adoption are Planned to be utilized but not sure when.

- 30% Plan to use Software-as-a-Service but not sure when.
- 20 % Plan to use Platform-as-a-Service but not sure when.
- 16.67% Plan to use Infrastructure-as-a-Service but not sure when.
- 13.33% Plan to use Storage-as-a-Service but not sure when.
- 10% Plan to use Private/Internal Cloud but not sure when.

Minority Serving Institutions (MSI) ability to looked forward and distinguished to use Software-as-a-Service as a priority. Maybe this survey is pushing researchers, students, and the public to recognize where resources are needed, and Software-as-a-Service holds grounds for future endeavors.

e) Plan to use within: 3-6 months- 6.67%: 6.67 % Say that these stages/plans of adoption are being Planned to be used within 3-6 months

- 6.67% Private/Internal Cloud.
- 6.67% Storage-as-a-Service.
- 6.67% Infrastructure-as-a-Service.
- 6.67% Platform-as-a-Service.
- 6.67% Software-as-a-Service.

f) Plan to use within: 3 months- 3.33%: 3.33% Say that these stages/plans of adoption are being Planned to be used within 3 months

- 3.33% Private/Internal Cloud.
- 3.33% Storage-as-a-Service.
- 3.33% Infrastructure-as-a-Service.
- 3.33% Platform-as-a-Service.
- 3.33% Software-as-a-Service.

g) Plan to use within: 6-18 months- 3.33%: 3.33% Say that these stages/plans of adoption are being Planned to be used within 6-18 months

- 3.33% Private/Internal Cloud.
- 0% Storage-as-a-Service.
- 3.33% Infrastructure-as-a-Service.
- 6.67% Platform-as-a-Service.
- 3.33% Software-as-a-Service.

2) Applications planned to deploy over a 12-month period: This question makes a relevant claim for CEB Cloud Services. Getting to know what Minority Serving Institutions

(MSI) plan to do as a whole can help mitigate activities involved in computing by transferring from a local machine location to cloud services online. We need to know what Minority Serving Institutions (MSI) have already done, that way we have cloud resources available on-top of already existing local resources.

- 29.87 % Say that Research Activities would be planned to be deployed over a 12-month period.
- 19.48% Say that Collaboration would be planned to be deployed over a 12-month period.
- 16.88% Say that Office productivity (e.g., Google Docks, Zoho) would be planned to be deployed over a 12-month period.
- 11.69% Say that Testing/QA/Staging would be planned to be deployed over a 12-month period.
- 7.79% Say that Web 2.0/Social Networking would be planned to be deployed over a 12-month period.
- 7.79% Say that Other(unspecified) would be planned to be deployed over a 12-month period.
- 3.90% Say that Monitoring/Application Performance Management would be planned to be deployed over a 12-month period.
- 2.60% Say that Security/Compliance would be planned to be deployed over a 12-month period.

3) The Initial adoption of cloud computing within MSI was done by: It is important for students or researchers to know the initial adoption of cloud computing in their institution. This is a question to see how literate the research community is with these services among their institutions. We find out how to distinguish as a whole among Minority Serving Institutions (MSI) where the discrepancies are, when it comes to cloud computing Literacy and direction.

- 29.17% Are not sure how the initial adoption of cloud computing was done on their campus.
- 25.00% Say Faculty for file sharing (i.e., Dropbox, OneDrive, Box.com).
- 16.67% Say A CIO/Senior IT Exec decision.
- 16.67% Say A departmental project.
- 12.50% Say A development team.
- 0% Say Cloud Infrastructure is not used at my institution.
- 0% Say other.

4) Current State of Adoption (12-month plans): This is a follow up question from the previous survey question The Initial adoption of cloud computing within Minority Serving Institutions (MSI). By understanding the initial adoption of cloud computing on your campus, you can give an interpretation of the current state of adoption or a 12-month plan. Here we are interested in a 12 month plan due to the existence of cloud computing resources available via contingent on the results of this survey.

- 52.94% Says the adoption of cloud research remains the same.
- 29.41% Says the adoption of cloud research not sure.
- 11.76% Says the adoption of cloud research includes additional applications within the same department.
- 5.88% Says the adoption of cloud research includes additional applications across departments.
- 0% Says the adoption of cloud research Organizational-wide mandate to adopt cloud computing.
- 0% Says the adoption of cloud research other.

5) How many employees work at each institution?: This question is important because it gives focus on underrepresented Minority Serving Institutions (MSI). The idea of minority-serving institutions Minority Serving Institutions (MSI) having adequate high caliber research opportunities, or state-of-the-art facilities to collaborate research have been researched for years but minimally implemented. The caliber of research for Minority Serving Institutions (MSI) should be fundamentally congruent with resources and funds equal to or of greater depth with tier 1 institutions. The higher the funds and resources allocated; the higher caliber of research accomplished.

- 41.18% Of Minority Serving Institutions (MSI) employee 500-1000.
- 35.29% Of Minority Serving Institutions (MSI) employee 100-500.
- 11.76% Of Minority Serving Institutions (MSI) employee 5001-10000.
- 11.76% Of Minority Serving Institutions (MSI) employee more than 10000.
- 0% Of Minority Serving Institutions (MSI) employee 1001-5000.

6) Mission critical applications for research: This question gives momentum to cloud computing and is the catalyst that drives cloud resources. There are projects that require sophisticated resources for big data. With this type of adaptation researchers can use this process to integrate research-oriented cloud ecosystems that operationalize funded research projects concepts within Minority Serving Institutions (MSI).

- 47.37% Say they do not consider cloud computing mission critical for their research.
- 28.95% Say they do consider cloud computing mission critical for their research.

- 23.68% Say they are unsure to consider cloud computing mission critical for their research.

B. Defining User Communities and User Desires

1) Define User Communities

a) Cloud Computing: Cloud Computing itself encompasses services rendered through the CEB to all the communities as services over a secured network purely online. On-demand services provided are HPC, storage, networking, analytics, databases, software, and AI delivered for faster innovation of particular research to achieve coherence and economies of scale [13]. Thus, *cloud Computing* drastically enhances the economic scale of research, and resources become flexible because the delivery mechanism encapsulates the services provided by overlapping platforms. The services provided are Infrastructure-As-a-Service (IaaS), Platform-As-a-Service (PaaS), Software-As-a-Service (SaaS), and Storage-As-a-Service (STaaS). These services overlap each other for full transparency of chosen delivery methods based on user domains.

b) Nanoscience: The study of extremely small things, the manipulation of materials on an atomic or molecular level to build microscopic devices. The Nanoscience community recognizes support of simulation programs in the cloud as to mitigate away from locally running software. The research tools and methods needed are aimed at fabrication methods and measurement tools along ways of measurement in regard to the characterization, design and synthesis within the ecosystems in place with goals responsible for commercialization of nanotechnology [15].

c) Bioinformatics: Bioinformatics is an interdisciplinary field entailed of Computer Science, Biology, Information Engineering, and Mathematical Science/Statistics that analyzes the interpretations of various data types. The role of *Bioinformatics* in cloud research extends to computation counterparts of analyzing and interpreting data. Large amounts of CPU time for tasks carried out in decomposed parallel development require users to work intermittently and facilitate continuous integration. This platform requires analysis to come to the data and to which the datasets expand in size.

d) Cybersecurity: Cybersecurity research is studying the defense of cloud systems from online attackers for the preservation of integrity, confidentiality, and availability of information in the cloud. These attacks take many forms i.e., Malware, Virus, Trojans, Spyware, Ransomware, Adware, Botnets, etc. Through *Cybersecurity* roles played in the cloud, there lies architecture built on top of layers spread across computing paradigms that hold different defenses on separate platforms. The technology used within this research entails of tools for protection from three main entity endpoint devices like: local/cloud/hybrid computers, routers, and smart devices. Also, technology used include, DNS filtering, antivirus software, malware protection, next-generation firewalls, and email [16]. As increased digitalization of human activities increase to improve lives by transferring data to the cloud, this increases vulnerability to being attached within the cloud.

2) Define User Desires

a) Storage-As-a-Service (STaaS): Reported in the survey results MSI students who had somewhat of a degree of cloud computing or knowledge base on their institution says *Storage-as-a-Service* is what is needed at the current state. This is a top priority on the existence of the CEB resources being allocated through the NSF grant to get the CEB Hub off the ground. Research suggest that storage is a top priority in the cloud for small and medium businesses, universities, and organizations outside of the CEB so there are resources already available to help transition from local computing to cloud computing. They are extremely scalable pay as you go system and easy to manage and maintain.

b) Software-As-a-Service (SaaS): This is the second service that serves the needs of Minority Serving Institutions (MSI) within a 12-month level plan to transition to cloud resources. A way to license hosted software for an extended period of time without purchasing or require a pay-as-you-go per user subscription of actual usage. *Software-As-a-Service* is an infrastructure that is most popular/common service over the internet. It correlates everything covertly that *(Paas)*, (*Iaas)*, and *(STaas)* is made up of infrastructurally which makes it more mature and to some extent more standardized or tailored software providers [17].

c) Platform-As-a-Service (PaaS): Survey conveys Minority Serving Institutions (MSI) plan to use *Platform-As-a-Service* within 6 to 18 months as predominant adoption plan for cloud research. It allows complementary software resources as a development and deployment environment to capitalize on this infrastructure as a set of tools to build on-top of CEB resources. Within the time frame *Software-As-a-Service* will pick up off the ground, and top-tier research can be authenticated with training and collaboration.

C. Understanding Firebase as a Backend Service

1) Firebase Authentication Services: Firebase backend services connected to the CEB authenticate users using username and password, phone numbers and federated identity, linking users accounts to social media providers into a single account on Firebase. Identification processes take an exhausting amount of time with traditional backend development, and Firebase takes that process out so you can focus on your users and not the sign-in infrastructure to support them. By knowing the identity of the user and creating an Authorization state we can customized an experience and keep their data secure, and we give them a choice of presenting login to the user. Either we can build our own interface or take advantage of Google opensource user interface (UI).

a) Firebase Authentication SDK: The identity of our users is crucial for the data to be securely saved in the cloud for a personalized experience across all devices. In order to accomplish this, Firebase Authentication tightly augments with other Firebase services leveraging standards of years of trial and error in the making of google services so it can be easily concatenated with a custom backend.

- Email and password-based Authentication – Create and Manage users through email and password sign-in. Password reset emails.
- Federated identity provider integration - (Google/Facebook/Twitter/GitHub/Apple/Microsoft/Yahoo).
- Phone number Authentication – SMS messages to phones.
- Custom auth system integration – Existing sign-in system to gain access to Firebase Realtime Database/Services.
- Anonymous Authentication - Create temporary anonymous account. If the user later chooses to sign up, you can upgrade to a regular account so the user can continue where they left off.

2) Firebase database system: Firebase's Cloud-Based Database query-based retrieve, sort, and filter data system supports real-time data syncing with no servers to deploy or maintain, it has real-time updates, and pay-as-you-go infrastructure. This is a NoSQL scalable infrastructure to store and sync data for server/client development. Firebase database systems carry mobile-first, real-time SDKs (software development kits) also in support of on-site storage for programs offline. Firebase offers two Database systems: (1) *Realtime Database* and (1) *Cloud Firestore*.

a) Realtime Database: Data is stored as JSON and managed as real-time instances that automatically receive updates within rich applications allowing secure access to the database. It provides a flexible language called the Firebase Realtime Database Security Rules which is an expression-based language used to define how your data access I/O, also defines how it can be structured. When paired with Firebase Authentication you can set up who has direct access to the communities and what services are provided through each community. This simple database system is set up for quick execution unlike an original legacy relational database in that it is set up for the whole user experience serving users of communities building real-time experiences without compromising on responsiveness.

b) Cloud Firestore: Like Firebases Realtime Database which allow for low-latency solutions for user devices requiring synced states across multiple mobile app platforms in real-time. Cloud Firestore builds on top Realtime Database software to create new, and more scaled data model offering rich, faster, and scalable queries that are further implemented based on the data. Following NoSQL pattern of database design, you store the data that support many different types, from simple strings and numbers to complex objects, in fields that are then stored in collections called containers that you can organize the build queries from. Querying in *Cloud Firestore* is expressive, flexible, and efficient. It creates shallow queries from documents without retrieving the entire collection, or subcollections. You are able to sort, limit, and filter into your queries to paginate your analysis in real time, also you are able to protect access to sufficient credentials through Firebase Authentication.

3) Firebase storage: Storage, sync, upload and backup time capsule, and key configurations for Macs, Windows, etc. and different mobile devices, which makes them immediately available, so you can have easier ways to collaborate. Firebase cloud services allow to share important files without the hassle of attachments, but instead gives a link option to download in order to preview, you can set passwords and expiration dates for added security. Having access to be able to preview files before downloading is a perk with the iCloud system for the exception, that, in order to edit them, you will have to download the attachments then upload back to Firebase. Also, the cloud services have the ability to save the latest version of any paper, so you can have the latest document available every time while the older versions are also saved for reference. It gives flexibility to share folders and files with colleagues or the ability to set permissions for security. Basically, having private and public sharing at the same time. As with other familiar services, Firebase can create new documents, but has the ability to assign tasks which stops the need for data task management teams.

4) Firebase machine learning: Machine Learning is a small section of AI which incoherently focuses on training algorithms for data analysis that automates analytical model building. A process where tasks can be created without human interference specifying every step of an algorithm. This is where computers learn on their own from represented data where there are no fully satisfactory algorithms to accomplish a task. Firebase recognizes the need for cloud-based processing, the real-time capabilities of standalone devices, or the flexibility of ML resources geared toward cloud-based or mobile-optimized devices. Cloud-based devices include services like text recognition, landmark recognition, and image labeling application programming interfaces, perform inference in the cloud so they require more computational power for HPC, also require more memory so they can perform with greater accuracy and precision. Mobile devices include services for custom model deployment by uploading to the firebase server creating on-demand capabilities, and AutoML Vision Edge that is an app to help create custom image classification models with user-friendly web interfaces. These services are suitable for real-time and low-latency applications provided by optimized to run on mobile devices that don't require a network connection and runs very quickly. Video processing is best suited for processing frames of video in real-time. Firebase makes it simple with a service called the mobile SDK package to provide insights about the data provided.

a) Firebase Predictions: Firebase Predictions leverage ML analytics data, creating behavior influenced clusters of users so you can make research decisions based on predicted behavior rather than historic behavior. Firebase Predictions is consolidated with other ML resources mentioned above for cloud-based or mobile-device customized user experiences so you can adjust the frequency of ads. You can increase retention by re-engaging a user who has predicted to opt-out of Firebase by sending a one-time campaign or automate sending messages for certain groups predicted to not return. Firebase also gives you the option to create your own predictions by creating a cluster based on prediction of any event defined in the analytics data in the near future. All this data is easily transferred to Googles *BigQuery* for further analysis for pushed into third party services.

b) Natural Language Processing (NLP) / Natural Language Understanding (NLU): Natural Language Processing reluctantly is concerned with finding the interactions between human language computers by analyzing large amounts of language data and processing the contents for the contextual nuances then extracting information and insights from the data then organizing the documents by categorizing them under the pretenses of linguistics. Text recognition applies to automating data entry like credit cards, business cards, and receipts. Also extracting text from pictures for translating documents and increasing accessibility. It is optimized for both photos and documents, recognizes text in hundreds of different languages. Firebase Language Identification is used to determine the language of user provided text, which naturally does not come with any uniform language information of a string of text. The capabilities of text translation are limited for casual and simple translations using dynamic storage for downloading and on-device requirements by managing language packets. Firebases Smart Reply API automatically replies to messages by generating multiple reply suggestions based on the context of strings of sentences through a conversation. This API is intended for just normal conversations. It does this by comparing the messages against a list of sensitive topics, then the model provides up to three suggested responses dependent on a sufficient level of confidence.

V. Preliminary Results

We started out addressing engagement of the CEB by (1) Defining User Communities, and (2) Define user desires based on survey result outcomes. The communities we are building are created from pain points within Minority Serving Institutions (MIS). We will continue address engagement of MSI's by measuring analytics of the CEB through response time through the retention/life cycle.

The results of the survey question Stage/Plans of adoption conclusively show that the overwhelming majority of Minority Serving Institutions (MSI) participating in this initial survey rule that cloud resources are novel. Meaning there is an unfamiliarity of resources beyond each institutions' limits on education. Based on this initial survey results, Storage-as-a-Service is a top priority when utilizing these cloud computing resources for the majority of Minority Serving Institutions (MSI) independently for the immediate moment. Next, Software-as-a-Service and Platform-as-a-Service is considered a priority for the future growth of the majority of Minority Serving Institutions (MSI) independently.

The adoption of infrastructural cloud services Cloud Ecosystem in a Box (CEB) across multiple campuses is integrated through a wide network that not only recognizes the need of increased viable research activities, but dedicated research in the classroom as well. Research Activities is ranked the most important at 29.87% to be utilized within the 12-month time frame. Collaboration, which according to the Survey Outcomes, holds 19.48% will also be an immediate undertaking within Minority Serving Institutions (MSI).

For the initial adoption of cloud computing on an MSI campus, here we intercept the knowledge gap and proceed to fill in the necessary timestamp that accrues when Cloud Computing education is not congruent with Minority Serving Institutions (MSI). There is a 29.17% majority of students that are not sure how the initial adoption of cloud computing was done on their campus. Next, we see that 25% of Minority Serving Institutions (MSI) say the initial adoption of cloud computing was done on their campus by Faculty for file sharing, 16.67 % say a CIO/Senior IT Executive decision, 16.67% also say a department project, and 12.50% say a Development Team. Here we see a portion of Minority Serving Institutions (MSI) collaboration on each individual campus. With the numbers being so low in comparison, we see the discrepancy with not only a shortcoming in cloud resources but an existence of ignorance (lacking knowledge or awareness) on each individual campus as well.

The results are conclusive with The Initial adoption of cloud computing within Minority Serving Institutions (MSI) when we see there is a knowledge gap of more than a 4^{th} of students that are not sure how the initial adoption of cloud computing was done on their campus. As a result, we see 52.94% of Minority Serving Institutions (MSI) say the adoption of cloud research remains the same, and 29.41% says they are not sure. As stated previously in The Initial adoption of cloud computing within Minority Serving Institutions (MSI) conclusions, outreach should be manifested as a key component to the growth of desired infrastructure to move the cloud services via existing cloud resources to new communities.

The majority of our small MSI' 41.18% make up 500-1000 employees. The other majority of 35.29% make up 100-500 employees. Due to this underrepresented institution research sector which consist of mainly teaching institutions, both small and large [1], outreach should be attainable to sustain and utilize these resources by establishing a mentoring network to embody a relationship between mentor and student bodies with the boundaries set within each sector.

47.37% Say they do not consider cloud computing mission critical for their research. In response to nearly half of Minority Serving Institutions (MSI) response that they do not consider cloud computing mission critical for their research, again there is a trend and a consistency within these survey results that points to an education relapse of cloud computing. Previous research regarding this particular domain not only fails to address the logistics of how to properly incorporate an ecosystem, but to share, analyze, and draw insights from mined data.

1) Storage-as-a-Service: Transferring data from a local machine to the cloud can be reluctantly smooth, given the proper guidance and planning. Training as a top priority for outreach will be mandated to take steps to reduce any risks involved. The first thing that needs to happen before any research group starts to transfer, is to have a viable back up stored on a local machine. Reason is you don't know of any (1) security risks, (2) network connection interruptions, or (3) improper file organization. During transferring of data your files are pretty much in limbo or left in the hacker's world. Interruptions related to the network that has the network connection or the strength of the network can play a part in the loss of data. After the successful transfer of data from local to the cloud, you may find your data all in disarray. There are ways to find and organize your data, but keep in mind that keeping a backup will enable you to locate the right ones immediately. Next, research needs to be taken place on who will be a suitable cloud service provider.

CEB Cloud storage allows for offline availability and synchronizing/sharing of files on all API's providing universal access to any and all types of data stored within the CEB Hub. It provides a service layer on top of the storage services already allocated by storing all the data and metadata from the users and provides access from different faucets (i.e., FUSE access, WebDAV, XROOTD, Synchronization client, Web access, Mobile devices). How this is done is by an internal database that keeps track of the migration status for all the users of the system through home directory services. The goal is to consolidate storage solutions (i.e., Batch System, Windows Terminal Servers, Personal computers, Small/Private Experiments) into one service that can be accessed from any API, "incognito" the CEB Hub [18].

2) Software-As-a-Service: How Minority Serving Institutions (MSI) migrate Software-As-a-Service within this 12-month period depends on research capitalization. Collaboration supports the CEB design to support many different users safely and securely with unlimited scalability to support changing research. Coordinated efforts within communities makes the foundation become the anchor for a rich partner ecosystem.

What CEB offers with *Software-As-a-Service* within this 12-month period are resources on many levels of platforms to designated API's. The CEB resources for the communities maintain software that is delivered remotely by providers. The CEB offers ecosystem integrators that use their reputation to convince critical as-a-Service Minority Serving Institutions (MSI) solutions especially in response to security concerns.

3) Platform-As-a-Service: The move to Platform-As-a-Service among Minority Serving Institutions (MSI) will increasingly evolve Big Data correspondence with scalable software delivered in the cloud with pay as you go investment. PaaS is a cloud-based environment for coding and deploying applications that make coding for developers and IT professionals much more efficient. The type of Platform-As-a-Service the CEB offers is called an open-cloud PaaS where it's not tied to a product or operating environment. It allows Minority Serving Institutions (MSI) to use a completely different platform therefore providing flexibility [19].

A. Cloud Ecosystem Design Infrastructure

1) Ecosystem life cycle: The best way to tackle our approach to gather access to cloud services is from a system engineering prospective, less from a developer perspective because the CEB Hub or onramp being built is a system with multiple interacting components performing functions that cannot be achieved by one component alone [20]. Therefore, we move forward with steps toward understanding and improving this multi-component faucet by designing and managing this complex system over its life cycles [21].

2) Bird's eye view: This lifecycle map is a user experience birds-eye view of an ecosystem approach to services rendered through the CEB. This diagram is taken from a source that expounds upon user experience/needs and world-class architecture in system design. This is the first step we must take to fully recognize what users of the CEB need and how to fully engage with them, also recognize future viability of emerging answers to questions and understanding of a cloud Ecosystem. The red circles represent the typical stages a user moves through when engaging with resources tailored to them specifically. The first five stages (Aware, Interest, Research, Consider, Purchase) represent from "prospect" to "user", then continuing until the user becomes aware of new services, capabilities, upgrades, and communities of the system their using [21].

"By looking at clusters, we can begin to see which stages are more important than others on a personal basis. We can look at the time required to move between stages as another element of the user experience, and again we may discover different users (Returning/Expert user or New-User) require different amount of time to transition from one stage to another [21]." Here we discover how the user interacts with specific elements of transition to uncover opportunities such as journey mapping to improving user experience leading to loyalty and returning research.

B. Pedagogy

1) What is behind the CEB: The pedagogy behind the CEB is a cloud service enabled by a complex system of unrelated independent components that work together to form a Symbian relationship of hardware and software integrated with consultants, cloud engineers, cloud customers and partners. In the center of this cloud service is a public provider (*Firebase*) that creates easy access for business or higher education to collaborate and share information with colleagues and clients. It offers storage, sync, upload and backup time capsule, and key configurations for Macs, Windows, etc. and different mobile devices, which makes them immediately available, so you can have easier ways to collaborate. Firebase cloud services allow to share important files without the hassle of attachments, but instead gives a link option to download in order to preview, you can set passwords and expiration dates for added security. Having access to be able to preview files before downloading is a perk with the Firebase Cloud system for the exception, that, in order to edit them, you will have to download the attachments then upload back to *Firebase*. Also, the cloud services have the ability to save the latest version of any paper, so you can have the latest document available every time while the older versions are also saved for reference. It gives flexibility to share folders and files with colleagues or the ability to set permissions for security; basically, having private and public sharing at the same time. As with other familiar services, *Firebase* can create new documents, but has the ability to assign tasks which stops the need for data task management teams. The CEB will provide researchers with tools like RStudio, Jupyter notebooks, Linux OS, Windows, and other applications that enable Research Infrastructure within the CEB hub.

2) Research infrastructure: The service model platform used in this particular CEB cloud framework is utilizing the Infrastructure-as-a-Service (IaaS) model. [4] In this type of architecture, the cloud provider (Firebase) manages IT infrastructures such as storage, high performance computing power, data warehousing, security, server and networking resources, database access, big data computing environments, monitoring/manageability, and more by way of our providers (via AWS, Google, Microsoft, Oracle). There is a way to address this through the way of a hypervisor (CEB), which is a (user friendly) virtual machine monitor that host multiple VM’s on a physical machine. This type of monitor allows simultaneous incorporation of different types of operation systems at the same time and provide instances of the diverse machines to users on demand [4]. With the Hub (CEB) portal activated, we see the CEB as an Infrastructure Provision cloud model. This means that the Minority Serving Institutions (MSI) utilize the cloud using the technology as a provisioning tool to supply on-demand virtualized resources. [4] The cloud model used in this project examines different types of characteristics customized for the various needs of our Minority Serving Institutions (MSI).

3) Data mining: Data mining refers to predictive analytics, a field of study recognized in Machine Learning that focuses on exploring data through unsupervised learning to discover previously unknown properties (Entity Resolution) in the data. When dealing with high volume big data projects, the successful implementation of the CEB life cycle become more difficult when dealing with issues related to *Data Governance*. [22]That is why resolving real issues related to the cloud services offered through CEB should identify the highest impact failures that can occur, then using work processes, risk management, and optimization methods to such services. [23].

a) Primary/Secondary data: Using the CEB, resources will be available for designing and populating secondary data resources that researchers can use to integrate already existing data with primary data. With this type of adaptation researchers can use this process to integrate research-oriented cloud ecosystems that operationalize funded research projects concepts within Minority Serving Institutions (MSI). Using the CEB will solve many challenges, mainly the challenge of meta-synthesis where scholars understand how different forms of evidence, variables, and analysis fit together to formulate a plan so the researchers can identify a specific research question, search for, collect, and combine evidence so the research question can be enhanced and to identify which variables from separate data sets may be made comparable to one another.

b) Descriptive metadata: Our descriptive metadata will support numerous STEM researchers’ efforts to identify and obtain data for comparative analysis across states/countries fit for compiling details on appropriate data sources, institutional locations and contact information for acquiring this data. The web based descriptive metadata will have the advantage of advanced cloud hypervisor capabilities with the CEB central hub (BOX) allowing researchers to upload further details based on their own use of data resources.

C. Research Focus

1) Collaboration and outreach: Cloud resources enable Collaboration at its highest level, and the CEB will address the cons of cloud computing through Collaboration, by expanding on how access to real-time engagement tools offer strategies of deployment. The benefits of Cloud Collaboration exist within each community and are exercised within each group working on a particular research project also within the maintenance and management of the CEB. As you scale back from a bird’s-eye view to see the logistics and the institutional boundaries, questions arise on how Outreach is attainable also how retention is manifested. Most existing national and international data are not designed to focus on the relationships between social groups relevant to our work, but through the CEB’s community/user-based discussions will rethink the use of these measures. Therefore, Outreach should be manifested as a key component to the growth of desired infrastructure to move the cloud services via existing cloud resources to new communities.

2) Measuring analytics: Firebase Cloud Messaging offers different ways and ranges of messaging options and capabilities allowing you to send notification messages that are displayed to the user or data messages to determine what happens in your application code used by the client application. Notification messaging is used for user re-engagement and retention. Firebase Cloud Messaging targeting is very versatile, you can distribute messages in three separate ways: to single devices, groups of devices, to user devices subscribed to their different communities by either Message delivery reports, or Notification funnel analysis. You can export your analysis to Googles BigQuery Database infrastructure for further analysis. Firebase Cloud Messaging also makes messaging available through client applications by sending back to your server over its batter-efficient channel set up from settings for relevant engaging messages, targeted messages by community, and flexible custom alert. This helps you engage your applications active users by sending them targeted messages encouraging them to user more of the application features for exploration and discovery.

Mailchimp is a cloud-based marketing automation platform designed for email campaign for a targeted audience. It offers Automated email creation and sending, it is easy to use, it has built in statistics and has a style guide and knowledge base. It is a tool that manages your mailing list for you, and also allows you to create your own custom email templates.

Social Media will be used at a level of networking and outreach to target our particular MSI groups through awareness.

The Backend Functionality configures a serverless framework in response to triggers or HTTPS requests that automatically run in the background. After you write and execute the function it immediately begins to manage as it is triggered.

D. Front-End Web Development

1) Becoming a member: In order for a user to become a member into the CEB portal, you have to get Authentication credentials from the backend set within the Authentication methods noted in the backend text. The credentials are then passed to the Firebase Authentication services, once verified there is a callback response to the client. Once Authenticated by the backend services of Firebase, you can manipulate access to storage, computing power, processing, etc. by modifying the rules set within Realtime Database Services and Storage Security Rules within the Firebase infrastructure.

2) Log In: Users login through the CEB Log in portal either by username and password or by a federated identity, once users are Authenticated through Firebase, the users are placed in their respective communities through the Oauth state, then the community helps them move their research into cloud ecosystem. Firebase backend services connected to the CEB authenticate users using username and password or federated identity, linking users accounts to social media providers into a single account on Firebase.

3) OAUTH: In order for firebase to know which community the logged in user has chosen, there is a script located within the Oauth state that pulls from the database. The Oauth function is also in charge of maintaining state for each user. It fires back at the backend services of firebase so that the user maintains state.

4) Database: At this point the Realtime Database is being used due to the small amount of prototyped user engagement, also its low-latency solution for synced states. This type of database is persisted locally, and able to allocate real-time events offline giving the end user a responsive experience. When online services are regained and the connection is secure, the changes made will merge any conflicts automatically that occurred while the client was offline.

VI. Scope and Timeline (12-Months)

Deliverable 1: Feb 2022

T1: Design methodology to deliver identifying quantitative and qualitative data-driven research and resources to the CEB community.

T2: Measuring data mining using message streams elements, the best tools to engage with them are channels tailored for each group.

o Sentiment Analysis

Deliverable 2: April 2022

T1: Identify content management strategies that best suits this unique community based on user feedback from each group.

T2: Local CEB members and lead correspondence officers will take the lead to develop policy procedures to work with campuses on sustaining CEB's computational research activities and resources for the long term.

o Science Gateways

Deliverable 3: June 2022

T1: Developing a funding strategy for new research hosted in the cloud ecosystem is a top priority for the longevity of the program.

T2: Design branding strategy, user guidelines, and content policies for the ecosystem.

o Code of Conduct

Deliverable 4: Aug 2022

T1: Finalize the CEB Community Groups based on user feedback from survey. Then tailor the users back into their communities.

T2: Test Beta version resources and functionalities

Deliverable 5: Oct 2022

T1: Expanding the first wave of users in the MSI community

T2: Identify outreach opportunities to educate

Deliverable 6: Dec 2022

T1: Promote CEB to the second wave of users in the MSI community.

T2: Consolidate a mentoring network between institutions, corporate partners, and certification organizations to enhance student-mentor relationships throughout the CEB ecosystem

o Certifications from Providers

VII. Implications

A. Where are we now

1) Conceptual: The conceptual contribution conveys focus towards ideas and communication about CEB. Participants from different disciplines and backgrounds within the sphere of different communities: (1) Cloud computing, (2) Nanoscience, (3) Bioinformatic, and (4) Cybersecurity, focused on Minority Serving Institutions (MSI) discussing proposing these research communities to build new research and educational practices to advance leveraging cloud provider ecosystems. These discussions emphasized computational collaboration concepts and introduced new research techniques to using these cloud provider ecosystems. Also, these discussions served to identify critical points of commonality or identify entity resolution within the data in already existing primary data and archived these details in the descriptive metadata to create a sustainable commodity for researchers all around the world, in particular resource challenged institutions Minority Serving Institutions (MSI).

a) Existing Initiative (Existing): Existing Initiative involved public talks, blogs on current topics and publishing in magazine outlets. Meetings convened inviting researchers, policy practitioners and non-profit members, cloud providers. The meetings focused on adapting and refining core ideas within CEB. Participants from different disciplines, and institutions will identify relevant research themes within projects with emphasis on articulating meta-level concepts that are robust and meaningful across national contexts. The first part of this project for year 1 we relied on participants from different disciplines to focus on establishing core ideas within CEB through a virtual meeting broadcasted in real-time via EVO. This meeting emphasized computational concepts on a national level by adapting and refining core ideas within CEB. The main goal of the first part was to showcase the work done and to introduce the community and new research techniques using cloud ecosystems. Within this portion of the project, we developed the conceptual framework for research migration into cloud ecosystems, to include a searchable web-based metadata for quantitative and qualitative data for participating institutions, including primary data from ongoing research specifically designed for researchers to operationalize data-oriented research concepts at these institutions.

b) Education (Current year 2020-2021): Education included classroom innovations, service-learning pedagogies, and graduate webinars to bring together researchers between annual meetings. We invited researchers and postdoctoral students to share their experiences and present individual research activities on both public/private cloud environments. The goal was to both showcase the work done, but also to introduce the community to each other and new research techniques using cloud ecosystems. Provided self-contained learning modules for national certifications, nanodegree programs around cloud computing/cybersecurity resources such as, Cloud Platform-as-a-Service, Cloud Security and Data Protection, and the practical tools used to store, and process access these systems.

2) Methodological: The *methodological* contribution conveys focus on survey result conclusions for user-based needs pertaining to mined research tools that will be mapped as a research Data-as-a-Service (DaaS) Data that already exists with emphasis on data-enabled areas of research by reaching out to corporate partners. This was the pivot point to which educational activities are broadened and enlightened through Data-as-a-Service (DaaS) rather primary or secondary, that have a case-based level on a national scale.

a) Networking (Current year 2020-2021): Networking Video archives of talks, working papers, call for participation in conferences, links to social media, newsletter updates, the descriptive metadata, elaboration of our in-common research design and shared educational resources. All presentation materials are broadcasted in real-time via EVO. Participants are tasked with identifying quantitative and qualitative data-driven research that exist for participating campuses to be leveraged in the new ecosystems. The PIs in collaboration with the core committee members designed and deployed a Hubzero portal (CEB) on Open Educational Resources Commons (OER) to highlight innovative intellectual, methodological elements associated with project activities. This Hubzero portal (CEB) serves as enduring library of unique content, while facilitating collaboration for the MSI audience around deploying research on the cloud. Also, establishing a mentoring network between institutions, corporate partners and certification bodies to foster student-mentor relationships within and across institutional boundaries.

b) Publishing activities (Current year 2020-2021): The second part of this project for year 1 constituted identifying quantitative and qualitative data that exist for participating campuses discussed in part one. The key innovation here is methodological in nature for the shear qualitative case-based data sets and national-level secondary data sets relevant to the CEB. The combined conceptual and methodological work makes the logic keen to realizing the intellectual merit that goes into the CEB, not to mention the intellectual innovations. During this second part we developed an extensible research design for robust mixed-methods research associated with leveraging these ecosystems for (1) bioinformatics, (2) cloud computing, (3) nanoscience, and (4) cybersecurity. Developing the Hubzero portal (CEB) served as a resource for experiential-based learning in cloud ecosystems around cyber infrastructure.

3) Bring it Home: Using the CEB, these Minority Serving Institutions (MIS) will begin to engage computationally and collaboratively within the cloud community it serves. With addressing the methods within this study and an expected scope of future goals, Visualization and Data Analysis System will increasing become more aware within these institutions and faculty will begin to gain value and be better prepared using CEB cloud resources. High Performance and Distributed high Throughput Computing will no longer be accepted into research only for R1 institutions. The CEB will restructure the way research is conducted and will encompass the minority research community, NSF funded ecosystems, and industry partners creating a triple convergence.

Through marketing the CEB, this is the first step to being proactive in outreach for the continuation of services rendered through the CEB. The first plan of action that was already preceded is to target Minority Serving Institutions (MSI) by developing a list of users based on NSF awards on their respective research areas of communities. Then the idea was to narrow this list by geographical location then send out a survey to measure the interest and needs of each community based on gender and demographics. Results from the survey helped us understand the dynamic within each dialect of data-driven research within each respective community of Minority Serving Institutions (MSI). Continuation of marketing throughout the life cycles of the CEB merits different styles of advertising and finding ways to target an indecisive audience. [24].

The CEB will provide at a national level the infrastructure for training program platforms by reinforcing the conceptual framework for research migration into the cloud ecosystems. It will be a direct centralized portal to provide users with: (1) applications to enable them to operate at real-time intervals within parallel computing power, (2) educational material such as links, mailing list, repositories, and YouTube channels pertaining to each Minority Serving Institutions (MSI) research to extend MSI networking through social media engagement, (3) Tools that enable data-driven research: programming interfaces, operating systems over hypervisor VM's, and a liaison/beacon for training.

Cloud services within each of these communities provides stable software platforms Infrastructure-As-a-Service (IaaS), Platform-As-a-Service (PaaS), Software-As-a-Service (SaaS), and Storage-As-a-Service (STaaS) with an associated larger community of providers ((via... AWS, Oracle Academy, Microsoft Azure, Google)) who can provide support and solutions specific to a researchers' domain, using these research specific public cloud providers that make up their own ecosystems [25].

B. *Next Steps*

Building the Communities will be incorporated in the CEB website as tools connected to the cloud providers provided cloud messaging, message bird, trigger emails functions, chat, storage, and machine learning. Upon success of the creation of a conceptual framework to incorporate searchable web-based metadata for computational research, and the developing of a Hubzero portal CEB that will be the center of ongoing research in learning cloud systems communities: (1) Cloud computing, (2) Nanoscience, (3) Bioinformatic, and (4) Cybersecurity. Broadening the scope and intellectual impact of the CEB will be established to build other branches of research. This research includes primary data from current research participating institutions, providing self-contained learning modules like Cloud Platform-as-a-Service, Cloud Security and Data Protection. These include public initiative, education, publishing activities and networking.

Measuring and understanding engagement on any platform is a complex undertaking that requires understanding multiple interconnecting ecosystems. To ensure the project was manageable in the timeframe of our funding, we initially created a survey instrument to select researchers at MSI, that are interested in migrating their research into cloud environments. This will allow us to design discipline level guides to facilitate this movement. In particular to measure engagement, we will capture the following elements, (1) user driven streams from each group i.e. nanoscience, cybersecurity, cloud computing and bioinformatics. (2) The project will be able to store publicly broadcasted messages, which are directed from users, particularly group followers; (3) The project will be able to determine the geographical location of the messages recorded; (4) The project will be able to retrieve user histories to aid in the fabrication of conversation instances; (5) The project will be able to extract all pertinent information from a tweet to aid computational processes; (6) The project will be able 'created_at' stamp which is a more mathematically friendly timestamp of the format MM/DD/YYYY ##:##:##, this will provide context when the message was created; (7) The project will be able to determine the participation of each user involved in a given conversation with a user group; (8) The project will be able to directly query a NOSQL database to keep all data structured and organized by individual user and research groups.

References

[1] Ramirez, G. Fox, A. Kuslikis, K. Barnes, D. Baxter and J. Foertsch, "Engaging Minority-Serving Institutions (MSIs) in Cyberinfrastructure (CI) through CI Days," p. 9.

[2] M. Dahan, J. Towns, T. Cockerill and N. Wilkins, "RG," Sept 2014. [Online]. Available: https://www.researchgate.net/publication/267867365_XSEDE_Accelerating_scientific_discovery/stats.

[3] S. Zweibel, "DH Box: A Virtual Computer Lab in the Cloud," CUNY Academic Works, New York, 2016.

[4] J. Walker, D. Freet, R. Agrawal and Y. Badr, "Open Source Cloud Management Platforms and Hypervisor Technologies: A review and Comparison," IEEE, Norfork, 2016.

[5] R. Tillinghast, D. Appel, C. Winsor and M. Mansouri, "STEM Outreach: A Literature Review and Definition," in 2020 IEEE Integrated STEM Education Conference (ISEC), New Jersey, 2020.

[6] K. K. P. Amy Aldous Bergerson, "CARES: Mentoring through University outreach," Journal of Higher Educaiton Outreach and Engagement, vol. 13, no. Number 1, p. 45, 2009.

[7] J. v. Byrne, "Outreach, Engagement, and the Changing Culture of the University," Journal of Public Service and Outreach, vol. 3, no. Number 2, pp. 3-8, 1998.

[8] M. Dennis, "Outreach initiatives in academic libraries, 2009 - 2011," Emerald, vol. 40, no. 3, pp. 363-383, 2010.

[9] K. A. G. Lisa S. Anderson, "Pre-university Outreach: Encouraging Students to Consider Engineering Careers," Global J. of Engng. Educaiton, Australia, 2003.

[10] D. Knighton, "Box Builds Ecosystem to Modernize Collaboration in Education," Canvas, LOS ALTOS, California, 2013.

[11] S. Fulton, "readwrite," 28 March 2012. [Online]. Available: https://readwrite.com/2012/03/28/box-launches-its-own-enterpris/.

[12] J. Sparapani, "TechTarget," 7 Feb 2018. [Online]. Available: https://searchcio.techtarget.com/blog/TotalCIO/Cloud-computing-ecosystem-Greater-than-the-sum-of-its-parts.

[13] Wikipedia, "Cloud Manufactoring," [Online]. Available: https://en.wikipedia.org/wiki/Cloud_manufacturing.

[14] P. Thomas, "A Sandbox in the Clouds: Software Testing and Development in Cloud Storage," 14 January 2020. [Online]. Available: https://en.wikipedia.org/wiki/Cloud_manufacturing. [Accessed 2 April 2021].

[15] P. University, "Pioneering naotechnology cloud--nanoHUB looks to future," Pioneering naotechnology cloud.

[16] Cisco, "What Is Cybersecurity?," Cisco, [Online]. Available: https://www.cisco.com/c/en/us/products/security/what-is-cybersecurity.html#~related-topics. [Accessed 22 04 2021].

[17] A. Valkonen, "Cloud Computing Ecosystem: Insights from an Exploratory Study in SaaS and Paas Value Networks," Aalto University, 2013.

[18] H. G. Labrador, G. Alexandropoulous, E. Bocchi, D. Castro, B. Chan, C. Contescu and M. Lamanna, "CERNBox: the CERN cloud storage hub," EDP Sciences, Meyrin, 2018.

[19] A. Cooke, "The 3 Main Types of PaaS," TrustRadius, [Online]. Available: https://www.trustradius.com/buyer-blog/the-3-main-types-of-paas. [Accessed 20 4 2021].

[20] C. o. Engineering, "College of Engineering," [Online]. Available: http://www.bu.edu/eng/departments/se/prospective-students/what-is-systems-engineering/.

[21] Leofrish, "Architecture and User Experience, Part 7: An Ecosystem Approach," Phase 2, 03 02 2017. [Online]. Available: https://www.phaseiidesign.com/posts/architecture-user-experience-part-7-ecosystem-approach/. [Accessed 03 05 2021].

[22] M. Rouse, "TechTarget," [Online]. Available: https://searchitchannel.techtarget.com/definition/cloud-ecosystem.

[23] "Wikipedia," 20 Feb 2019. [Online]. Available: https://en.wikipedia.org/wiki/Systems_engineering.

[24] MailChimp, "MailChimp," [Online]. Available: https://mailchimp.com/.

[25] A. M. O. A. P. H. Hugh P. Shanahan, "Bioinformatics on the Cloud Computing Platform Azure," Plos One, 22 7 2014. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0102642. [Accessed 20 4 2021].

BEAM: A Network Topology Framework to Detect Weak Signals

Hiba Abou Jamra, Marinette Savonnet, Éric Leclercq
Computer Science Department - EA 7534
University of Burgundy
9, Avenue Alain Savary, F-21078 Dijon - France

***Abstract*—In these days, strategic decision making and immediate action are becoming a complex task for companies and policymakers, since the environment is subject to emerging changes that might include unknown factors. When facing these challenges, companies are exposed to opportunities for growth, but also to threats. Therefore, they seek to explore and analyze large amounts of data to detect emerging changes, or so-called weak signals, that can help maintaining their competitive advantages and shaping up their future operational environments. But due to the increasing volume of daily produced data, scalable and automated computer-aided systems are needed to explore and extract these weak signals. To overcome the automation and scalability challenges, and capture early signs of change in a big data environment, we propose a framework for weak signals detection relying on the network topology. It is implemented under the Cocktail project framework whose goal is to create a real-time observatory of trends, innovations and weak signals circulating in the discourses of the food and health sectors on Twitter. This method analyses quantitatively the network local structure using the graphlets (particular type of motifs) to find weak signals. It provides accordingly qualitative elements that contextualise the identified signals, which will allow business experts to interpret and evaluate their dynamics to determine which ones may have a relevant future. After testing this method on different types of networks (we present two of them in this paper), we proved that it is able to detect weak signals and provides a quantifiable signature that allows better decision making.**

***Keywords*—Weak signals; network analysis; network topology; graphlets**

I. Introduction

In recent years, we have noticed an increasing demand to anticipate emerging issues or so-called weak signals, which could help making strategic decisions for future opportunities or threats. In a complex and dynamic world where information circulates rapidly, companies must monitor their environment, in order to adapt and anticipate the market. Monitoring the environment is important in detecting and acting on weak signals, just as it is important to pay attention to transmitted signals at the periphery of the companies' activities without being distracted by the mass of available information. Let's take the example of a soft drink manufacturer who competes with manufacturers of water and new types of beverages instead of traditional colas. The solution to mitigate the risks consists in broadening one's field of vision and thinking by analyzing different sources, such as social networks, blogs and specialized newspapers where one's competitors and customers can express themselves. The information produced by social networks allows decision-makers to have a peripheral and broader vision, and constitutes a source for finding weak signals announcing future threats or opportunities for the company [1], [2]. However, identifying these weak signals usually hidden amidst a large amount of noisy and strong information, and interpreting them when events are not known in advance, becomes a complex problem for researchers [3].

As early as 1975, Ansoff was among the first researchers exploring the field of weak signals. Afterwards in [4], he explained that a weak signal contains partial information available at the time a response is needed. This information must be completed before the signal becomes strong and has an impact on the company. Thus, the identification of a weak signal may announce a future event, its nature (opportunity or threat), its potential impact and the delay before it occurs. The process of carrying out these steps has been described by Ansoff as a graduated response by amplifying weak signals in relation to the level of information. Hiltunen [5] introduced the concept of future sign in which she defined the weak signal using a three-dimensional model: signal, event and interpretation. She aimed to describe a weak signal using the criteria of its visibility, diffusion and amplification. Other terms and definitions qualifying weak signals by author and field of research, are given in Table I.

TABLE I. Different Terms from the Literature to Define a Weak Signal, Classified by Fields of Research

Author	Field of research	Description and terms
Godet (1994)[6]	Economy, strategic prospection	A factor of a big change hardly perceptible at present; Component of a strong trend in the future.
Coffman (1997)[7]	Business management studies	Initiator of an important event, of a future trend; Signal beyond perception, or within perception but unrecognizable.
Hiltunen (2010)[8]	Event-related future studies	Early information or first symptoms of an emerging event; The emerging event itself.
Welz (2012)[3]	Social media, innovation management	Incomplete and imprecise information that is often not perceived; Evidence of strong discontinuities of present trajectory.

We hereby choose as definition for a weak signal:

Weak signals are information that provide an indication of upcoming or emerging events that may have significant implications. The information provided is often imprecise, ambiguous and incomplete.

We propose a method called BEAM[1] to detect weak signals and to help business experts in their interpretation. Although this method can be applied on different data sets, we implemented it under the context of Cocktail project to help companies to monitor their image on social networks (especially on Twitter), by detecting weak signals that might be a source of future threats or opportunities. Moreover, we provide a description of weak signal dimensions based on Hiltunen's model. The first two dimensions are translated by the signal support which is a graph built from the data of the study corpus, and a phenomenon/cause[2] which is a precursor of the event. For example, a tweet issued on April 14, 2021 by the official account of Game of Thrones television series, announces a potential remaking of the final season diffused earlier in 2019. The account tweeted "Winter is coming", just as it had done in previous years to announce a new season. Two hours later on the same day, the account again tweeted a video remembering a character's journey in the show. We can consider these tweets as the cause that triggered the weak signals announcing a new (or a remake) season of this show. The third dimension, interpretation, that we use in BEAM, depends on the competencies of experts or decision makers to give meaning to the detected weak signals.

Most weak signal detection approaches study the emergence of keywords using text-mining techniques. Other approaches look to identify the time points at which a behavior of change occurs, so they tend to extract interesting patterns from time series data. Another problem rises is that the data sources used to explore weak signals contain large volumes, which makes their processing and analysis a costly and time-consuming task, so that traditional methods often fail to unveil some of the strategic hidden information. Therefore, it is crucial to adapt a maximum of automation in the detection process. In our proposed approach, we do not consider these techniques neither those that are event detection-based.

Our method relies on the network topology since it allows to explore the local structure of the network and helps finding patterns that can constitute a quantifiable property of the weak signal. Our contributions are the following: 1) we automate the task of weak signals detection, and 2) we provide complementary elements that will help business experts in the interpretation of the detected signals, so that they finally judge their usefulness in a strategic decision or action.

The rest of this paper is organized as follows. Section II presents some of the related works of weak signals detection. We introduce the main steps of our method BEAM in Section III. Section IV presents two use cases and their results and Section V presents some limitations and a discussion. Finally, we conclude the paper and discuss some directions for future work in Section VI.

II. Related Work

In this section, we provide a synthetic description of some major works on weak signal detection, classified according to their detection techniques. Three main categories can be identified: 1) methods based on keyword mapping; 2) methods based on topic modeling; and 3) methods based on network motifs.

The first work is due to Yoon [9] who proposed a quantitative method based on text mining with keyword analysis to identify topics as weak signals in the field of solar panels. The method relies in the search of keywords having a low frequency of occurrence which reflects the visibility of the signal, and a high and increasing rate corresponding to its diffusion. The developed method proposes two metrics: the degree of visibility (DoV) represents the frequency of the keywords in the set of documents, the degree of diffusion (DoD) represents the frequency of the documents. A keyword with a low DoV and a low DoD is considered as a weak signal. He used these metrics to build the Keywords Emergence Map (KEM) and the Keywords Issue Map (KIM). However, Lee and Park [10] then Krigsholm and Riekkinen [11] raise two pitfalls: 1) the problem of uncertainty when the same keyword is at the limit of several quadrants or appears in different quadrants for both maps; 2) the problem of interpretation that occurs when there are several meanings for a given keyword. In addition, keywords related to weak signals are, in general, isolated terms and the absence of relationships and context limits the information in further interpretation. Griol-Barres et al. [12], [13] use a third measure, the degree of transmission (DoT) to assess the importance of keywords. DoT takes into account the documents' h-index where keywords appear. Kim and Lee's approach [14] is based on "document/keywords" matrices. A weak signal is seen as a rare or unusual keyword (outliers) and it is not related to existing topics (cohesion). Analyses are implemented with the Local Outlier Factor (LOF) algorithm [15].

The main interest of using topic modeling techniques is to infer topics from a corpus of documents. In [16], Maitre et al. used Latent Dirichlet Allocation (LDA) with Word2Vec to detect a topic related to a weak signal. For these authors, a weak signal is characterized by a small number of words per document, present in few documents, and unrelated to other topics. Other works have also focused on the detection of weak signals using dynamic LDA as in [17] and [18].

Some researchers have been interested in identifying specific patterns in the networks, called motifs, which could be considered as precursors of events. Baiesi [19] presented a method that studies the correlations between graphs from earthquakes, using tools on network theory. He found that simple patterns such as triangles are an interesting type of major event precursors because they were found in all three studied earthquakes. Kwon et al. [20] use Minimum Spanning Tree (MST) and apply betweenness centrality on the graphs. They consider that a weak signal appears as the node with the lowest centrality score. Moreover, Kim et al. [21] propose the NEST model that collects information from expert networks worldwide. They applied clustering, pattern recognition, and cross-impact analysis using a Bayesian network.

Others considered that weak signals can characterize real-world events, therefore the authors in [22] used tweets frequency and sentiment analysis in an attempt to detect weak signals, using tweets collected during the London riots in 2011. Another study using keyword and sentiment analysis was discussed in [23], where the authors proposed a model

[1]BEAM: a ray or shaft of light beams from the searchlights, a collection of nearly parallel rays (such as X-rays) or a stream of particles (such as electrons), a constant directional radio signal transmitted for the guidance of pilots.

[2]phenomenon = observed fact, normal or surprising event

for predicting a movie performance after its release, using the tweets frequency rates and sentiment analysis.

In contrast to the above studies, our work considers the diffusion and amplification of the signal as important criteria in the detection process (as seen in [5]). In addition, when dealing with social data, it is difficult to apply topic modeling and LDA techniques. Social data consists of short texts (for example 280 characters at most in a tweet), and often contains abbreviations and spelling errors. Moreover, as opposed to these works, our method is not limited to user-generated content analysis but considers a multi-dimensional approach consisting of various data analysis and visualization methods - both quantitative (detection) and qualitative (interpretation). Therefore, we rely on the works related to the identification of specific patterns from the network, since they support our hypothesis that graphlets, which are network particular patterns, can be weak signals.

III. Proposed Framework

Our goal is to provide a weak signal signature; this signature is a quantifiable property of signals that is characteristic of a weak signal [24]. We want to detect weak signals from social network data particularly, that can be represented as an interaction graph between entities. Therefore, the proposed method is based on a topological analysis of the network, in which we have chosen the graphlets as an operational description of weak signals. Indeed, graphlets present characteristics generally associated with weak signals. According to [25] and [26] they are:

- fragmentary and not very visible because they are small sub-graphs;
- meaningless, in fact considering only a sub-graph of at most 5 nodes does not mean much in the mass of data produced by social networks;
- interpretable by business experts by means of their predefined shapes and orbits.

We assume that graphlets allow to automate the task of detecting weak signals in a large volume of data while leaving room for interpretation by experts, thus eliminating the black box effect that a fully automated method could have.

Our method is based on a standard data processing pipeline that includes data collection and exploration, before the detection of weak signals, as illustrated in Fig. 1 and detailed in the rest of this section. In the first step, we prepare the collected raw data, and build the study corpus which is then divided into snapshots of the same duration, for a more refined study. In the second step, we examine the data of the study corpus by applying algorithms and measures for network topology analysis for each snapshot. Then, we identify precursors using different filtering and selection methods. A precursor is an observable and clear fact present in an organization's operational process and caused by existing factors in the process [27]. In the last two steps, we define criteria to qualify the identified precursors as weak signals, and provide contextual elements to interpret them.

A. Raw Data Preparation

BEAM applies to all data types that can be modeled in the form of a graph. For example, we can find such data on the platform https://www.data.gouv.fr/fr/ which makes available French public data. Data from social networks like Twitter and Facebook can be modeled as a labelled graph. However, their semantics is complex because the relations are multiple and context dependent. For example, thanks to a graph we can represent relationships between individuals that can be reciprocal or not, hostile or favorable. Specialized forums and scientific blogs like StackOverflow or MathOverflow can also be modeled as a graph representing interactions between entities.

TABLE II. Example of Interactions Between Entities

u	v	t	i
1	2	1643207148	RT
1	2	1643207149	Q
2	3	1643207149	RT
1	3	1643207150	RT
2	1	1643207150	Q
2	3	1643207150	F
2	4	1643207150	R

BEAM takes as input data a list of interactions where each row represents a four-component tuple of the form (u, v, t, i), where u and v are the identifiers of the entities in contact (e.g., accounts of Twitter users, or two hashtags that are in the same tweet), t determines the time at which the interaction took place. The time can take different formats such as a string or a timestamp epoch[3], etc. The i component represents the type of interaction. For example, Twitter has often been modeled as a graph in which the nodes represent users (or hashtags and urls), and the edges represent different types of interactions that exist between the nodes, like a user retweeting (RT), quoting (Q), replying (R), or following (F) another user, or the co-occurrence of hashtags. As well as replies or mentions on other users' Facebook posts, questions and answers on a scientific forum, etc. Table II shows an example of interactions. If several interactions took place at the same time, we find several rows with the same value on the t component in the interaction list. In this table, the first two rows represent interactions between entities 1 and 2 at two different dates ($t = 1643207148$ and $t = 1643207149$) with two different types of interactions. The last four rows of the list represent four existing interactions between different pairs of entities for the same time ($t = 1643207150$) with different types of interactions.

In the following, we describe the first three phases of the data preparation step.

1) Raw Data Cleansing and Filtering: The global corpus is composed of the raw list of interactions before being processed and filtered. Based on the raw list and the context of the study, it is possible to reduce the data in this corpus by specifying criteria to select only the data that matches the chosen criteria, such as filtering tweets containing hashtags about a topic of interest or some interactions of interest i. Several existing interactions (sometimes having different types) between the same pair of entities at the same time, are grouped together.

[3]Number of seconds elapsed since January 1, 1970 at 00:00; e.g. "26/01/2022 15:25:48" corresponds to timestamp 1643207148.

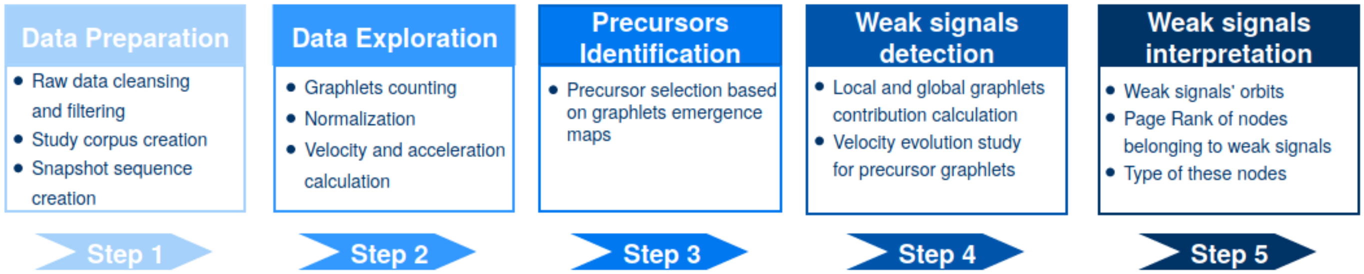


Fig. 1. The Main Steps of our Method BEAM.

This phase is important because the detection of weak signals only makes sense in a well-defined context, and the quality of the filtered data determines the reliability of the decision-making process.

2) Creation of the Study Corpus: Once the data has been filtered according to the chosen criteria, it is transformed into a new set that corresponds to the study corpus. Generally, the study corpus comprises a smaller volume of data than the global corpus. When we do a posteriori weak signals detection, the end of the study period must be determined.

3) Snapshot Sequence Creation: The interaction data can be represented in the form of a graph $G(V, E)$ where V is the set of entities and $E \in V \times V$ is the set of edges modeling the relationship between the pairs of entities, if $(u, v) \in E$ then the nodes u and v are linked together. Temporal interactions between a set of nodes over the time on a period T = $[\alpha, \beta]$ can be formulated by:

$$\mathfrak{X} = \{(u, v, t)\ t \in T \text{ and } u, v \in V\},$$

such that $(u, v, t) \in \mathfrak{X}$ indicates that u interacts with v at time t, these data can be ordered over time t. Temporal interactions do not admit a unique representation: some researchers study them using augmented graphs that integrate temporal information, others still, study them as link streams, others still with a temporal sequence of static graphs [28]. The link stream view was formally defined by Latapy et al. [29]: a link stream $L = (T, V, E)$ is defined by a time interval $T \subset \mathbb{R}$, a set of nodes V, and a set of edges $E \subseteq V \times V \times T$, where $(u, v, t) \in E$, denotes that the nodes u and v have interacted at time t. The augmented graph view represents temporal interactions within a finite set $\mathfrak{X}$, in the form of a single graph. A graph is created in which a node is a pair (v, t), with $t \in T$ and $v \in V$, and in which the node (u, t_i) is related to the node (v, t_j) if they interact, i.e. if $t_i = t_j = t$ and $(u, v, t) \in \mathfrak{X}$, or if they are contiguous in time, i.e. if $u = v$ and $t_j > t_i$ [30]. A second representation is to put temporality on the links.

The representation that we adopted to analyze the temporal interactions is a sequence of $s \in \mathbb{N}$ static graphs: $\mathfrak{S} = \{S^i : i \in \{1, \ldots, s\}\}$ where S^i called snapshot i, is the undirected and unweighted graph (see top part of Fig. 3) containing all the interactions that occurred between times $t_i = \alpha + i\Delta t$ and $t_{i+1} = \alpha + (i+1)\Delta t$, Δt is a constant duration for all snapshots representing one day, one hour, 30 minutes, 10 minutes, etc. The aim of Δt is to connect nodes as a function of time, such that two consecutive snapshots S^t and S^{t+1} are Δt-adjacent. Note that we have grouped the interactions in a time window.

Formally,

$$S^i = (V_i, E_i) \begin{cases} V_i = \{u_i : (u_i, v_i, t) \in \mathfrak{X} \text{ and } t \in [t_i, t_{i+1}[\} \bigcup \\ \qquad \{v_i : (u_i, v_i, t) \in \mathfrak{X} \text{ and } t \in [t_i, t_{i+1}[\} \\ E_i = \{(u_i, v_i) : (u_i, v_i, t) \in \mathfrak{X}, t \in [t_i, t_{i+1}[\\ \text{and } u_i, v_i \in V_i\} \end{cases}$$

Where u_i, v_i are the nodes u and v in the snapshot S^i.

B. Data Exploration

In the following, we explain our BEAM method for finding a quantifiable property of weak signals based on graphlets.

1) Graphlets: Graphlets were first introduced by [31]. A graphlet is a connected induced[4] non-isomorphic[5] subgraph (2 to 5 nodes) chosen among the nodes of a large graph. There are 30 different graphlets ranging from G_0 to G_{29}[6]. An essential element in the context of graphlets are the orbits [25]. They represent the positions (or roles) occupied by the nodes of these subgraphs. There are 73 different positions (from O_0 to O_{72}) for the 30 graphlets. The graphlets up to 4 nodes are presented in Fig. 2, with their corresponding orbits numbered from 0 to 14; in a same graphlet, the orbits having a same color are exchangeable.

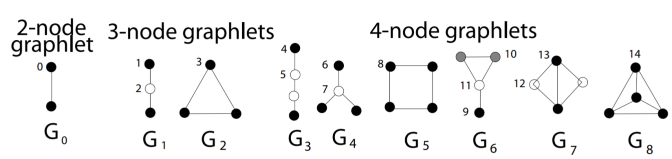


Fig. 2. Representation of the First 9 Graphlets Going from 2 to 4 Nodes.

There are several algorithms for enumerating graphlets and orbits in a graph [32]. In order to choose the most appropriate

[4]In graph theory, an induced subgraph is a subset of the nodes and **all** their connecting edges of an original graph.

[5]In graph theory, an isomorphism of two graphs G and H is a correspondance between the node sets of G and H, such that if any two nodes are adjacent in G, they are adjacent in H. Graphlets are non-isomorphic because they do not have the same form.

[6]In this article, we use the term graphlet for each type out of 30, however it does not represent its occurrence.

algorithm to count graphlets and orbits in the studied graph structures, we defined four criteria:

1) the exact enumeration of graphlets;
2) the enumeration of the orbits;
3) an acceptable complexity;
4) the availability of the source code.

The first criterion ensures the completeness of the graphlets. The second one ensures the interpretability of the results by studying the shape and role of the nodes in the graphlets. Orca satisfies our requirements. It was proposed by Hočevar and Demšar in 2014 [33], which is an exact enumeration algorithm. It counts graphlets up to 5 nodes and enumerates the orbits. It uses an analytical approach based on a matrix representation and works by setting up a system of linear equations per node of the input graph that relates different frequencies of orbits. Its source code is available at: https://rdrr.io/github/alan-turing-institute/network-comparison/src/R/orca_interface.R

With Orca, we count the number of graphlets for each snapshot. Each snapshot S^t is then represented as a vector of thirty elements (G_0^t, G_1^t, G_{29}^t), where G_x^t is the number of the graphlet G_x in the snapshot S^t.

2) Normalization: We apply a normalization procedure on the number of graphlets in order to reduce their values to a particular magnitude. The normalized values are used in the subsequent calculations. This step is of great importance because it should not cover the weak signals but make them comparable to others. The chosen procedure is the one proposed by D. Goldin and P. Kanellakis [34] in which they study the similarity between two queries in a temporal database. A query returns a sequence X of real numbers $(x_1, \ldots, x_n)$. Two reals a, b define a transformation $T_{a,b}$ on X by relating each x_i with $a \times x_i + b$. $\overline{X}$ represents the normal form of X, calculated by:

$$\overline{X} = T_{\sigma,\mu}^{-1}(X) = T_{\frac{1}{\sigma},-\frac{\mu}{\sigma}}(X)$$

Where $\mu(\overline{X})$ = 0 and $\sigma(\overline{X})$ = 1, μ being the mean and σ the standard deviation.

Applying this normalization procedure for each snapshot S^t where s is the number of snapshots, each component G_x^t of its vector with $x \in \{0, \ldots, 29\}$ is therefore normalized by:

$$\overline{G_x^t} = \frac{G_x^t - \mu(G_x)}{\sigma(G_x)}$$

with $\mu(G_x)$ the mean of each graphlet G_x for all snapshots, given by:

$$\mu(G_x) = \frac{1}{s}\sum_{t=1}^{s} G_x^t$$

and $\sigma(G_x)$, the standard deviation, calculated by:

$$\sigma(G_x) = \sqrt{\frac{\sum_{t=1}^{s}(G_x^t - \mu(G_x))^2}{s-1}}$$

3) Estimation of the Signal Reinforcement: Diffusion and Amplification: Velocity and acceleration evolution are quantitative features that allow us to evaluate the diffusion and amplification of the signal. We use these measures to identify event precursors among the graphlets.

From the obtained normalized values, the evolution of all $\overline{G_x^t}$ components is studied via their velocity and their acceleration. Our objective is to highlight the graphlets which emerge before others. For each snapshot and each graphlet G_x, we compute its velocity as the difference between the normalized value of the graphlet in snapshot S^{t+1} and the normalized value of the same graphlet at snapshot S^t. As for the acceleration, it is also calculated in the same way by making the difference between the velocities.

We therefore acquire a numerical matrix representing for each snapshot, the normalized value of the graphlets $\overline{G_x^t}$, their velocity $\overline{V_x^t}$, and their acceleration $\overline{A_x^t}$. This matrix is presented in the Table III which illustrates an example of the values of some graphlets per snapshot. Note that the velocities can only be computed from snapshot S^2, and the accelerations from snapshot S^3.

TABLE III. Example of Normalized Values of Graphlets, their Velocity and Acceleration for the Snapshots of the Study Period

Snapshot	Graphlet	$\overline{G_x^t}$	$\overline{V_x^t}$	$\overline{A_x^t}$
S^1	G_1	-0.952	NA	NA
S^1	G_2	-0.222	NA	NA
S^2	G_1	-0.666	-0.121	NA
S^2	G_2	0.456	0.678	NA
S^3	G_1	-0.786	0.938	1.059
S^3	G_2	0.758	0.302	-0.376
S^4	G_1	0.152	0.975	0.037
S^4	G_2	0.758	-0.315	-0.617
....	....	....	....	....
S^s	G_{29}	0.503	0.129	0.225

Fig. 3 summarizes the results obtained from the topological characteristics of the graphs for precursor discovery.

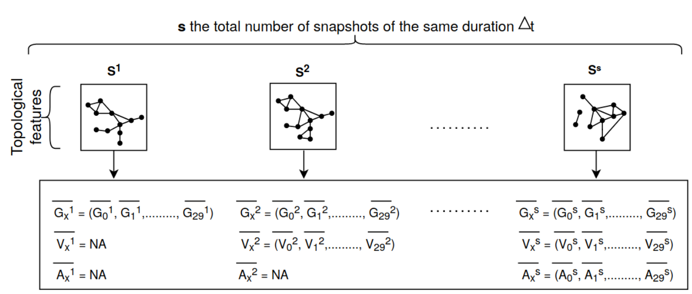


Fig. 3. Representation of the Calculation of Graphlets, Velocities and Accelerations for the Snapshots of the Study.

C. Precursors Identification

First, we present the principles of precursor identification and then its implementation set up. To identify precursors, we rely on Hiltunen's approach in which she considers two types of precursors: early information and first symptoms. The first type represents new information that appears suddenly, such as the announcement of a new product or invention. The second type represents a remarkable change that is difficult to interpret.

We have illustrated this representation in Fig. 4 where we consider four zones/quadrants. The "noise" area corresponds to graphlets with very small or negligible velocities and accelerations. The three remaining areas correspond to graphlets with 1) high velocity and low acceleration i.e. Q4, 2) low velocity and high acceleration i.e. Q1, 3) high values for both criteria i.e. Q2. If a graphlet is located in the areas Q1 and Q4 of the figure, we consider it as early information, whereas if it is located in the Q2 area, we consider it as first symptoms because the values of its velocity and acceleration are much higher. The graphlets located in these three zones are precursors of events because they are observable and clear facts or remarkable by their velocities and/or accelerations. The graphlets located in the noise zone are not considered as event precursors, because they represent unclear information which appears randomly but is not meaningful.

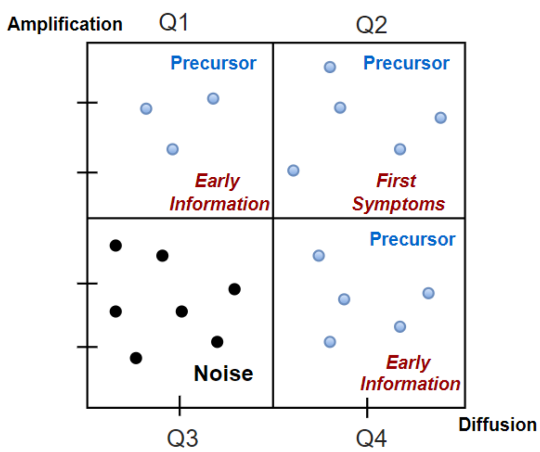


Fig. 4. Structure of the Graphlets' Emergence Map for a Particular Snapshot S^t.

Our aim here is to propose a division into graphlet emergence zones in the form of the four quadrants, not necessarily of equal size, as presented above. We propose two solutions:

1) **Slicing with respect to a top k:** The first possibility consists for each of the s snapshots S^t in
   - ranking their velocities $\overline{V_0^t}, \overline{V_1^t}, \overline{V_2^t}, \ldots, \overline{V_{29}^t}$ in ascending order;
   - settting a top k value for which the k graphlets have the highest velocities.

   Similarly, we set a top k for the graphlet acceleration criterion.
2) **Slicing with respect to a threshold μ:** The second possibility consists in setting a threshold for each of the s snapshots S^t equal for example to the average of the velocities and in selecting the velocities greater than or equal to this value, $\overline{V^t} \geq \mu(\overline{V^t})$. A similar division is made for the accelerations.

For a snapshot S^t, we create an emergence map that represents on the x-axis the velocity values, and on the y-axis the acceleration values. The structure of this map is based on the representation of the signals as seen in Fig. 4, where we place the graphlets in the four quadrants of the map. The division into quadrants depends on the choice of one of the two slicing methods presented above.

D. Identification of Weak Signals

The aim of this step is to qualify the graphlets as weak signals. We want to keep the True Positives and introduce the True Negatives as weak signals. False Negatives are graphlets that are neither precursors nor weak signals, and False Positives are graphlets that have been identified as precursors, but are not weak signals, they should be eliminated at the end of this step. The Table IV qualifies the results obtained by BEAM and estimates them.

TABLE IV. Qualification of the Results Obtained by BEAM

	Weak signal	$\neg$ **Weak signal** (noise, strong signal)
Precursor	True positive	False positive
$\neg$ **Precursor**	True negative	False negative

To qualify graphlets as weak signals, our BEAM method relies on:

1) Graphlets contribution in weak signals;
2) Evolution of the signal's diffusion.

The first measure aims to detect the weak signals at a certain time t, and the second one confirms or denies the detected signals, for an interval of time following t.

1) Contribution Ratio Calculation: We look in this phase to measure the contribution of a graphlet according to all graphlets to verify whether they are weak signals or not. The qualification criterion here is a ratio calculation. We propose two calculations: a local ratio that takes into account only the current time window, and a global ratio that takes into account all the studied windows. The number of graphlets in the chosen time window t is divided by the total number of graphlets for this time window:

$$R_{Local}(G_x) = \frac{G_x^t}{T_{Local}(G)}$$

Where $T_{Local}(G) = \sum_{x=0}^{29} G_x^t$.

The total number of a graphlet in the set of studied snapshots is divided by the total number of graphlets for all this set of snapshots, as follows:

$$R_{Global}(G_x) = \frac{\sum_{t=1}^{s} G_x^t}{T_{Global}(G)}$$

Where $T_{Global}(G) = \sum_{t=1}^{s}(\sum_{x=0}^{29} G_x^t)$, with s the number of processed time windows.

The resulting ratios $R_{Local}(G_x)$, $R_{Global}(G_x)$ are ranked in ascending order to qualify graphlets as weak signals if they are at the top of the list; the other graphlets are eliminated. This calculation is related to the **rareness characteristic** of weak signals since it selects the weakest ratios. Finally, at the end of these different analyses, we can provide per snapshot S^t a signature of the weak signal as a vector of at most 29 components: $(G_x^t, \ldots, G_y^t)$ with $x, y \in \{0, \ldots, 29\}$.

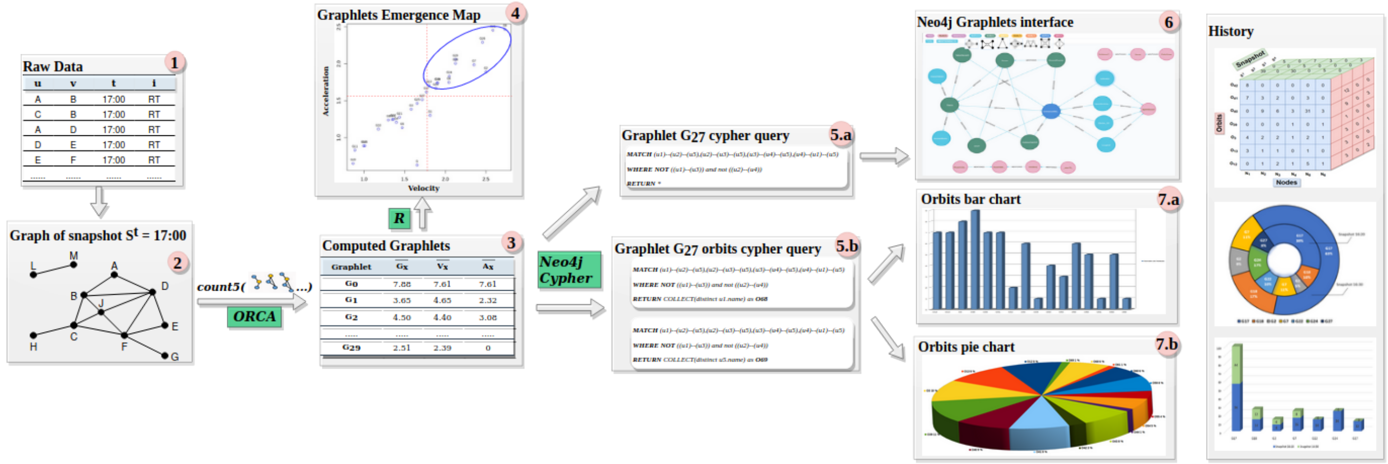


Fig. 5. BEAM user Interface for a Particular Snapshot. On the Right a Historical Panel for All Snapshots.

2) *Evolution of Graphlet Velocities:* The aim here is to propose a visual support to the identification of weak signals that can be carried out in parallel with the creation of emergence maps. This support consists in monitoring the evolution of the graphlet velocities over the time. To do this, we build a heatmap to visualize the evolution of the values of this criterion over a window of snapshots surrounding S^t : $\left[S^{t-1}, S^{t+3}\right]$. For example, if a graphlet G_x is found among the precursors in S^t, and if its velocity decreases in S^{t+1} and this decrease remains in S^{t+2}, then we consider that it is a false alarm thus we will not qualify it as a weak signal. On the contrary, the graphlets whose velocity increases in the studied interval, are True Positives and qualified as weak signals. Similarly, those that were not identified among the precursors but their speed increases, are True Negatives.

E. Interpretation of Weak Signals

Searching for relevant data for important event anticipation has several challenges including the complexity of detection and interpretation. At first sight, weak signals are fragmented pieces of information, so hard to be interpreted. However, if they were combined with context or additional elements, the interpretation task may become clearer and more relevant. Hiltunen [5] considers that the interpretation dimension means the information user's understanding of the future signal.

Our aim in this step is to make sense of the identified weak signals to help business experts in better and useful decision making. Several works tried to address this dimension like Lee and Park [10] who employed keyword clustering and topic selection based on keyword co-occurrences, or J. Kim [14] who proposed to observe the rarity and the paradigm unrelatedness of weak signals. In BEAM, we consider elements characterizing the nodes like their name, their type, as well as their role, etc. The role of a node can be characterized by its orbit, or its position in the graphlets. We therefore take advantage of this criterion to identify the central, cross or internal nodes roles. With Orca, we can also count how many times a node is playing these roles. BEAM also provides to business experts different visualizations that may help interpreting the identified signals, as shown in Fig. 5. This figure illustrates an example of one snapshot being studied by an expert, where parts 1 and 2 represent the raw interactions data at this snapshot, and the corresponding generated graph. Once Orca is executed, it allows the user to fetch all computed graphlets (their normalized number, their velocity and their acceleration) shown in part 3, and allows him to spot the precursor graphlets on the emergence map (part 4). In parts 5.a and 5.b, the expert may look further for a particular precursor graphlet using the Neo4j Cypher queries implemented in BEAM. Consequently, he has two visualization options, either 1) with a view that highlights the precursor graphlet (here G_{27} for example) in the initial graph, as in part 6, or 2) with a chart view representing the number of nodes in each of the precursors' orbits (including G_{27}), as shown in the bar and pie charts (parts 7.a and 7.b). At the right of this figure, there is a historical panel that allows the user to go back to a visited visualization, or to visualize the data based on a sequence of snapshots (in the form of a data cube, a clustered bar, or pie chart).

IV. Empirical Case Study

In order to evaluate the proposed method, we performed experiments on different datasets. We will describe in this section the empirical studies carried out on the Game of Thrones Twitter network (GOT), and a sensor network holding interactions between elementary school students. The purpose behind these studies is to 1) ensure the robustness of the method by verifying that weak signals are detected prior to a certain event (GOT), but are not detected when the event is already foreseen and punctual (elementary school); and 2) confirm its reproducibility through 6 episodes of GOT.

A. Game of Thrones Dataset

With long-tail streaming, HBO has pegged the last season of GOT as averaging around 44.2 million viewers per episode. The final season of GOT has upset many viewers because of changes in writers, shortening the season (this last one had only six episodes diffused once per week), and a surprising turn of events. This is why the fan club was more into criticizing this final season. It was released on April 14 2019, and ended on May 19 2019. The episodes were diffused live at 9 p.m.

American time, and the collected data corresponds to the tweets published at the same timestamp. The original data sample including all episodes of the season, has a total number of 46 481 705 tweets including 34 094 365 retweets.

1) Study Corpus: We describe the study carried on the first episode with a period limited to 14-04-2019 from 1 p.m. to 7 p.m. The graph represents 223 383 retweets between 189 741 users corresponding to audience, journalists, media, etc. This graph is divided into a sequence of snapshots with a duration Δt equal to 10 minutes.

2) Precursors and Weak Signals: For each snapshot, we counted the number of 5-node graphlets and normalized them, then computed their corresponding velocity and acceleration. In the following, we will present only the obtained results for the snapshot of 5 p.m. Fig. 6 illustrates the graphlets' emergence map of 5 p.m built based on these criteria. The 13 graphlets that are circled in red are selected as precursors, since they belong to one of the aforementioned zones Q1, Q2 or Q4. The remaining graphlets are considered noise being positioned in the Q3 zone.

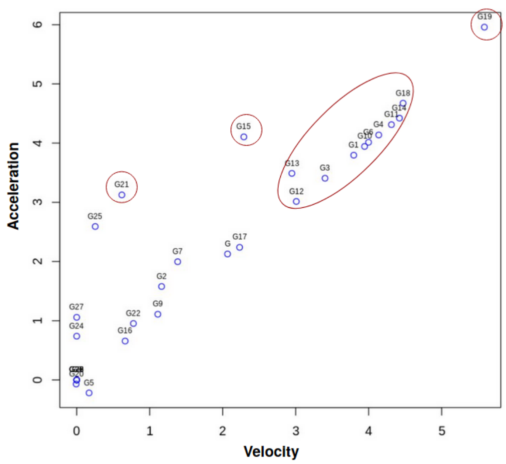


Fig. 6. Precursors Selected Among the Graphlets in the Snapshot of 5 p.m.

Once the precursors are selected, the next step is to check their relevance by identifying the false positives, preserving the true positives and adding the true negatives. For this, we first calculated the local and the global contributions of the graphlets for each snapshot, and ranked the resulted ratios in ascending order. For the 5 p.m. snapshot, we selected the top 5 graphlets according to the ranked ratios, that are respectively G_{21}, G_{15}, G_{13}, G_{12} and G_3. We moved next to confirm the obtained contribution ratios, by analyzing the evolution of graphlets' velocities in the snapshots surrounding 5 p.m.

We monitored the evolution of graphlets (precursor ones in particular) velocities during an interval of snapshots going from 4:50 to 5:30 p.m., as shows the Fig. 7. The x-axis of the heatmap in this figure represents the different snapshots, and the y-axis represents some of the precursor graphlets. The legend on the right of the figure represents the graphlets velocity values between -2 and +5. The blue colors in the heatmap are the graphlets G_x such that $\overline{V_x^t} \geq +5$, and the red colors are those such that $\overline{V_x^t} < 0$. For example, when considering the 5:00 p.m. snapshot, the graphlet G_{19} shows a significant increase in velocity up to +5, and in the next snapshot, it decreases rapidly to -2. Another evolution is the one of G_{16}, that shows continuous decrease in its velocity for all snapshots. We revised the global contributions of these graphlets and noticed that they have higher ratios than the top 5 mentioned above. These results are therefore consistent, and indicate that these graphlets are false alarms. As for the remaining graphlets, we see that they show a significant rise in their velocities at 5 p.m., followed by a minor decrease at 5:10 p.m., then another rise at 5:20 p.m. The decreases at 5:30 p.m. do not affect the analysis results, because it is enough to observe one remarkable rise after the current snapshot (i.e. 5 p.m.). These last results are consistent with those obtained from the contribution calculation, hence we qualify the top 5 graphlets from the list as weak signals. Table V illustrates the five qualified weak signals, their shape and their global contribution.

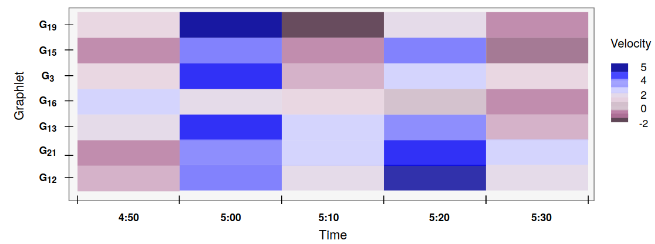
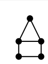
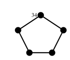
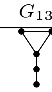
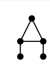


Fig. 7. Velocity Evolution of Some Precursors in the Snapshots Around 5 p.m.

TABLE V. TOP 5 GRAPHLETS QUALIFIED AS WEAK SIGNALS FOR THE SNAPSHOT OF 5 P.M.

Graphlet	G_{21}	G_{15}	G_{13}	G_{12}	G_3
Graphlet shape					
Graphlet global contribution	0.0009	0.0174	0.0235	0.0260	0.0283

3) Weak Signals Interpretation: Our objective here is to help in the interpretation of identified weak signals, by providing contextual elements like user accounts. We first start by calculating the number of orbits (nodes positions) of graphlets with Orca, then analyze these positions in weak signal graphlets. This is an important task because it can help explaining the users' role in assessing and supporting this final season.

Table VI illustrates an extract of the obtained results from this step at the 5 p.m. snapshot. The accounts `@dcucomics`, `@jonatas_maia12` and `@9GAG` appear mostly in a peripheral position in graphlet G_3 (O_4). The remaining accounts in this table occupy a central position (O_5). The third and fourth columns of the table represent respectively the different node types (from comic accounts, to online platforms or journalists), and their corresponding ranks after executing the PageRank algorithm on 7000 users. We note that the official accounts of the series are communicating more than the others since they occupy the most central positions. In addition, the account `@TylerIAm` has a you-tube channel and diffuses live broadcasts from 12 to 3 p.m., so from his PageRank we can see that he retweeted very much about GOT two hours after the broadcast.

TABLE VI. EXTRACT OF IMPORTANT NODES BELONGING TO G_3 ORBITS, AND THEIR PAGE RANK FOR THE SNAPSHOT OF 5 P.M.

Graphlet-Orbit	Account	Account type	PageRank
O4	@dcucomics	DC universe (fictional universe produced by Warner Bros) fan account	27
	@jonatas_maia12	Product designer	61
	@9GAG	Online platform, viral and funny videos	248
O5	@GameofThrones	Official account of GOT	1
	@Thrones_Memes	GOT memes account	3
	@TylerIAm	Journalist	4
	@LordSnow	Got character (John Snow)	12

We then conducted a refined analysis on the account `@TylerIAm` using the Neo4j Cypher queries implemented in BEAM. Here we went from a graphlet type to 4 graphlet instances, and displayed the users linked to this account on a G_3 instance. As shows Fig. 8, `@Woodlawnwonder`, a blogger, occupies the other central position of G_3, and the peripheral positions are occupied by the remaining accounts. We considered the instance containing `@joestudz18`, an artist and concept illustrator, as well as `@holy_schnitt` who is a social media star. We assume that the blogger account activity for example, should be monitored by business experts.

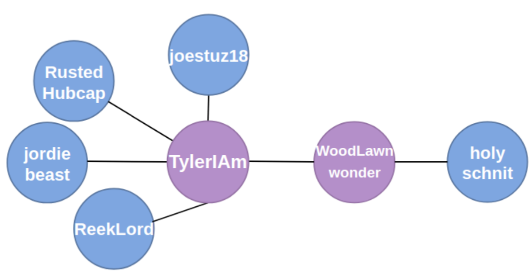


Fig. 8. Important Accounts Occupying Orbits of an Instance of G_3 in the Snapshot of 5 p.m.

We also noticed an interesting fact about nodes appearing in G_{15}. These nodes belong to accounts located in Brazil, including a comic and digital creator `@cleytu`, an entertainer `@fabwiano`, the Brazilian version of HBO channel `@HBO_Brasil`, and a journalist `@LethyciaDias_`, etc. We then executed the Louvain algorithm for the whole episode (where all snapshots are combined). We noticed that all these users belong to the same community. We therefore allow business experts to give their final assessment regarding these results and the type/location of the users, since the shape of this graphlet can put into evidence an important social structure [26].

We performed the same experiment on the retweets published on the remaining 5 episodes of GOT, to test the reproducibility of our method. The results obtained showed that the weak signal graphlets are detected few hours before the episode is broadcast. Moreover, we noticed that the same graphlets are found in several episodes as weak signals.

According to the results of our study, the official GOT accounts (`@GameofThrones`, `@ThronesMemes` and `@LordSnow`) communicated supporting information (they were motivating the audience to watch GOT, and highlighting the crucial role of the character Jon Snow). Therefore, the weak signals carried by these accounts can be opportunities, which is important that they emerge to verify if the speech is well perceived. On the other hand the comic accounts `@dcucomics` and `@9GAG` can be threats because they can make fun of GOT or even divert its image, so they must be watched to mitigate their negative impact on the audience.

B. *Elementary School Interactions Dataset*

We conducted another experiment on students interactions in an elementary school in Lyon, France, on two consecutive days in October 2009 [35]. The school day runs from 8:30 a.m. to 5:30 p.m., with a lunch break between 12:00 p.m. and 2:00 p.m., and two 20-25 minute breaks at around 10:30 a.m. and 3:30 p.m. that take place in a common playground. Two students are considered to be interacting if they are within 3 to 5 feet of each other for at least 20 seconds.

Through this experiment, we aimed to:

1) verify that no weak signal is found when the events are known (as the breaks are scheduled at specific times, without particular triggers, . . .);
2) verify that the error rate is reduced, i.e. that the false positives are eliminated by our method.

1) Study Corpus: The raw data file was downloaded from the official site[7]. We divided the corpus into snapshots of 10 minutes duration each, and choose to work with five snapshots that correspond to the times before the lunch break (from 11:20 a.m. to 12 p.m.) of the first school day.

2) Precursors and Weak Signals: We focused our analysis on the snapshot of 11:40 a.m. and we were able to spot precursors on the corresponding graphlets' emergence map. Thereafter, we evaluated the located precursors using the graphlets contribution ratios, as well as their velocity evolution. The velocities heatmap showed that for this particular snapshot, the graphlets reach high values, but then decrease rapidly in the next snapshot (11:50 a.m.) and continue to decay until the last snapshot. This decrease supports the hypothesis that the identified precursors at this snapshot were false alarms, hence shall not be qualified to weak signals. Although we did not expose weak signals from this experiment, the dataset is a ground truth example on which we relied to confirm the method's objectives.

C. *Results Validation: Cross-Correlation*

This is a supplementary step that aims to validate the discovery of weak signals by studying the relation between them and a potential event[8]. In this step, we observe the data no more as a sequence of graphs, but instead as time series.

[7] http://www.sociopatterns.org/datasets/primary-school-temporal-network-data/
[8] We consider an event as a situation in which the number of interactions reaches its maximum value.

A time series X is created from the number of interactions selected between all pairs of nodes. It is a sequence of n elements $X = (x_i)_{1 \leq i \leq n} = (x_1, x_2, \ldots, x_n)$.

We rely on the Cross-correlation to validate the intrinsic properties of BEAM by studying the relation between the graphlets weak signals and the original time series of interactions (built from the raw data). Cross-correlation[9] is a linear measure of similarities between two time series X and Y, which helps evaluate the relation between these series over time [36]. An offset/lag h is associated with this measure, knowing that if $h < 0$ then X could predict Y, and if $h > 0$ then Y could predict X.

We applied this function on the GOT dataset, by first building the time series corresponding to the users' retweets, then building a time series for each one of the 30 graphlets. From the obtained results, we noticed that the graphlets that were qualified as weak signals (see Table V), presented positive correlations (between 0.5 and 0.7) with the initial series with a lag of 10 or 20 minutes (Δt = 10 minutes). Fig. 9 illustrates the correlogram of graphlet G_{13} with the initial retweets series. The graphlets identified as false alarms in the previous steps of the method, like G_{16} and G_{19} for instance, did not present correlations.

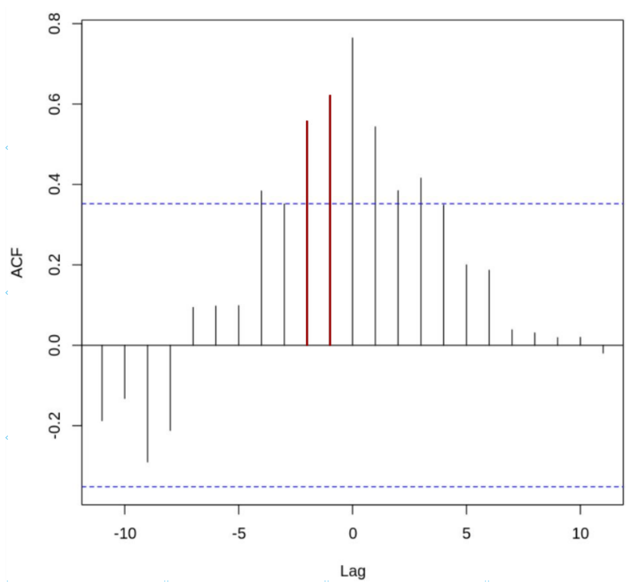


Fig. 9. Positive Correlation with Negative Lags of 10 and 20 Min Between G_{13} and the Retweets Time Series (GOT Dataset.)

We executed the Cross-correlation on the elementary students interactions dataset, and from the obtained correlograms, we did not find any negative-lag correlation between the initial time series and the precursor graphlets. These results confirm that when the event is already planned and punctual, there are no weak signals that can announce the occurrence of such event, hence the detection phenomena corresponds in this case to false positives that we were able to identify.

[9]implemented with the R package *tseries*: https://www.rdocumentation.org/packages/tseries/versions/0.1-2/topics/ccf

V. Limitations and Discussion

Several works relied on different techniques to detect future weak signals. After applying BEAM on several data sets, the findings have been quite consistent in establishing that this method is able to detect weak signals prior to an important event. Using BEAM, the business expert does not spend time anymore to choose selection filters, prepare data and extract remarkable information for analysis. Instead, he relies on the previously detected information and assesses if it should be qualified as future weak signals or not.

However, BEAM still presents some limitations related to: 1) The constitution of study corpus. 2) The snapshots duration Δt. 3) The division of graphlets' emergence maps. 4) The interpretation of the detected signal through its recognition.

These limitations are related to filters that hinder the analysis of weak signals and constitute barriers to their interpretation. These barriers were discussed by Ansoff [4] who suggested that weak signals must pass three filters 1) the monitoring (surveillance), 2) the mentality and 3) the power filter, before potentially triggering an action or a decision. The monitoring filter corresponds to the capacity of the weak signal to be detected or discovered, in the midst of all other perceived information, by one or more actors within the organization. The mentality filter refers to the capacity of the signal to be recognized, after being detected, as relevant information with regard to the examined situation. Finally, the power filter refers to decision-making once the signal has been detected and its relevance recognized. The people in charge in the organization can decide for example not to make this signal a priority, despite the underlying risk.

Limits 1, 2 and 3 of BEAM are related to the monitoring filter. The first one results from an irrelevant choice of keywords, accounts or hashtags (in case of Twitter data) during the creation of the study corpus. It can be resolved by adding a feedback loop in BEAM, to allow business experts to return to the first step and modify the selection criteria, if the interpreted signals are not significant for them. Regarding the second one, the duration of snapshots Δt should be well-chosen for an expert to analyze the detected signals and provide his decision. To do so, we performed an experiment to measure the time in which a tweet becomes viral (i.e. when it reaches the highest number of its retweets). This limitation is also depending on the number of tweets in the studied snapshots, if it is high, Δt should be limited otherwise increased. The third limitation relates to the choice of the top k or the threshold that divides the graphlets' emergence maps for the precursors selection. This one was resolved with the contribution calculation with which graphlets are added or removed from the list of weak signals based on their ratio. The fourth limitation is related to the mentality and power filters, thus cannot be resolved within BEAM, since it is up to the business expert to provide his final interpretation and decision regarding the detected signals.

VI. Conclusion

In this article, we presented our method BEAM whose objective is to help business experts to detect and interpret the identified weak signals in order to enable them to make decisions and plan future strategies. First, we find graphlets in a graph of interactions between entities, which represent clear and observable facts, quantifiable using measures of diffusion and amplification that characterize them as precursors. We examine the contribution of graphlets to eliminate false alarms (False Positives) and qualify True Positives and True Negatives as weak signals. Once these signals are identified, the shapes of the graphlets and their orbits help the business experts in their interpretation. In BEAM, we have chosen graphlets as an operational description to give a signature of the weak signal, this choice allows an automation of the detection task while enabling a final judgment by the business experts for a better decision making.

Next, new experiments will be performed on other types and larger networks to resolve the remaining limitations of the method. In further research, we would like to explore if graphlets can be used to indicate phase transitions of an information emergence between transition points. This will help in analyzing the weak signal amplification process. BEAM combines data analysis and visualization tools to guide business experts in detecting and interpreting weak signals, and offers a great potential for decision-making in most business organizations. Moreover, the detection of weak signals offers promising leads for innovation.

Acknowledgment

This work is supported by the Investissements d'Avenir program, Cocktail ISITE-BFC project (Initiatives Science Innovation Territoires Économie en Bourgogne-Franche-Comté), ANR contract 15-IDEX-0003, https://projet-cocktail.fr/.

References

[1] G. Day and P. Schoemaker, "Peripheral Vision: Sensing and Acting on Weak Signals," *Long Range Planning*, vol. 37, p. 117–121, 04 2004.

[2] J. Wu and A. Pinsonneault, "Enhancing peripheral vision through social media use: A social network perspective," in *AMCIS*, 2011.

[3] K. Welz, L. Brecht, A. Pengl, J. V. Kauffeldt, and D. R. Schallmo, "Weak signals detection: Criteria for social media monitoring tools," in *ISPIM Innovation Symposium*. The International Society for Professional Innovation Management (ISPIM), 2012, p. 1.

[4] H. I. Ansoff and E. J. McDonnell, "Implanting strategic management," 1990.

[5] E. Hiltunen, "The future sign and its three dimensions," *Futures*, vol. 40, no. 3, pp. 247–260, 2008.

[6] M. Godet, *From anticipation to action: a handbook of strategic prospective*. UNESCO publishing, 1994.

[7] B. Coffman, "Weak signal research, part I: Introduction," *Journal of Transition Management*, vol. 2, no. 1, 1997.

[8] E. Hiltunen, "Weak Signals in Organisational Futures," *Aalto University School of Economics, Aalto*, 2010.

[9] J. Yoon, "Detecting weak signals for long-term business opportunities using text mining of Web news," *Expert Systems with Applications*.

[10] Y.-J. Lee and J.-Y. Park, "Identification of future signal based on the quantitative and qualitative text mining: a case study on ethical issues in artificial intelligence," *Quality & Quantity*, vol. 52, no. 2, pp. 653–667, 2018.

[11] P. Krigsholm and K. Riekkinen, "Applying Text Mining for Identifying Future Signals of Land Administration," *Land*, vol. 8, no. 12, 2019. [Online]. Available: https://www.mdpi.com/2073-445X/8/12/181

[12] I. Griol-Barres, S. Milla, and J. Millet-Roig, "System Implementation for the Detection of Weak Signals of the Future in Heterogeneous Documents by Text Mining and Natural Language Processing Techniques," in *ICAART*, 2019.

[13] I. Griol-Barres, S. Milla, A. Cebrián, H. Fan, and J. Millet, "Detecting Weak Signals of the Future: A System Implementation Based on Text Mining and Natural Language Processing," *Sustainability*, vol. 12, no. 19, 2020. [Online]. Available: https://www.mdpi.com/2071-1050/12/19/7848

[14] J. Kim and C. Lee, "Novelty-focused weak signal detection in futuristic data: Assessing the rarity and paradigm unrelatedness of signals," *Technological Forecasting and Social Change*, vol. 120, pp. 59–76, 2017.

[15] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: identifying density-based local outliers," in *ACM sigmod record*, vol. 29, no. 2. ACM, 2000, pp. 93–104.

[16] J. Maitre, M. Ménard, G. Chiron, A. Bouju, and N. Sidère, "A Meaningful Information Extraction System for Interactive Analysis of Documents," in *International Conference on Document Analysis and Recognition (ICDAR)*, 2019, pp. 92–99.

[17] D. M. Blei and J. D. Lafferty, "Dynamic topic models," in *Proceedings of the 23rd International Conference on Machine Learning*, ser. ICML '06. New York, NY, USA: Association for Computing Machinery, 2006, p. 113–120. [Online]. Available: https://doi.org/10.1145/1143844.1143859

[18] L. Pépin, P. Kuntz, J. Blanchard, F. Guillet, and P. Suignard, "Visual analytics for exploring topic long-term evolution and detecting weak signals in company targeted tweets," *Computers & Industrial Engineering*, vol. 112, pp. 450–458, 2017. [Online]. Available: https://doi.org/10.1016/j.cie.2017.01.025

[19] M. Baiesi, "Scaling and precursor motifs in earthquake networks," *Physica A: statistical mechanics and its applications*, vol. 360, no. 2, pp. 534–542, 2006.

[20] L.-N. Kwon, J.-H. Park, Y.-H. Moon, B. Lee, Y. Shin, and Y.-K. Kim, "Weak signal detecting of industry convergence using information of products and services of global listed companies-focusing on growth engine industry in south korea," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 4, no. 1, p. 10, 2018.

[21] S. Kim, Y. Kwon, Y. Jeong, S.-B. Choi, J.-K. Park, and S.-W. Hong, "Nest: a model for detecting weak signals of emerging trends using global monitoring expert network." in *EKAW (Posters and Demos)*. Citeseer, 2010.

[22] C. Charitonidis, A. Rashid, and P. J. Taylor, "Weak signals as predictors of real-world phenomena in social media," in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, 2015, pp. 864–871.

[23] S. Asur and B. A. Huberman, "Predicting the future with social media," in *2010 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology*, vol. 1. IEEE, 2010, pp. 492–499.

[24] D. Georgiadis and M. Raubal, "An interdisciplinary review on weak signal detection: Future resilient systems working paper# 3," 2020.

[25] N. Pržulj, "Biological network comparison using graphlet degree distribution," *Bioinformatics*, vol. 23, no. 2, pp. e177–e183, 2007.

[26] R. Charbey, L. Brisson, C. Bothorel, P. Ruffieux, S. Garlatti, J.-M. Gilliot, and A. Mallégol, "Roles in social interactions: graphlets in temporal networks applied to learning analytics," in *International Conference on Complex Networks and Their Applications*. Springer, 2019, pp. 507–518.

[27] P. Korvers, "Accident precursors: Pro-active identification of safety risks in the chemical process industry." 2004.

[28] P. Holme and J. Saramäki, "Temporal networks," *Physics Reports*, vol. 519, no. 3, p. 97–125, Oct 2012. [Online]. Available: http://dx.doi.org/10.1016/j.physrep.2012.03.001

[29] M. Latapy, T. Viard, and C. Magnien, "Stream Graphs and Link Streams for the Modeling of Interactions over Time," *CoRR*, vol. abs/1710.04073, 2017. [Online]. Available: http://arxiv.org/abs/1710.04073

[30] V. Kostakos, "Temporal graphs," *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 6, p. 1007–1023, Mar 2009. [Online]. Available: http://dx.doi.org/10.1016/j.physa.2008.11.021

[31] N. Pržulj, D. G. Corneil, and I. Jurisica, "Modeling interactome: scale-free or geometric?" *Bioinformatics*, vol. 20, no. 18, pp. 3508–3515, 2004.

[32] P. Ribeiro, P. Paredes, M. E. Silva, D. Aparicio, and F. Silva, "A survey on subgraph counting: concepts, algorithms, and applications to network motifs and graphlets," *ACM Computing Surveys (CSUR)*, vol. 54, no. 2, pp. 1–36, 2021.

[33] T. Hočevar and J. Demšar, "A combinatorial approach to graphlet counting," *Bioinformatics*, vol. 30, no. 4, pp. 559–565, 12 2014.

[34] D. Q. Goldin and P. C. Kanellakis, "On similarity queries for time-series data: constraint specification and implementation," in *International Conference on Principles and Practice of Constraint Programming*. Springer, 1995, pp. 137–153.

[35] J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, L. Isella, J.-F. Pinton, M. Quaggiotto, W. Van den Broeck, C. Régis, B. Lina *et al.*, "High-resolution measurements of face-to-face contact patterns in a primary school," *PloS one*, vol. 6, no. 8, p. e23176, 2011.

[36] B. D. Ripley and W. Venables, *Modern applied statistics with S*. springer, 2002.

Soft-sensor of Carbon Content in Fly Ash based on LightGBM

Liu Junping[1], Luo Hairui[2], Huang Xiangguo[3], Peng Tao[4], Zhu Qiang[5], Hu XinRong[6], He Ruhan[7]
Hubei Provincial Engineering Research Center for Intelligent Textile and Fashion
Engineering Research Center of Hubei Province of Clothing Information
School of Computer Science and Artificial Intelligence
Wuhan Textile University, Wuhan, China[1, 2, 4, 5, 6, 7]
Hubei Provincial Exchange, Wuhan, China[3]

Abstract—The soft-sensor method of carbon content in fly ash is to predict and calculate the carbon content of boiler fly ash by modeling the distributed control system (DCS) data of thermal power stations. A novel data-driven soft-sensor model that combines data pre-processing, feature engineering and hyperparameter optimization for application in the carbon content of fly ash is presented. First, extract steady-state data by data mining technology. Second, twenty characteristics that may affect the carbon content in fly ash are identified as variables by feature engineering. Third, a LightGBM prediction model that captures the relation between the carbon content in fly ash and various DCS parameters is established and improves the prediction accuracy by the Bayesian optimization (BO) algorithm. Finally, to verify the prediction accuracy of the proposed model, a case study is carried out using the data of a coal-fired boiler in China. Results show that the proposed method yielded the best prediction accuracy and closely approximates the non-linear relationships between variables.

Keywords—LightGBM; carbon content; fly ash; soft-sensor; feature engineering; Bayesian optimization

I. INTRODUCTION

The unburned carbon content in fly ash reflects the combustion efficiency of a coal-fired boiler. The combustion condition of coal can be better evaluated by analyzing the unburned carbon content in fly ash [1]. Real-time monitoring of carbon content in fly ash helps keep the carbon content in fly ash within a reasonable range, thus reducing the cost of power generation and improving the economy of generating units.

Currently, the methods for detecting the carbon content of fly ash are divided into two categories: physical measurement methods and soft-sensor methods. Physical methods commonly include the loss-on-ignition method [2], Laser-induced breakdown spectroscopy [3], microwave absorption method. [4] etc. Physical solutions are not widely available due to technical or cost reasons [5-6]. Machine learning methods have been widely used in human life, industrial production and power generation [7-9]. Distributed control system (DCS) is a computer control system for centralised management and decentralised control of the production process [10]. The distributed control system contains many sensors, which record the information of system operation. By analyzing this information, we can predict the operation status of the system [11]. The soft-sense method organically combines the production process knowledge through mechanism analysis, which can quickly and accurately reflect the carbon content in fly ash under different working conditions, and has a high economy.

Currently, there are three main problems with soft-sensor methods for the carbon content in fly ash:

1) The boiler combustion process is a multivariable variable, nonlinear and highly coupling thermal process [12]. For example, the DCS records variables such as air volume, air pressure, and air temperature for each coal mill outlet. These variables are highly correlated with boiler combustion prediction modeling, resulting in a certain amount of variable redundancy, affecting the model estimation accuracy, and increasing the computational complexity. Therefore, it is necessary to apply feature engineering to reduce the impact caused by redundant variables.

2) Most current research tests have limited data and working conditions. They do not effectively represent the complete operational status of the boiler.

3) The accuracy of these algorithms is limited.

II. RELATED WORK

Zhou et al [13]. established an artificial neural networks (ANN)-based soft-sensor model for the carbon content in the fly ash of a 300MW utility tangentially firing coal burned boiler and verified the validity of the model by multi-state thermal experiments. Wang et al [14]. proposed building a prediction model with support vector regression(SVR) for carbon content in fly ash and showed through experiments that the carbon content in fly ash model using SVR has reliable generalization and is suitable for online modeling. In machine learning, finding appropriate data processing methods, such as removing noise data and extracting suitable features, will help to improve the accuracy of prediction [15]. To address these issues, Zhu et al [16] performed a sensitivity analysis of the related features for the carbon content in fly ash, using the Garson algorithm for variables selection before modeling. Wang [12] collects the factors influencing the carbon content in fly ash constitute the initial variables, and the optimal variables are selected by the random forest-based variable selection method. The machine learning model contains many super parameters, such as penalty, learning rate and loss function. A suitable combination can effectively improve the prediction accuracy of the model [17]. Feng [18] improves the model generalization ability by using the genetic algorithm to optimize the values of each neural network parameter. Peng [19] proposed an adaptive perturbation quantum particle swarm optimization algorithm (AQPSO) with a support vector machine to jointly predict the carbon content in fly ash and improve the prediction accuracy of the SVR model by ADQPSO.

LightGBM [20] is an ensemble learning algorithm, Developed by Microsoft in 2017. It is an advanced implementation of the distributed gradient boosting decision tree (GBDT) framework. The GBDT [21] algorithm is the core of LightGBM, which iteratively sums weak estimators to generate robust estimators by computing the negative gradient of the loss function. Lightgbm integrates GOSS (Gradient-based One-Side Sampling) algorithm and EFB (Exclusive Feature Bundling) algorithm based on GBDT. GOSS algorithm can lead to a more accurate gain estimation than uniformly random sampling, and the EFB algorithm provides a nearly lossless approach to reduce the number of effective features [20]. LightGBM algorithm extensively applied in many regression problems [22-23].

Hyperparameters play a valid role in the accuracy of regression prediction algorithms. In practice, it is necessary to continuously adjust the hyperparameters, train the model under different sets of hyperparameters, and determine the best hyperparameters by comparing the model's performance. Therefore, finding the appropriate hyper-parameters has become a critical issue in machine learning [24].

Bayesian optimization (BO) is a very effective global optimization algorithm. BO is very suitable for solving highly complex optimization problems. Their objective functions could not be expressed, or the functions are non-convex, multimodal, and computational expensive [25]. BO can actively select appropriate evaluation points according to the relevant results of the current unknown function to avoid unnecessary sampling. At the same time, Bayesian optimization can use historical search information to improve search efficiency. [26]. BO has achieved better results than other hyperparameter tuning methods in the Black-Box Optimization Challenge 2020 [27].

III. Proposed Work

In this study, a new soft-sensor method for measuring the carbon content of fly ash is proposed by analyzing and experimenting with a total of 3,272,872 DCS data from an electric boiler from October 23 to November 30, 2020. The method combines data mining, feature engineering, LightGBM, and BO algorithm. A flowchart of the applied methodology is proposed in Fig.1. By comparing with other feature selection methods, and prediction models, experiments show that the prediction results of the presented approach are closer to the actual working conditions of the carbon content of fly ash, which improves the soft-sensor accuracy and ensures the reliability and accuracy of the soft-sensor method.

The data processing part is the operation of apparent outlier removal and re-sampling of the acquired DCS data.

A. Apparent Outlier Removal

First, the raw data was examined, and remove the data are outside the reasonable range. For example, the actual load recorded by the DCS has some invalid data at the beginning due to plant shutdown, etc. As shown in Figure 2, the data in the red area are unreasonable. By removing apparent outliers, the natural distribution of the variables can be captured.

The load changes drastically since thermal power units need to adjust the power generation capacity according to the grid load during operation. The thermal power units are constantly changing the working conditions, such as steady-transition-steady. This will result in a miss correlation between data. This effect can be minimized by data-resampling the data in an appropriate period. In this study, the datasets were re-sampled into 6-minute intervals. The actual load’s scatter plot, before and after re-sampling, is shown in Fig.3, Fig.4. The re-sampled data is smoother and more similar to regular operation, as is shown in Fig.4.

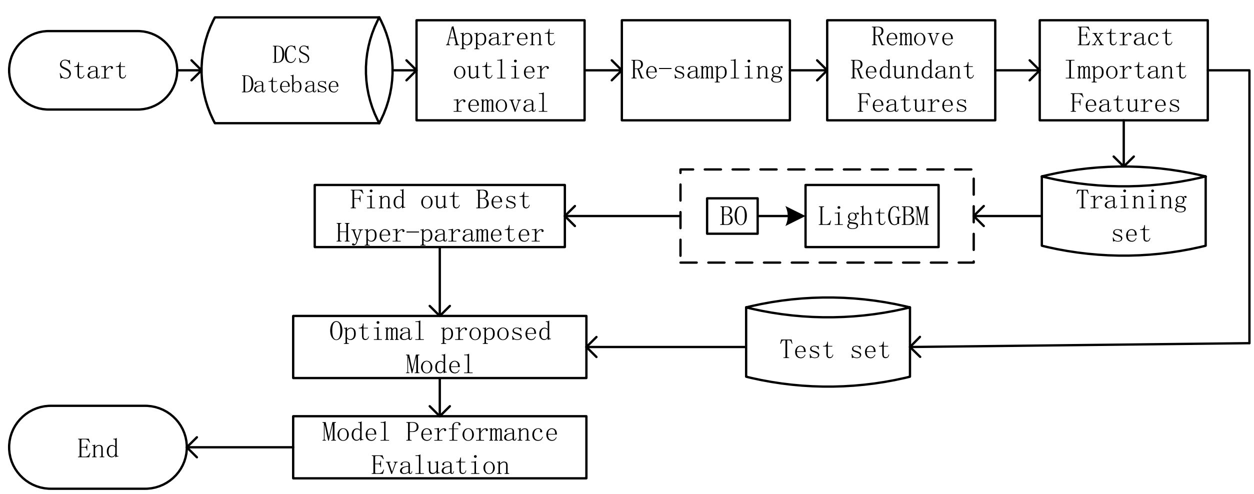


Fig. 1. Methodology Flowchart.

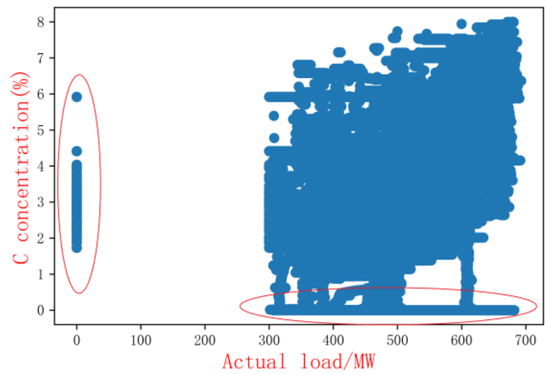


Fig. 2. Raw Load Data Scatter Plot.

B. Data Re-sampling

Feature engineering [28] is scoring each potential feature based on specific feature selection metrics and selecting representative variables from a given dataset to improve the final prediction. Feature engineering is crucial in the model design, as irrelevant or redundant data features will harm the model's performance. By reducing the number of variables, noisy and irrelevant data are removed, and the algorithm can run fast as the number of variables is reduced. There are generally three feature selection methods: filter method based on statistical information, wrapper method, and embedded method [29]. This study uses the correlation matrix(based on the filter method) and wrapper method to deal with variables.

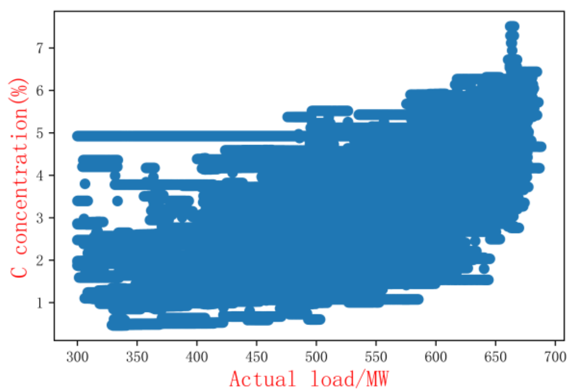


Fig. 3. Scatter Plot before Re-sampling.

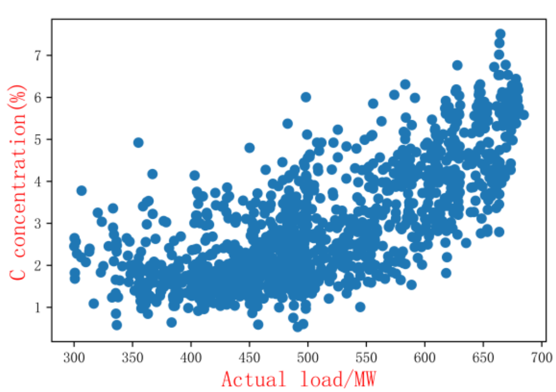


Fig. 4. Scatter Plot after Re-sampling.

According to the characteristics of multivariable variables, nonlinear and highly coupling thermal process, Firstly, the features with strong coupling are found through the correlation matrix, and the variables with low correlation with the carbon content of fly ash are eliminated. The essential variables are further extracted by the wrapper method to reduce the computational complexity of the model.

C. Remove Redundant Features

The correlation matrix (CM) is a table that is constructed to quantify the dependence between variables, as shown in equation (1), and the correlation coefficient indicates the positive or inverse relationship between the target variables [30]. The correlation matrix identifies and deletes redundant features in the dataset. Fig.5 shows one of the generated correlation matrices, which presents the correlation between the six features, The features from top to bottom are 'air temperature', 'air volume', 'steam temperature', 'steam temperature 2', 'air temperature 2', and 'air pressure'. If the correlation coefficient between the two variables is more significant than 0.95, they are compared with the carbon content in fly ash, and the variable with the smaller correlation coefficient is removed. For Fig.5, the features ' air temperature', 'air volume', and 'steam temperature ' were removed. In this way, the number of 71 DCS variables was reduced to 45.

$$r = \frac{\sum (x_i - x_{ave})(y_i - y_{ave})}{\sqrt{\sum (x_i - x_{ave})^2 \sum (y_i - y_{ave})^2}} \tag{1}$$

where r is the correlation coefficient, x_i is value of feature x, y_i is value of feature y, x_{ave} is mean value of the feature x, y_{ave} is mean value of the feature y.

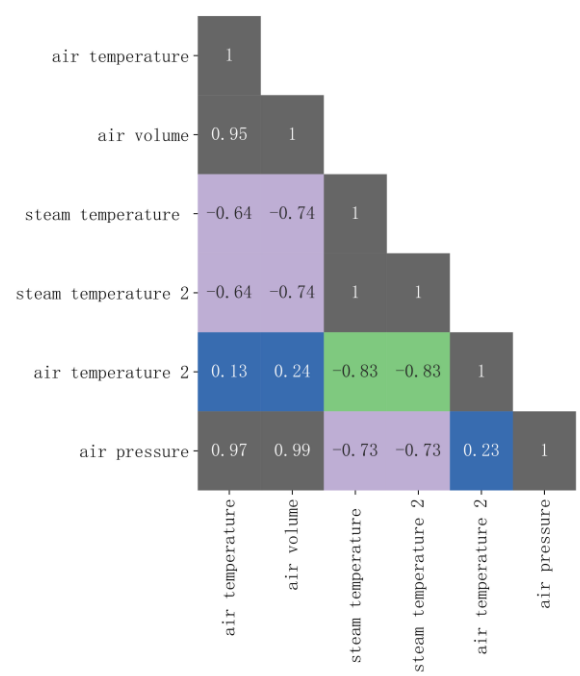


Fig. 5. Part of the Correlation Matrix.

D. Extract Important Features

The wrapper method is a feature selection method according to a specific prediction model, and this method uses recursive feature elimination (RFE). It is a greedy optimization algorithm that selects the best feature subset by repeated iteration. For the last step, the 45 variables selected by the correlation matrix are then used to determine the best performing 20 variables by the wrapper method.

E. Establishment of the Prediction Model

Before modeling, we will process features through correlation matrix and wrapper method, eliminate variables with highly coupling through correlation matrix, and select essential features subset by wrapper method, to reduce the computational complexity and further improve the expressiveness of the model.

LightGBM has many hyperparameters, and a reasonable choice of hyperparameters can improve prediction. While using the lighthGBM model to predict the carbon content of fly ash, the BO algorithm is used to continuously adjust the hyperparameters of the lighthGBM model to improve the prediction accuracy of the modelThe process of generating the optimal model BO_LightGBM is shown in Fig.6. During the model satisfaction assessment, cross-validation is set 5, the evaluation function is a root mean squared error.

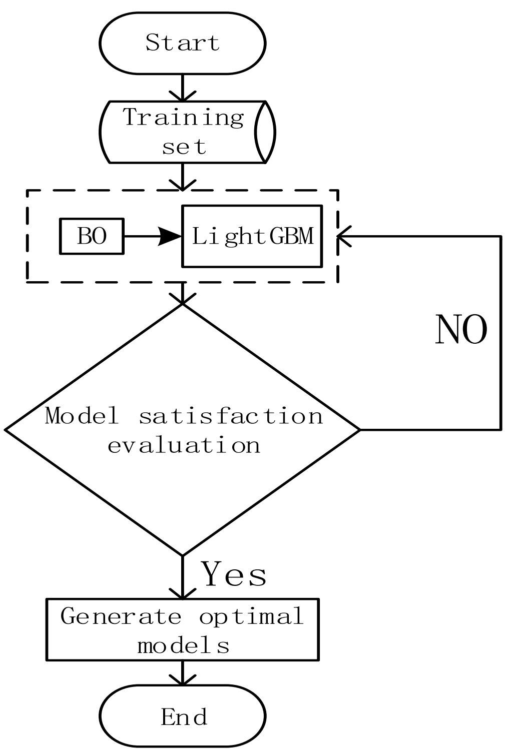


Fig. 6. Hyperparameter Tuning Flow Chart.

IV. Validation and Results

A. Performance Metrics

The regression evaluation indexes in regression analysis have mean squared error (MSE), root mean squared error (RMSE), mean absolute percentage error (MAPE), and coefficient of determination(R^2). $\mathrm{RMSE}=\sqrt{\mathrm{MSE}}$, In this study, RMSE, MAPE, and R^2 were selected as performance metrics, and these indices can be calculated as Eq. (3)(4)(5). R^2 is adapted to measure the approximation degree of the data to the prediction value, the closer R^2 is to 1, the better the fitting effect of the model. The smaller MSE, MAPE means the more accurate the prediction.

$$\mathrm{RMSE}=\sqrt{\frac{1}{\mathrm{m}}\sum_{i=1}^{m}(y_i-\hat{y}_i)^2} \tag{3}$$

$$\mathrm{MAPE}=\frac{1}{\mathrm{m}}\sum_{i=1}^{m}\left|\frac{y_i-\hat{y}_i}{\hat{y}_i}\right|\times 100\% \tag{4}$$

$$R^2=1-\frac{\sum_{i=1}^{m}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{m}(y_i-\bar{y}_i)^2} \tag{5}$$

where y_i are the actual values, $\hat{y}_i$ are the predicted values,and $\bar{y}_i$ is the mean of y_i (i=1,2,..., n).

B. Performance Comparison of Feature Processing Methods

Methods frequently used in feature processing are Random Forest (RF) [31] and Pearson correlation coefficient (PCC) [32]. The experiment compares the method proposed in this paper with RF and PCC. The optimal 20 features of the three methods are modeled for prediction while ensuring that the selected data are consistent with the boiler's steady-state operating conditions. In this process, the correlation matrix eliminates the features with high correlation, retains 45 features with low coupling, and then uses the packaging method to retain 20 features. The processed results of each of the three methods are used as input to the LightGBM model. The experimental results are listed in Table I.

The random forest method model is less effective, as shown in Table I. The model treated with PCC outperformed the RF. After using the correlation matrix to process the features, the R^2, MAPE, and RMSE were significantly optimized, and after further processing of features by wrapper method, the R^2 was improved to 0.822, MAPE was reduced to 16.5%, and RMSE was reduced to 0.509. It is proved that the fitting effect of the model using the method proposed in this paper is further enhanced, and the error is further reduced, which effectively solves the problem of high correlation and strong coupling between variables.

TABLE I. Prediction Results after Feature Processing

Method	**R^2**	**MAPE**	**RMSE**
RF	0.71	19.22%	0.644
PCC	0.78	17.47%	0.573
CM	0.814	16.69%	0.523
CM+Wrapper	0.822	16.50%	0.509

C. Comparison of Model Prediction Performance

To validate the superiority of the proposed method, three methods, including LM- Garson-BP [16] AQPSO-SVR [19] FPA-RF [12], the three latest methods are compared as benchmarks.

- LM-Garson-BP: The LM-Garson-BP methods used sensitivity analysis to select the feature variables and then used BP neural networks for predictive modeling and genetic algorithms to optimize the connection weights, number of neurons, and number of hidden layers.
- AQPSO-SVR: The AQPSO-SVR method first adds adaptive perturbation to the quantum particle swarm optimization (QPSO) algorithm and uses this improved algorithm to find the optimal hyper-parameters of the support vector regression (SVR).
- FPA-RF: The FPA-RF method first uses the random forest method to filter features, then uses the random forest as a prediction model and uses the flower pollination (FPA) algorithm to optimize the hyperparameters of the random forest.

MAPE, RMSE, and R^2 are selected as evaluation indexes. The experimental results are shown in Table II, and the prediction comparison results are listed in Fig. 7 and Fig. 8.

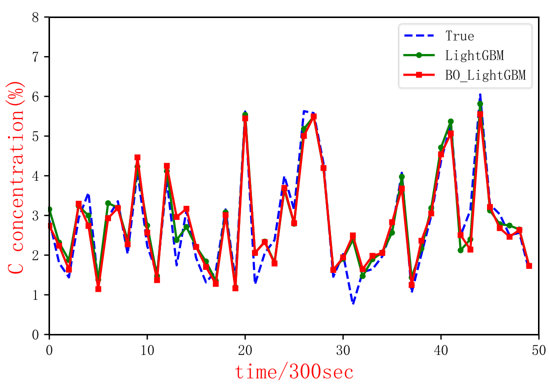


Fig. 7. Prediction Comparison of LightGBM Model and BO_LightGBM.

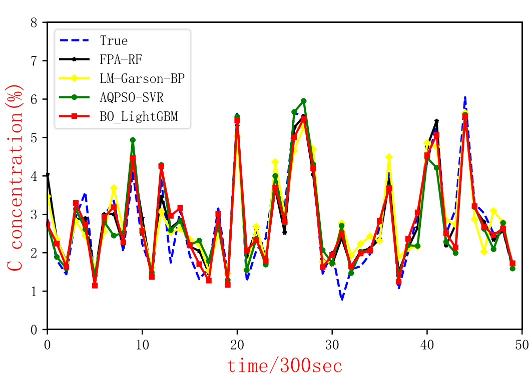


Fig. 8. Prediction Comparison of different Combined Models.

TABLE II. PREDICTION RESULTS OF DIFFERENT MODELS

Model	R^2	MAPE	RMSE
LM-Garson-BP [16]	0.722	19.69%	0.658
AQPSO-SVR [19]	0.786	18.22%	0.560
FPA-RF [12]	0.696	19.37%	0.652
LightGBM	0.822	16.50%	0.509
BO_LightGBM	0.831	16.02%	0.494

Table II shows the performance comparison between the proposed method and other methods. The obtained results are representative. The method proposed in this paper achieves the lowest MAPE, RMSE, and the highest R^2. The method in this paper reduces RMSE by 2.9%~24.9% and MAPE by 2.9%~18.6% compared with the above methods, indicating a further reduction of errors and improvement of measurement accuracy. The R^2 was improved by 1.1%~15.1%, indicating that the prediction curves were better fitted and the method in this paper was more accurate and reliable. Specifically, LM-Garson-BP, AQPSO-SVR, and FPA-RF all use heuristic algorithms for hyperparameter tuning and combine with regression models for prediction, which improves the prediction accuracy of the corresponding models to some extent.

From the perspective of hyper-parameter tuning, The BO algorithm can find the next evaluation position based on the information obtained for the unknown objective function when facing a complex optimization problem with hyperparameter tuning that is non-convex, multimodal, and computational, to reach the optimal solution the fastest [25]. The BO algorithm avoids the issues of ineffective use of iterative feedback information and the slow search speed of the algorithm. From the perspective of the prediction model, the LightGBM algorithm objective function adopts the second-order Taylor expansion, which can fully learn the model, add regular terms, reduce the complexity of the model, prevent overfitting, support parallel and distributed computing, and effectively improve the prediction accuracy. Therefore, the prediction results are better compared to the four models compared.

V. CONCLUSION

In this study, a data-driven approach integrating various machine learning algorithms and data mining techniques is used for the first time to analyze the relationship between the carbon content of fly ash and various operating parameters of boilers. This method has practical significance for guiding boiler production; collecting data for 37 days of complete working conditions and comparing our feature processing method with the PCC method, the RF method. The performance of the model is compared with LM-Garson-BP, AQPSP-SVR, and FPA-RF models. The results demonstrate that the method used in this paper exhibits better prediction results.

In future work, consider combining DCS data and coal type characteristics to improve accuracy by more data information. In addition, due to the high correlation between DCS data and time, it is worthwhile to study more deeply how to mine valuable information from these unstructured time series data and find the intrinsic correlation between the time series data.

ACKNOWLEDGMENT

This research is supported by National Natural Science Foundation of China (52176110); Scientific Research Project of Education Department of Hubei Province (D20191708); Intellectual Property Promotion Project of Colleges and Universities of Hubei Province (GXYS2018009).

REFERENCES

[1] Li R, Wei K, Huang Q. A novel method for precise measurement of unburnt carbon in boiler fly ash by ECSA® based on TG-MS[J]. Fuel, 2020, 264: 116849.

[2] Brown R C, Dykstra J. Systematic errors in the use of loss-on-ignition to measure unburned carbon in fly ash[J]. Fuel, 1995, 74(4): 570-574.

[3] Liu R, Deguchi Y, Nan W, et al. Unburned carbon measurement in fly ash using laser-induced breakdown spectroscopy with short nanosecond pulse width laser[J]. Advanced Powder Technology, 2019, 30(6): 1210-1218.

[4] Liu H, Tan H, Gao Q. Microwave attenuation characteristics of unburned carbon in fly ash[J]. Fuel, 2010, 89(11): 3352-3357.

[5] CHENG Qi-ming, HU Xiao-qing, et al.Summary of Measurement Methods of Carbon Content in Fly Ash[J]. Journal of Shanghai University of Electric Power, 2011, 27(5): 519-524.

[6] H.Y. Liu, H.Z. Tan, Q.A. Gao, et al. Microwave attenuation characteristics of unburned carbon in fly ash, Fuel 89 (2010) 3352–3357.

[7] Xiao, Jianli. "SVM and KNN ensemble learning for traffic incident detection." Physica A: Statistical Mechanics and its Applications 517 (2019): 29-35.

[8] Development of soft sensors for isomerization process based on support vector machine regression and dynamic polynomial models[J]. Journal of Robotics & Machine Learning,2019, 149:95-103.

[9] Liu R, Chen P, Wang Z. Quantitative analysis of carbon content in fly ash using LIBS based on support vector regression[J]. Advanced Powder Technology, 2021, 32(8): 2978-2987.

[10] Amin S M. Smart grid: Overview, issues and opportunities. advances and challenges in sensing, modeling, simulation, optimization and control[J]. European Journal of Control, 2011, 17(5-6): 547-567.

[11] Yan W, Tang D, Lin Y. A data-driven soft sensor modeling method based on deep learning and its application[J]. IEEE Transactions on Industrial Electronics, 2016, 64(5): 4237-4245.

[12] WANG Fang1, MA Suxia1, WANG He. Prediction model of carbon content in fly ash using random forest variable selection method[J]. Thermal Power Generation, 2018, 47(11): 89-95.

[13] ZHOU Hao,ZHU Hong-bo, et al. Artificial neural network modelling on the unburned carbon in fly ash from utility boilers[J].Proceedings of the CSEE,2002(06):97-101.

[14] WANG Chun-lin, Zhou Hao, et al. Support vector machine modeling on the unburned carbon in fly ash[J]. Proceedings of the CSEE,2005(20):72-76.

[15] Zhang P. A novel feature selection method based on global sensitivity analysis with application in machine learning-based prediction model[J]. Applied Soft Computing, 2019, 85: 105859.

[16] Zhu Jinqi,Niu Xiaofan,Xiao Xianbin. Prediction models of the carbon content of fly ash in a biomass boiler based on improved BP neural networks [J]. Renewable Energy Resources, 2020,38(02):150-157(in Chinese).

[17] Wen X. Modeling and performance evaluation of wind turbine based on ant colony optimization-extreme learning machine[J]. Applied Soft Computing, 2020, 94: 106476.

[18] Feng Xugang, Qian Jiajun, Zhang Jiayan. Prediction method of unburned carbon content in fly ash based on genetic neural network with sensitivity analysis [J].Journal of electronic measurement and instrumentation,2016,30(7): 1083-1089(in Chinese).

[19] PENG Dao-gang, LI Dan-yang, et al.Research on Prediction of Carbon Content in Boiler Fly Ash Based on ADQPSO－SVR[J]. Computer Simulation,2020,37(03):72-77(in Chinese).

[20] Ke G, Meng Q, Finley T, et al. Lightgbm: A highly efficient gradient boosting decision tree[J]. Advances in neural information processing systems, 2017, 30: 3146-3154.

[21] J. Friedman. Greedy function approximation: a gradient boosting machine. Annals of Statistics, 29(5):1189–1232, 2001.

[22] Chen C, Zhang Q, Ma Q, et al. LightGBM-PPI: Predicting protein-protein interactions through LightGBM with multi-information fusion[J]. chemometrics and intelligent laboratory systems, 2019, 191: 54-64.

[23] Ju Y, Sun G, Chen Q, et al. A model combining convolutional neural network and LightGBM algorithm for ultra-short-term wind power forecasting[J]. Ieee Access, 2019, 7: 28309-28318.

[24] Probst P, Boulesteix A L, Bischl B. Tunability: importance of hyperparameters of machine learning algorithms[J]. The Journal of Machine Learning Research, 2019, 20(1): 1934-1965.

[25] Cui JX, Yang B. Survey on Bayesian Optimization Methodology and Applications. Journal of Software, 2018, 29(10): 3068-3090(in Chinese).

[26] Wu J, Chen X Y, Zhang H, et al. Hyperparameter optimization for machine learning models based on Bayesian optimization[J].Journal of Electronic Science and Technology, 2019, 17(1): 26-40.

[27] Turner R, Eriksson D, McCourt M, et al. Bayesian optimization is superior to random search for machine learning hyperparameter tuning: Analysis of the black-box optimization challenge 2020[J]. arXiv preprint arXiv:2104. 10201, 2021.

[28] Tsai C F. Feature selection in bankruptcy prediction[J]. Knowledge-Based Systems, 2009, 22(2): 120-127.

[29] Yvan S , I Iñaki, Pedro L . A review of feature selection techniques in bioinformatics[J]. Bioinformatics, 2007(19):2507-2517.

[30] Asuero A G, Sayago A, Gonzalez A G. The correlation coefficient: An overview[J]. Critical reviews in analytical chemistry, 2006, 36(1): 41-59.

[31] Breiman L. Random forests[J]. Machine learning, 2001, 45(1): 5-32.

[32] Benesty J, Chen J, Huang Y, et al. Pearson correlation coefficient[M]//Noise reduction in speech processing. Springer, Berlin, Heidelberg, 2009: 1-4.

Research on Intelligent Natural Language Texts Classification

Chen Xiao Yu[1]
Computer School
Beijing Information Science & Technology University
Beijing, China

Zhang Xiao Min[2]
Academy of Agricultural Planning and Engineering
Ministry of Agriculture and Rural Affairs
Beijing, China

***Abstract*—Natural language texts widely exist in many aspects of social life, and classification is of great significance to its efficient use and normalized preservation. Manual texts classification has the problems such as labor intensive, experience dependent and error prone, therefore, the research on intelligent classification of natural language texts has great social value. In recent years, machine learning technology has developed rapidly, and related researchers have carried out a lot of works on the texts classification based on machine learning, the research methods show the characteristic of diversification. This paper summarizes and compares the texts classification methods mainly from three aspects, including technical routes, text vectorization methods and classification information processing methods, in order to provide references for further research and explore the development direction of the texts classification.**

Keywords—Machine learning; natural language texts; text vectorization; classification information processing

I. INTRODUCTION

Intelligent classification of natural language texts has many application scenarios in social life, and related research has important social value. In recent years, the research of natural language texts classification based on machine learning has received a lot of attention. Researchers tried to develop corresponding intelligent texts classification methods for different types of natural language texts, and obtained many valuable research results. The research objects involved the fields of government affairs, justice, energy and electricity, transportation, medical care and health, agriculture, science and technology, intellectual property, business, etc. The factors affecting the applicability of texts classification methods mainly include two points; firstly, the differences in the characteristics of different types of texts may affect the applicability of classification methods; secondly, the possible limitations of the training data used in model training might limit its applicability to other data of the same type.

From the perspective of technical route, the appropriate technical routes for different texts classifications are usually different due to the differences in text types and text samples. In general, the technical routes usually involve data collection, data preprocessing, text vectorization, classification information processing, classification and other links. In relevant research reports, through comparative analysis, the technical routes which were finally adopted and achieved good results were different in the links and the basic technical methods involved. For example, for Chinese texts, it is usually necessary to add word segmentation link before text vectorization; in some classification scenarios, the classification effect could be optimized by adding feature selection or feature extraction link to reduce dimensionality before text vectorization; in the classification information processing, there are also widely differences in the appropriate information processing levels for different types of text data.

From the perspective of specific technical method, the applicability of technical method to text types has the characteristics of difference and diversity. A method may have different applicability when targeting different types of texts. Zhu F. P. et al. has achieved good classification results in the classification of news texts in the shipbuilding industry using the method based on SVM [1]; Zhao M. et al. compared the classification effects of LSTM, SVM, and CNN methods in the classification of dietary health texts, the results showed that LSTM had the best effect in the corresponding scenario, which was better than SVM [2]. The same type of texts may be classified effectively based on different types of methods, though there might be some differences in the classification accuracy. For example, Bao X. et al. compared the classification effects of the multi-instance learning method, SVM and KNN in the classification of patent texts, and the results showed that the multi-instance learning method could achieve good classification results [3]; Wen C. D, et al. used the BIGRU method in the classification of patent texts, and also achieved good results [4].

Natural language texts classification methods usually include two core parts, one is the text vectorization method, and the other is the classification information processing method. Text vectorization refers to the representation of natural language text in the form of vectors for further data processing, the technical methods used in related research include word frequency-based methods, distributed static word vector-based methods, and distributed dynamic word vector-based methods and so on. The technical methods based on word frequency express text in the form of vector mainly by counting the frequency of word appearing in the text, the principle is relatively simple and the implementation is relatively convenient, but the context-related information could not be preserved in the text vectorization; the distributed word vector methods, which understand and represent words based on context information, could effectively retain context-related information in text vectorization; the distributed dynamic word vector methods are different from the static methods, the main

Funding Project: Beijing Information Science & Technology University-Computer School- Scientific Research Business Work Funds (5029923412).

difference is that the distributed dynamic word vector methods could distinguish the polysemy of a word in different contexts. Classification information processing is the process of further processing for the information in the vectors to obtain more accurate classification prediction information after texts are represented in the form of vectors. The technical methods involved in current related research include classical machine learning methods such as SVM, NB, the methods based on signal type of neural network, the methods based on vertical combination of multiple types of neural networks, the methods combining attention mechanism, the methods based on ensemble learning strategies, etc. In general, the methods used in current natural language texts classification research usually involve the combination of a variety of basic technologies, and with the continuous development of the basic technology and the breadth and depth of the texts classification research, the basic technologies and the combination strategies are becoming more and more diverse.

This paper analyzes and compares the research methods of natural language texts classification based on machine learning from three aspects: technical routes, text vectorization methods, and classification information processing methods. On the one hand, it is expected to provide reference support for subsequent related research and application, and on the other hand, it is expected to explore the technical development direction of this field based on the analysis of high-level research reports in recent years.

II. TECHNICAL ROUTE OF INTELLIGENT NATURAL LANGUAGE TEXTS CLASSIFICATION

The links involved in natural language texts classification include data acquisition, data preprocessing, splitting into words (for Chinese), text vectorization, feature selection, feature extraction and dimensionality reduction, classification information processing and classification. Among them, data preprocessing, splitting into words (for Chinese), text vectorization, classification information processing and classification are usually necessary links.

Natural language texts classification modeling involves multi-link collaboration. Data acquisition methods mainly include crawling data from internet, internal data and public data sets; data preprocessing mainly includes data desensitization, deduplication, removing invalid data, removing incomplete data, completing incomplete data, etc.; the link of splitting into words is usually based on jieba, and if it is combined with a thesaurus based on professional fields, better result might could be achieved; the methods commonly used in text vectorization mainly include BERT, word2vec, and TD-IDF; feature selection can reduce the vector dimension, the methods commonly used include chi-square test, removal of low-frequency words, etc.; the use of feature extraction and dimensionality reduction may improve the classification effect, related methods include PCA, etc.; classification information processing is used to find out the connection between the text content and the classification to which it belongs, related methods mainly include neural-based methods, ensemble learning-based methods in which, different types of technologies could be used in combination; in order to obtain classification results, the softmax technical method was commonly used.

The differences and diversity exist in the technical routes of the machine learning texts classification for different types of texts or different datasets. We summarize a technical route based on the commonality summary and difference supplement, which covers the main links involved in the texts classification, and in a specific study, the technical route could be used as a basis for further applicability deletion or adjustment, the technical route is shown in Fig. 1.

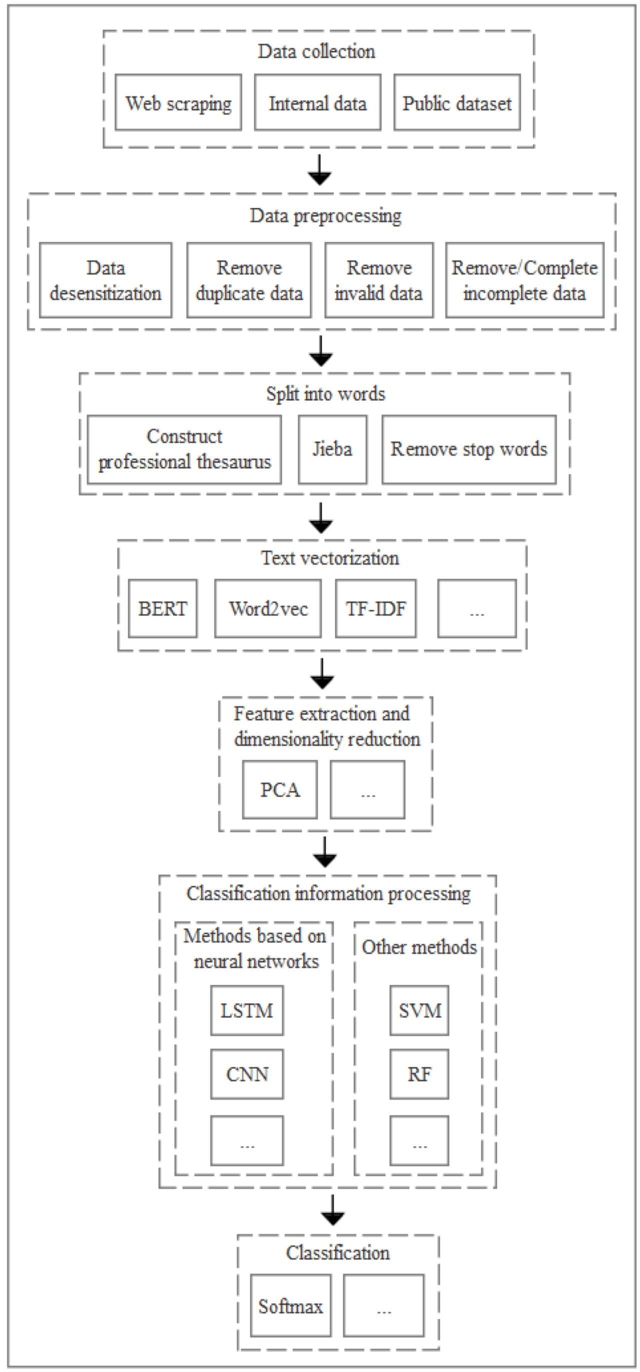


Fig. 1. The Technical Route of Texts Classification Research.

III. Text Vectorization Methods

Text vectorization is one of the core links of natural language texts classification. After preprocessing, text data usually needs to be expressed in vector form firstly, and then classification information processing would be operated. Text vectorization methods could be mainly divided into four categories: methods based on distributed dynamic word vectors, methods based on distributed static word vectors, methods based on topic and methods based on word frequency, as shown in Table I.

The technical methods based on word frequency include one-hot, TF-IDF, etc. One-hot is the simplest text vectorization technical method, whose basic principle is to use an N-dimensional vector to represent text based on the size of thesaurus, the N is the number of the words in the thesaurus, for a piece of text data, each dimension value of the N-dimensional vector corresponds to a word in the thesaurus, and the value range of the dimension variables is 0 or 1, if the word corresponding to a dimension variable appears in a piece of text data, its corresponding value would be 1, and if it does not appear, the corresponding value would be 0; one-hot is simple in principle and convenient in implementation, which also has many shortcomings, for example, when the scale of thesaurus is relatively large, the dimension of the vector will also be expanded accordingly, which is not conducive for data processing and classification. One-hot could be understood as the simplest word frequency-based text vectorization technical method, furthermore, TF-IDF is a kind of some more complex text vectorization technical method based on word frequency, compared to other technical methods, TF-IDF still has the characteristics of simple principle and convenient implementation, which is widely used in current text vectorization. Text vectorization methods based on word frequency could often achieve good application results in texts classification, which also have some important disadvantages, including: (1) the extraction of text features would ignore contextual information, (2) the position information where the word appears would be lost in text vectorization, etc.

The text vectorization methods based on distributed word vector have received extensive attention in recent years. The basic principle of distributed word vector method is to understand and represent word based on the context, therefore, compared with the methods based on word frequency, it could effectively solve the problem of the loss of contextual information in the vectorized representation of text. The context-based text vectorization methods could usually more effectively extract the classification information from text data, thereby improving the classification effect, and in different methods of this type, there are also differences in the scope of the context used, based on which the technical methods could be divided into partial text information-based methods and all text information-based methods; relatively, all text information-based methods usually have the advantages in relational information extraction.

Distributed word vector methods include distributed dynamic word vector methods and distributed static word vector methods. Distributed static word vector methods understand and represent words based on context, however, it cannot solve the problem of polysemy, that is, the same word may have different meanings in different contexts, in the text vectorization based on static word vector, a word could only have one representation, so it is impossible to distinguish the different interpretations of a word in different contexts. The main difference between the distributed dynamic word vector methods and the distributed static word vector methods is that the dynamic methods could distinguish the different interpretations of a word.

Topic-based text vectorization methods are another type of commonly used methods in natural language texts classification, including LDA (latent Dirichlet allocation), improved LDA, etc. The basic principle of LDA-based texts classification methods is to firstly extract topic information based on the texts, and then implement texts classification based on the topic information.

TABLE I. TEXT VECTORIZATION METHODS

No.	Category	Method	Using Examples	References
1	Distributed dynamic word vectors	BERT (bidirectional encoder representations from transformers)	Social e-commerce text, power grid equipment defect text	[5], [6]
		ALBERT	Patent text (ALBERT VS Word2vec, global vectors for word representation)	[4]
2	Distributed static word vectors	Word2vec	Railway signal equipment breakdown short text, healthy diet text (Word2vec VS term frequency-inverse document frequency, bag-of-words model), rice knowledge text (Word2vec VS one-hot, term frequency-inverse document frequency, hashing), ultra-short commodity text, news text	[7], [2], [8], [9], [10]
		CLW2V (character level Word2Vec)	Railway text (CLW2V VS term frequency-inverse document frequency, Word2vec)	[11]
		Fasttext	Coal mine accident case text	[12]
3	Topic	LDA (latent Dirichlet allocation)	Patent text	[13]
		MULCHI-labeled LDA (improved LDA)	Science and technology video text	[14]
4	Word frequency	Bag-of-words model	Hypertension medical record text	[15]
		TF-IDF (term frequency-inverse document frequency)	People's congress report text, cultural tourism text	[16], [17]

In summary, the methods used in text vectorization have developed from simpler methods such as ont-hot and word frequency-based methods, to the methods based on static word vectors, and then to the methods based on dynamic word vectors. Zhao M. et al. compared the text vectorization methods of word2vec, TD-IDF, and bag-of-words in their research on dietary health texts classification, and the research results showed that word2vec had the best effect in corresponding scenario [2]; Wen Ch. D. et al. compared ALBERT, word2vec, and glove text vectorization methods in their research on patent texts classification, and the results showed that the ALBERT method had the best effect [4].

Text vectorization is the basis of natural language texts classification, with the development of related technologies and the extensive and in-depth development of related research, the development of the text vectorization methods could be used in texts classification currently presents two trends, firstly, new technical methods emerge in an endless stream, the new technical methods could usually better retain classification-related information in text vectorization transformation, thereby providing a good foundation for accurate classification; secondly, in the classification of different types of texts, the applicable text vectorization methods reflect the diversified development trend, the emergence of new technical methods has enriched the options, and different technical methods have different characteristics and applicability, increasing the diversity of text vectorization technical methods.

IV. Classification Information Processing Method

Classification information processing is another core link of natural language texts classification. After the text data is represented in the form of vector, the including classification information needs to be extracted or processed to realize texts classification. The methods mainly include neural network methods, attention machine combined methods, ensemble learning methods, active learning methods and other methods, as shown in Table II.

TABLE II. Classification Information Processing Methods

No.	Category	Method	Using Examples	References
1	Neural network	LSTM (long-short term memory network)	Eating healthy text (LSTM VS SVM, CNN)	[2]
		BILSTM (bidirectional LSTM)	Electricity audit text	[18]
		CNN (convolutional neural network)	Electrical equipment defect text	[19]
		MCNN (multi-pooling convolutional neural network)	Short railway signal equipment failure text	[7]
		CCNN (combined CNN)	News text	[10]
		CNN-SVM (CNN-support vector machines)	Nursing adverse event text	[20]
		CRNN (improved convolutional recurrent neural network)	Police text	[21]
		MNN (multilayer neural network)	Government website mailbox text (MNN VS naive Bayes, random forest, decision tree)	[22]
		BIGRU (bidirectional gated recurrent unit)	Patent text	[4]
		Method based on DCNN (deep CNN)	Rice knowledge text (DCNN VS BILSTM, attention-BIGRU, RCNN, DPCNN)	[8]
2	Attention machine	BILSTM-attention	Power grid equipment failure text, railway traffic accident text, commodity text	[23], [24], [25]
		CNN-NLSTM-attention (method based on CNN, nested long-short term memory network and attention mechanism)	News text	[26]
		LS-GRU (an improved GRU deep learning framework)	Medical text (LS-GRU VS BIGRU, LSTM, GRU)	[27]
		ERNIE (enhanced representation from knowledge integration)	People's congress report text	[16]
3	Active learning	SVD-CNN combined with improved active learning (CNN based on singular value decomposition algorithm combined with improved active learning)	Barrage text	[28]
4	Ensemble learning	Beam search ensemble	Short medical text	[29]
5	Other methods	NB (naive Bayes)	Cultural tourism text, agricultural text	[17], [30]
		SVM (support vector machines)	Hypertension medical record text, ship industry news text	[15], [1]
		Deep random forest	Super short commodity text (deep random forest VS k-nearest-neighbor, decision tree)	[9]
		Method based on multi-instance learning framework	Patent text (multi-instance learning VS support vector machines, k-nearest-neighbor)	[3]

Neural networks are widely used in the classification information processing in natural language texts classification, including classical neural networks, improved neural networks for specific classification problems and the combined neural networks formed by stacking different types of neural networks. There are two main advantages of the neural network methods applied to the classification information processing. On one hand, the neural network methods could usually extract the deep internal correlation information between text content and category with high quality through black-box information processing, compared with the traditional methods such as NB and SVM, neural network methods could commonly analyze classification information deeply, and thus achieve better texts classification result. On the other hand, the neural network methods have the advantage of being more flexible, in dealing with practical classification problems, the applicability could take classic neural network models (such as CNN, RNN, etc.) as basis to make suitability adjustment, and better classification results could be achieved by constructing improved neural network models to better adapt to text characteristics; different types of neural networks could also be combined vertically, using the combined neural network for texts classification information processing could realize multi-level classification information processing and possibly achieve more effective classification information extraction; neural networks could be well combined with attention mechanism to improve the processing effect for classification information, this kind of methods have attracted wide attention in related research in recent years, which could often effectively improve the classification effect of natural language texts, the combination with attention mechanism further increases the flexibility and applicability of neural network methods. In general, the advantages of deep extraction of classification information and flexibility make the neural network methods the most usable basic methods in the classification information processing of natural language texts.

Except the neural network methods, classical machine learning methods such as RF and SVM still play an important role in natural language texts classification information processing, the advantages of traditional machine learning methods are mainly that they are more convenient to implement and the principles are relatively simple, at the same time, although different types of methods have general differences in information extraction capabilities, the diversity of natural language text characteristics makes traditional machine learning methods could also be more suitable for some specific classification scenarios, which could also possibly obtain better classification results.

Ensemble learning method is an important kind of methods suitable for the classification information processing of natural language texts, the basic principle of ensemble learning is to use different basic technologies to independently train multiple models, and combine the outputs of multiple independent models through a certain strategy to obtain the final classification output. The basic technical methods used in ensemble learning could be traditional machine learning methods such as NB, or relatively complex neural network methods such as BILSTM; ensemble learning could comprehensively apply the advantages and applicability of various types of machine learning technologies to improve the effect of texts classification. In the classification scene with diverse text features, based on ensemble learning strategy, to select different types of machine learning technologies reasonably as the basic methods could improve the applicability of the overall model and achieve high-quality texts classification result.

In summary, the methods used in the classification information processing have developed from basic methods such as NB, SVM, etc., to the methods based on neural networks, and then to attention combined methods, etc. Zhao M. et al. compared LSTM, SVM methods in the classification of diet and health texts, the results showed that LSTM has the best effect in corresponding scene [2]; Liu Y. et al. combined convolutional neural network (CNN), nested long short-term memory network (NLSTM) and attention mechanism in the study of news texts classification, and achieved good results using the CNN-NLSTM-Attention method [26]. In addition, the ensemble learning method could effectively combine the advantages of multiple classification models to improve the classification accuracy, which has received more attention in recent years [29]. Relevant theories and methods are becoming more and more abundant, further research and application could start from the texts characteristics to refer to research reports of the same type of texts for method selection or the basis for optimization, so as to efficiently and accurately implement texts classification modeling.

V. Text Types

Natural language texts classification commonly involves multiple links, in each link, the relevant basic technical methods have different applicability in different classification problems, and the main reason is the diversification of natural language text features; in texts classification research, the application could take the text types as a basis, and refer to existing related research reports to select technical methods or build a basis for the further improvement.

In order to provide convenient reference for subsequent relevant researchers, we have sorted up the research reports cited in this paper according to the text types of their research objects, which involves the fields of energy and environment, justice, government affairs, transportation, medical care and health, agriculture, science and technology, commerce and so on, as shown in Table III. This paper conduct analysis and research mainly from the perspectives of the technical routes of natural language texts classification and the main technical methods suitable for the core links, which does not comprehensively cover the text types already involved in natural language texts classification research, it is expected that subsequent researchers could make quick reference based on the text types table organized in this article or take it as a basis for further improvement.

TABLE III. TEXT TYPES OF NATURAL LANGUAGE TEXTS CLASSIFICATION

Field	Text type	References
energy and electricity	coal mine accident case text	[12]
	power information communication customer service system text	[31]
	power grid equipment text	[6], [23]
	electricity audit text	[18]
	power equipment defect text	[19]
transportation	railroad accident text	[24]
	railway text	[7]
	ship industry news text	[1]
medical and health	image report text	[27]
	medical short text	[29]
	nursing adverse event text	[20]
	hypertension medical record text	[15]
	eating healthy text	[2]
agriculture	rice knowledge text	[8]
	agricultural text	[30]
government affairs	people's congress report	[16]
	government website mailbox text	[23]
judicial	judgment document	[32]
	police text	[21]
Intellectual property	patent text	[4], [3], [13]
science & technology	academic paper	[33]
	science and technology video text	[14]
business	social e-commerce text	[5]
	super short commodity text	[9]
	commodity text	[25]
society	news	[10]
		[26]
		[34]
other fields	digital library text	[35]
	barrage text	[28]
	travel text	[17]

VI. CONCLUSION AND OUTLOOK

In general, the research on natural language texts classification received a lot of attention in recent years, and the research reports involved many important aspects of social life, including intellectual property, government affairs, justice, energy and electricity, transportation, medical care and health, agriculture, science and technology, commerce and so on, which constructed a favorable foundation for further research and applications. In terms of texts classification methods, the methods show diversification, and different methods commonly have different characteristics and applicability differences, which are closely related to text characteristics; follow-up research could take text characteristics as the basis to select and further improve classification methods.

The core links in natural language texts classification include text vectorization and classification information processing, in recent years, the theories and technologies in these both two aspects have made great progress; the text vectorization methods have developed from the basic methods such as the methods based on word frequency to the methods based on distributed static word vectors, and then to the methods based on distributed dynamic word vectors; the classification information processing methods have developed from the methods based on a signal technology such as SVM, CNN or RNN to the methods based on vertical integration of multiple technologies and the methods based on ensemble learning. The emergence of new theories and technologies provide more options for subsequent texts classification research, and also constructed more advantageous basis for further development of related theories and technologies; and at the same time, due to the diversification of text characteristics, the development trend of texts classification methods could still be diverse development.

In terms of similar studies comparison, the text classification modeling based on machine learning has received extensive attention in recent years, and at the same time, researchers have carried out some review and summary works. The relevant researches mainly summarized and analyzed from the aspects of research progress and development trends, and there are few systematic analyses for the links involved in text classification. This paper systematically analyzed the links involved in natural language text classification research, and summarized main research methods based on the core links. It's expected that this paper could efficiently provide reference for subsequent related research.

REFERENCES

[1] Zhu Fang Peng, Wang Xiao Feng, Text classification for ship industry news [J], Journal of Electronic Measurement and Instrumentation, 2020, 34 (01): 149-155.

[2] Zhao Ming, Du Hui Fang, Dong Cui Cui, Chen Chang Song, Diet health text classification based on word2vec and LSTM [J], Transactions of the Chinese Society for Agricultural Machinery, 2017, 48 (10): 202-208.

[3] Bao Xiang, Liu Gui Feng, Yang Guo Li, Patent text classification method based on multi-instance Learning [J], Information Studies: Theory & Application, 2018, 41 (11): 144-148.

[4] Wen Chao Dong, Zeng Cheng, Ren Jun Wei, Zhang Yan, Patent text classification based on ALBERT and bidirectional gated recurrent unit [J], Journal of Computer Applications, 2021, 41 (02): 407-412.

[5] Li Ke Yue, Chen Yi, Niu Shao Zhang, Social E-commerce text classification algorithm based on BERT [J], Computer Science, 2021, 48 (02): 87-92.

[6] Tian Yuan, Yuan Ye, Liu Hai Bin, Man Zhi Bo, Mao Cun Li, BERT pre-trained language model for defective text classification of power grid equipment [J]. Journal of Nanjing University of Science and Technology, 2020, 44 (04): 446-453.

[7] Zhou Qing Hua, Li Xiao Li, Research on short text classification method of railway signal equipment fault based on MCNN [J]. Journal of Railway Science and Engineering, 2019, 16 (11): 2859-2865.

[8] Feng Shuai, Xu Tong Yu, Zhou Yun Cheng, Zhao Dong Xue, Jin Ning, et al. Rice knowledge text classification based on deep convolution neural network [J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52 (03): 257-264.

[9] Niu Zhen Dong, Shi Peng Fei, Zhu Yi Fan, Zhang Si Fan, Research on classification of commodity ultra-short text based on deep random forest [J]. Transactions of Beijing Institute of Technology, 2021, 41 (12): 1277-1285.

[10] Zhang Yu, Liu Kai Feng, Zhang Quan Xin, Wang Yan Ge, Gao Kai Long, A combined-convolutional neural network for Chinese news text classification [J]. Acta Electronica Sinica, 2021, 49 (06): 1059-1067.

[11] Lu Bo Ren, Hu Shi Zhe, Lou Zheng Zheng, Ye Yang Dong, Character-level feature extraction method for railway text classification [J]. Computer Science, 2021, 48 (03): 220-226.

[12] Yan Yan, Yang Meng, Zhou Fa Guo, Ge Yi Fan, Comparison of text classification methods of coal mine accident cases based on Fasttext network [J]. Coal Engineering, 2021, 53 (11): 186-192.

[13] Liao Lie Fa, Le Fu Gang, Zhu Ya Lan, The Application of LDA Model in Patent Text Classification [J]. Journal of Modern Information, 2017, 37 (03): 35-39.

[14] Ma Jian Hong, Fan Yue Xiang, Science and technology video text classification based on improved labeled LDA model [J]. Computer Engineering, 2018, 44 (09): 274-279.

[15] Hu Jing, Liu Wei, Ma Kai, Text categorization of hypertension medical records based on machine learning [J]. Science Technology and Engineering, 2019, 19 (33): 296-301.

[16] Yu Hang, Li Hong Lian, Lü Xue Qiang, Text classification of NPC report contents [J], Computer Engineering and Design, 2021, 42 (06): 1772-1778.

[17] Wang Xiang Xiang, Fang Hui, Chen Chong Cheng, Classification technique of cultural tourism text based on naive Bayes [J]. Journal of Fuzhou University (Natural Science Edition), 2018, 46 (05): 644-649.

[18] Chen Ping, Kuang Yao, Hu Jing Yi, Wang Xiang yang, Cai Jing. Text categorization method with enhanced domain features in power audit field [J]. Journal of Computer Applications, 2020, 40 (S1): 109-112.

[19] Liu Zi Quan, Wang Hui Fang, Cao Jing, Qiu Jian, A classification model of power equipment defect texts based on convolutional neural network [J]. Power System Technology, 2018, 42 (02): 644-651.

[20] Ge Xiao Wei, Li Kai Xia, Chen Ming, Text classification of nursing adverse events based on CNN-SVM [J]. Computer Engineering & Science, 2020, 42 (01): 161-166.

[21] Wang Meng Xuan, Zhang Sheng, Wang Yue, Lei Ting, Du Wen, Research and application of improved CRNN model in classification of alarm texts [J]. Journal of Applied Sciences, 2020, 38 (03): 388-400.

[22] Wang Si Di, Hu Guang Wei, Yang Si Yu, Shi Yun, Automatic transferring government website e-mails based on text classification [J]. Data Analysis and Knowledge Discovery, 2020, 4 (06): 51-59.

[23] Tian Yuan, Ma Wen, Attention-BiLSTM-based fault text classification for power grid equipment [J]. Journal of Computer Applications, 2020, 40 (S2): 24-29.

[24] Han Guang, Bu Tong, Wang Ming Ming, Zheng Hai Qing, Sun Xiao Yun, et al. Text classification of railway traffic accidents based on dual-channel bidirectional long short term memory network [J]. Journal of the China Railway Society, 2021, 43 (09): 71-79.

[25] He Bo, Ma Jing, Li chi, Research on commodity text classification based on fusion features [J]. Information Studies: Theory & Application, 2020, 43 (11): 162-168.

[26] Liu Yue, Zhai Dong Hai, Ren Qing Ning. News text classification based on CNLSTM model with attention mechanism [J]. Computer Engineering, 2019, 45 (07): 303-308+314.

[27] Li Qiang, Li Yao Kun, Xia Shu Yue, Kang Yan. An improved medical text classification model: LS-GRU [J]. Journal of Northeastern University (Natural Science), 2020, 41 (07): 938-942+961.

[28] Qiu Ning Jia, Cong Lin, Zhou Si Cheng, Wang Peng, Li Yan Fang. SVD-CNN barrage text classification algorithm combined with improved active learning [J]. Journal of Computer Applications, 2019, 39 (03): 644-650.

[29] Zhang Bo, Sun Yi, Li Meng Ying, Zheng Fu Qi, Zhang Yi Jia, et al. Medical text classification based on transfer learning and deep learning [J]. Journal of Shanxi University (Natural Science Edition), 2020, 43 (04): 947-954.

[30] Zhao Yan, Li Xiao Hui, Zhou Yun Cheng, Zhang Yue. A study on agricultural text classification method based on naive bayesian [J]. Water Saving Irrigation,2018(02):98-102.

[31] Yu Xue Hao, Zhao Zi Yan, Ma Ying Long, Zheng Rong Rong, Xi Zi Yue, et al. Multi-label text classification for power ICT custom service system based on binary relevance and gradient boosting decision tree [J]. Automation of Electric Power Systems, 2021, 45 (11): 144-151.

[32] Weng Yang, Gu Song Yuan, Li Jing, Wang Feng, Li Jun Liang, Li Xin, Paragraph context-based text classification approach for large-scale judgment text structuring [J]. Journal of Tianjin University (Science and Technology), 2021, 54 (04): 418-425.

[33] Xue Feng, Hu Yue, Xia Shuai, Xu Jian Dong, Research on short text classification based on paper title and abstract [J]. Journal of Hefei University of Technology (Natural Science), 2018, 41 (10): 1343-1349.

[34] Hu Yu Lan, Zhao Qing Shan, Chen Li, Niu Yong Jie, A Fusion network for Chinese news text classification [J]. Journal of Chinese Information Processing, 2021, 35 (03): 107-114.

[35] Xu Tong Yang, Yin Kai. Text classification of digital library based on deep learning [J]. Information Science, 2019, 37 (10): 13-19.

Image Analysis of Heat-Affected Zone of Laser-Cut Heat-Resistant Paper using Otsu Thresholding Technique

Shalida Mohd Rosnan[1]
Graduate School of Life and Environmental Sciences
University of Tsukuba
Tsukuba, Japan

Kong Peifu[2]
Degree Programs in Life and Earth Sciences
University of Tsukuba
Tsukuba, Japan

Toshiharu Enomae[3], Nakagawa-Izumi Akiko[4]
Faculty of Life and Environmental Sciences
University of Tsukuba
Tsukuba, Japan

Abstract*—Since ancient times, natural fibers have been essential in paper production and packaging fabrication. However, beauty-marring carbonization, or a heat-affected zone (HAZ) generated during the laser cutting process of paper materials led to an intriguing discussion on the possibility of reducing this defect zone. Thus, paper loaded with aluminum hydroxide [$Al(OH)_3$] (AH) was prepared and tested with laser cutting. There were two input parameters of laser processing: the ratio of laser power to a maximum and the cutting speed. The study discussed the HAZ area of the paper with AH loaded at 0–40% on a dry pulp basis. The HAZ area was measured through image processing software. The Otsu thresholding technique (OTT) was applied to HAZ area determinations. The results from the image analysis signified that the smallest HAZ area was successfully achieved on samples with AH loaded at 40%. The optimal condition for the sample with 40% AH loaded was 60% power ratio and 20 mm/s in cutting speed. Based on the results, the cutting speed was the most significant parameter to produce the smallest HAZ area; therefore, the laser processing parameters were optimized to achieve a minimum HAZ area, and it was possible to reduce its dark color appearance of the material surfaces. Based on this study, it was found that the application of the Otsu thresholding technique was of significance to the HAZ area determination and reduction of the time consumption for the image analysis.

***Keywords*—Heat-affected zone; image analysis; image processing; laser cutting; thresholding**

I. Introduction

The digital image may be described as the two-dimensional function which refers to f (x, y). To make it easy to understand, pairs of x and y are plane or spatial coordinates, and the amplitude of function f at any coordinates of x, y is called gray level or scale. Thus, when x, y, and f are all finite, we call the function f a digital image. The digital image processing is performed with a computer [1] its techniques has been widely thriven and they are now used for all kinds of tasks in various areas such as image recognition technology along with computer vision and thresholding techniques [2]. Image analysis is an area related with image processing and the extraction of useful information from an image is beneficial to various fields ranging from computer vision applications to medical bioimaging analyses. Vision is the most advanced human senses; however, our vison is limited to the visual band of the electromagnetic spectrum. Thus, imaging machines such as electron microscopy, and computer-generated images used to operate on images [1]. Thus, digital image processing and analysis so popular nowadays and encompasses a varied field of applications. There is almost no technical area that is not impacted by the digital image processing. In our case, it is crucial to extract useful information from images that can help determine the HAZ and improve the laser cutting process. In the laser cutting process, image processing techniques and image analysis are helpful for HAZ determination.

The laser cutting process has recently become a fascinating element in paper-based product fabrication. This is because of the wide range of applications from cutting to engraving [3], perforating, and designing complex geometrics [3]–[8]. Moreover, due to the growing personal needs as well as product quality and high productivity, it is convenient to apply laser cutting technology. However, carbonization and heat-affected zone (HAZ) seem to lead to an intriguing discussion on whether reducing the HAZ in the laser cutting process is possible. Laser cutting is an efficient method for cutting materials as it can be operated readily. Through understanding, high-quality cutting optimal parameters can be obtained. Even though laser cutting has been applied since the 1990s for cutting, perforating, and creasing paper materials, unfortunately, only a few works in literature have been discussed using the laser process of paper products. The mechanisms of laser cutting for paper or other wood-based materials are a thermochemical decomposition [8][9]involving vaporization, meaning the paper material reacts to heat generated by laser beam irradiation and evaporates from the surface.

The study of laser cutting quality focused on two laser processing output which are HAZ and kerf of the cuts. However, in this study, only the HAZ output will be discussed. To analyze the HAZ quality, image analysis needed to determine the optimal setting of the laser process. Over the years, image processing and image analysis studies in the laser cutting has been grown [10] [11]. Even there was various image processing techniques reported in the literatures, this study focused on the Otsu thresholding. In computer vision and image processing, Otsu thresholding technique (OTT) which is named after Nobuyuki Otsu, is performed to obtain automatic image thresholding. It is a simple form of algorithm. The algorithm returns a single threshold intensity that can separate pixels into two regions: foreground and background [12]–[14].

This paper is of significance on the study of HAZ and the laser cutting field. By analyzing the HAZ images and determining its HAZ area, the determination of laser cuts can be done faster and cost effectively. The use of digital image analysis has been widely thriven in laser cutting studies [15][16]. However, the literatures only reported the HAZ measurement values by measuring the mean HAZ width without any further discussion on how to determine the HAZ area. Thus, this study aimed to explore and investigate the image processing and image analysis of HAZ using ImageJ global thresholding techniques. This study reported the image processing technique and image analysis used to improve HAZ image analysis that beneficial to the laser cutting field.

II. Literature Review

In the laser cutting, especially during the heating and vaporizing processes, the paper material absorbed heat energy, carbonized, and produced a HAZ. Thus, when laser cutting is performed, the process of material degradation will cut cellulose molecular chains to its shorter elements comprised of small chains of polysaccharides, carbon dioxide, carbon, and water vapor [17] [18].

Paper materials are used in packaging development because they are environmentally friendly materials because of no carbon emission and play roles as alternatives to petroleum-based plastics for packaging. The application of paper-based packaging has become widespread in alignment with a campaign to zero single-use plastics consumption and goals. For understanding the application of paper in packaging and its fabrication using laser cutting, the nature of paper must be understood first. A natural and renewable resource like plant fibers is eco-friendly, low cost, and has soft features and properties [19]. The composition of plant fibers is mainly cellulose, hemicellulose, and lignin. These three components determine the plant fiber properties. Plant fibers are a significant raw material that influences the thermal properties of paper. Some paper products, such as heat-resistance label paper and heat-resistance release paper, require fiber with high thermal stability [20]. In the thermal analysis conducted by Soares et al. 2001 [21][4], the pyrolysis of hemicellulose and cellulose readily happened. It has been reported that the loss of hemicellulose occurred between 220ºC and 315ºC, while that of cellulose occurred between 250ºC and 400ºC. However, lignin degrades at a wide temperature range from 160ºC to 900ºC, meaning it is more difficult to decompose. Consequently, plant fibers have poor heat-resistance, and its decomposition begins at 220ºC, limiting its application at high temperatures. The mechanism of laser cutting for paper or other wood-based materials is a thermochemical decomposition process [18]. The cutting mechanism is vaporizing, which means the paper material will react to the laser beam and evaporate when the heat of the laser beam reaches the surface. Thus, when laser cutting is performed, the process of material degradation will reduce cellulose from large cellulose molecules to its small elements, which are small chained of hydrocarbon, carbon dioxide, carbon, and water vapor [9]. Recently, natural fiber has become essential in paper-based product fabrication and laser cutting applications.

Aluminum hydroxide [$Al(OH)_3$] (AH) is known as a green inorganic flame-retardant [22] and has been applied as a heat-resistance compound used in flame retardant systems [23] for years. For instance, AH is used in the making of heat-resistance and flame-retardant papers [20]. Many methods have been proposed to modify the paper with AH. One of them was a placement of AH on the surfaces and inside walls of fibers. By implementing the method, AH was successfully distributed evenly on the inside and surfaces [24]. During exposure to high temperatures, AH starts to decompose while absorbing heat sourced from the radiant energy of the laser and eliminating water. Then the emitted water molecules absorb heat to evaporate and decrease the temperature around the HAZ. This mechanism leads to the improvement of the thermal properties of the modified fibers and paper. Based on the Yang, F. et al. study [24], the paper samples modified with AH content succeeded in resisting the heat by determining its thermal properties using the thermal gravimetric analysis (TGA). However, their study did not link with the application of laser irradiation process or discussed any thermal process [16].

Generally, the basic mechanisms of laser cutting are exceptionally easy to understand. Infrared light contained in a high-intensity beam is generated by a laser [25]. The beam focuses onto a material surface through a lens. The focused beam functions to heat the material and caused it to melt in case of plastics and vaporized [17]. Then, the molten material will be ejected from the area by pressurized gas. The schematic of laser cutting can be seen in Fig. 2. This study also discussed the effects of various combinations of laser power ratio and cutting speed on the area of HAZ. The samples with different contents of AH loaded were tested. The laser cutting process has been applied to them, including the control sample (without AH loaded) to determine whether it resists the heat and reduces the HAZ production during the laser cutting process. The laser cutting process has been applied to them, including the control sample (without AH loaded) to determine whether it resists the heat and reduces the HAZ production during the laser cutting process.

Based on the literature, in determining the optimal setting of HAZ, the cutting speed should be maximized to prevent excessive heat transfer depending on this parameter [26]. It also highlighted that laser power variations show less impact, and cutting speeds show more impact on HAZ. Furthermore, based on previous literature studies, the quality of a cut can be obtained by determining one of the outputs which is HAZ [27]. The significant reason that those two parameters were chosen has been discussed in literatures [4][5][8]. In this study, laboratory-made paper sheets loaded with AH were cut using CO_2 laser processes at different laser power ratios and various cutting speeds. These processed samples were micrographed with a focus on the HAZ to analyze it using image processing software ImageJ [28]. Even though many methods were presented in the literature using various software tools, ImageJ was chosen because it is the most popular open-source software provided with a detailed manual and tutorials [22].

In image processing, there is no general agreement among authors regarding when and where the image processing stops. Also, in other related areas such as image analysis and computer vision, and when it starts. However, it is important to highlight that in image processing there are types of computerized processes involved; segmentation, classification (recognition) of individual objects, extracted edges and contours [1]. Image segmentation refers to the algorithms class that can partition the image into different pixel groups or segments. In that case, image thresholding is the best and simplest image segmentation as it partitions the image into two pixels groups which are foreground and background [29].

Thresholding is an essential application, especially in any field of study that relies on image detection, inspection, and classification. The thresholding performance depends on the object-background difference in gray level, noise, and the size and position of the object [30]. The process of selecting the optimal thresholding algorithm is the most important step in image processing and analysis. Image processing has been applied in many fields with computers to process digital images through basic operations or advanced algorithms. Thus, we can extract various information from the original images through image processing and analysis. However, the manual threshold process was not preferable in the image analysis for its several limitations. Especially with a high possibility of high bias while performing the image processing [31]. It is also time-consuming to perform a manual threshold for analyzing many images. Therefore, the most crucial part of binarization is identifying the best threshold algorithms and comparing them with a least-biased comparison [23] [24].

There are 16 universal thresholding algorithms offered by ImageJ, which are Huang, Huang2, Intermodes, IsoData, Li, Mean, MinError(1), Moments, Otsu, Percentile, RenyiEntropy, Shanbhag, Triangle, and Yen. However, this study focused on the Otsu thresholding technique (OTT) based on its advantages in terms of simplicity, robustness, and the most widely applied thresholding algorithm [30] [32]. Apart from those reasons, we mainly chose OTT because we focused on determining the HAZ by identifying the optimal thresholding that is highly capable of detecting the area and edge of HAZ. In our case, HAZ images need optimal thresholding to avoid erroneous detection of an edge, dark pixels in the background, and an excessive number of pixels for the edge [33]. Therefore, in our case, OTT was the best option because of its performance in detecting edges and the HAZ area properly. The equation and demonstration of OTT application has been discussed in literature [33]. The OTT is based on the histogram of grayscale images, the main idea is to maximize the between-class variance (foreground and background region) in the histogram. By using OTT, we get the best threshold value that maximizes the variance of HAZ area and background.

OTT was proposed by Nobuyuki Otsu [12] in 1975. Since then, Otsu's technique has become the most widely used in image thresholding algorithms. The concept of OTT can be simplified into the following steps: (1) processing the input image, (2) obtaining image histogram (pixels distribution), (3) computing the threshold value (t), and (4) replacing the pixels obtained from the image by partitioning white and black regions. One of the advantages of OTT is its simplicity in calculating the threshold values because it only involves single-dimensional intensity data; therefore, the process helps in reducing the time for computational processing. OTT is useful in performing image segmentation because it involves a simple mathematical expression in its algorithm. Thus, it requires less computational time compared to other thresholding techniques. Owing to the advantages, the improvement of OTT has been discussed in Kalathiya et al. study [34]. The OTT (iterative and custom approaches) modified by them are also demonstrated in their study.

Over the years, the scientific community has explored the thresholding techniques in image processing. Apparently, Otsu technique appeared as the most prominent for this purpose. ImageJ Otsu global thresholding algorithms is an implementation of the OTT [12]. The histogram is classified into two classes and the inter-class variance is minimized. The Otsu thresholding can be found in ImageJ software under Auto Threshold options. Gonzalez and Woods have been discussed the Otsu method in depth in the literature. The Otsu's algorithm summarized as; (1) computation of the normalized histogram of the input image, (2) computation of the cumulative sums, (3) computation of the cumulative means, (4) computation of the global mean, (5) computation of the between-class variance term, (6) Obtaining the Otsu threshold and (7) computation of the global variance [1].

However, our study mainly focuses on applying the OTT for extracting HAZ. Although many studies in the literature have discussed the characteristics of HAZ generated in a laser cutting process, most of the recent literature mentioned more about the use of image processing to measure the HAZ area without any further discussion on how it is done. Therefore, this study is useful for the laser cutting process field of study in terms of image detection and analysis. Through this kind of study, HAZ can be easily determined, optimizing the laser process and its relations to input parameters, and minimizing the HAZ can be achieved.

III. Materials and Methods

A. Paper Sheet Preparation

The material used in the experiments was laboratory sheets prepared from pulp fibers and loaded with AH to reinforce heat resistance. The fiber materials used were commercially available hardwood pulp (Eucalyptus species). AH in the special grade was purchased from Fujifilm Wako Chemicals, Osaka, Japan. Alkyl ketene dimer (AKD) (AD1608, Seiko PMC Corporation, Tokyo, Japan) as a sizing agent and polyamine polyamide epichlorohydrin (PAE) (WS4030, Seiko PMC Corporation) as a cationic polymer for retaining the size. To prepare laboratory sheets, a 1.0% AKD emulsion was first prepared by weighing 1 g of AKD emulsion stock solution (25% solids) and diluting it 25 times. Next, a 1% cationic polymer solution was prepared likewise from the PAE stock solution (25% solids). A total of 27 g of air-dried hardwood pulp was weighed and disintegrated in 1.6 L of water using a disintegrator specified in the ISO standard for 5 min and poured into a sieve for drainage. Then, the pulp was squeezed to 10% solids and beaten to 5000 times on a PFI mill. The pulp beating process swells pulp fibers, and therefore improves fiber flexibility and mechanical strength[35]. Next, the beaten pulp slurry was diluted to 1.6 L of water and disintegrated at 3000 rpm with the disintegrator for 1 min. Then, the suspension was diluted to 11.25 L of water, and 7.5 ml of sizing agent (1.0 % AKD) was added and stirred for 1 min, followed by adding AH (10%, 20%, and 40%) and 7.5, 15, and 30 mL of retention aid solution (PAE), respectively for sample B, C, and D as shown in Table I. It was then mixed for 1 more min. AKD works as a sizing agent, and PAE works as a retention aid when the pulp is mixed with the AH component. Paper sheets with a targeted basis of about 100 g/m^2 were prepared using a standard sheet former (ISO 5269-1:2005), pressed with the hydraulic press at 0.34 MPa, set between metal rings, and stored in a controlled environment (23ºC) for restraint air drying.

TABLE I. THE COMPOSITION OF SOLUTION

Sample	**Composition of solution for hand sheet paper making**		
	AH to dry pulp mass ratio (%)	***AKD (mL)***	***PAE (mL)***
A	0	0	7
B	10	7.5	7.5
C	20	7.5	15
D	40	7.5	30

B. Laser Processing Setup

The experiments were conducted with a CO_2 laser cutting machine (PODEA 20W ZERO Corsa, PODEA Co. Ltd, Saitama, Japan) with a maximum output power of 20 W with a spatial resolution of 1000 dpi. A schematic of a laser beam head and cutting mechanism is shown in Fig. 1. The vector cutting speed is 20 mm/s at a maximum. One straight cut, each of 5 mm in length (Fig. 2a and 2b), was performed for each experimental run. Then, the results of the laser cuts were evaluated in terms of the HAZ area.

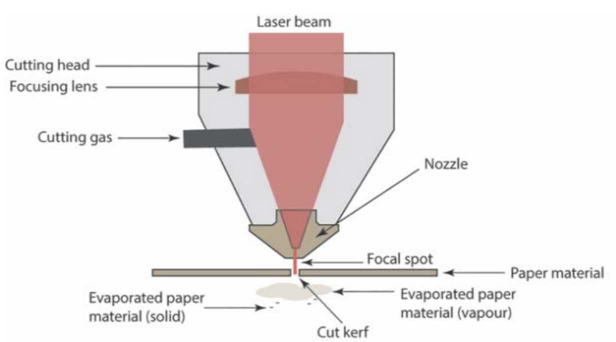


Fig. 1. A Schematic of a Laser Beam Head and Cutting Mechanism.

C. Cutting Parameter Choice

There were two laser parameters chosen: (1) laser power ratio expressed in percentage and (2) cutting speed (mm/s). The laser power ratio was chosen from 20%, 40%, 60%, 80%, and 100% to the maximum and the cutting speed was chosen from 4, 8, 12, 16, and 20 mm/s. The process parameters were performed for each paper condition. The laser cuts with the smallest areas of HAZ were chosen for optimization [16].

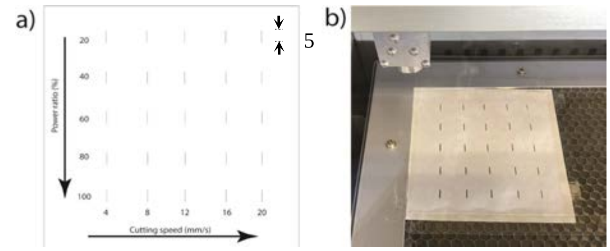


Fig. 2. (a) Illustration of Laser Cut Design (b) Laser Cutting Process.

D. Scanning Electron Microscopy (SEM) and Optical Microscopy

SEM images of heat resistance paper samples were taken using Hitachi Tabletop Microscopes (TM4000Plus). An industrial stereo microscope OLYMPUS SZX10 (Olympus Corporation, Tokyo, Japan) was used to observe the boundary of HAZ. First, optical images were captured using a digital single-lens reflex camera (EOS Kiss X8i, Canon Inc., Tokyo, Japan). Then, the image processing and analysis were performed using ImageJ software to identify and measure the HAZ area by performing binarization and thresholding algorithms.

E. Image Segmentation and HAZ Area Measurement

The identification of the area and edges of HAZ followed the procedure in Table Ⅱ. First, original HAZ images were converted to the grayscale format via the RGB format. Next, image segmentation was performed with the thresholding values (*t*) determined using OTT.

TABLE II. THE IMAGE PROCESSING AND IMAGEJ COMMANDS

Image Processing	**ImageJ Commands**
Grayscale conversion	Image > Type > 8 bit
Thresholding	Image > Auto (global) Threshold > Select method (All 16 global threshold methods have been tested and optimal thresholding selected.)
Particles analysis	Analyze > Analyze particles

IV. Result and Discussion

A. Scanning Electron Microscopy (SEM) and Optical Microscopy

Samples were coated with platinum (Pt) using a spatter before observation with Hitachi Tabletop Microscopes (TM4000Plus). Fig. 3a-d shows the SEM images of the laboratory sheets with and without AH. Fig. 3a shows no particles in Sample A with no AH loaded. Obviously, AH particles are observed to be distributed increasingly in order of increasing AH dose from Samples B (10%), then C (20%) until D (40%).

Fig. 4 shows a HAZ area clearly. The clear definition of HAZ can be defined and referred to the dark burn area that clearly visible under the optical microscope. By analyzing the image, we can determine the HAZ area for each laser process parameter and setting. Each image resulted from the various setting of the laser power ratio percentage and cutting speed can be studied through image processing and image analysis.

B. Image Segmentation

Among the results of image segmentation available with ImageJ for the HAZ images, OTT outperformed the other 15 methods (Fig. 5) in terms of sensitivity and accuracy in detecting the intended areas in the foreground from the background obviously even by visual comparison to the original image. The OTT algorithm is one of the most used methods in binarization [31] [36]. The OTT algorithm refers to the bi-modal image histogram. We focused on OTT not just because it is used widely in many computers vision applications but also because its performance on image segmentation was by far the highest. In this case, the pixels of a HAZ became the foreground. The purpose of thresholding was to separate the HAZ as the foreground from the background. All pixels of a HAZ with intensity lower than a threshold value are considered as the background. After the binarization process, the HAZ area was identified and measured. Based on these values, HAZ was studied for each batch of samples (A to D in Table Ⅰ). In the field of laser process study, it is crucial to determine and understand the HAZ. Moreover, it is important to understand the HAZ area and its correlation with the laser process parameters. Through this process, accurate extraction of HAZ has become possible using the image analysis data. The results showed the smallest values of HAZ area achieved in the condition of optimal laser process parameters for each sample (Table Ⅲ).

As digital image processing is an important tool to analyze the HAZ images. The main step used to perform image analysis in digital image processing through the flowchart is shown in Fig. 6. The process of OTT started with the acquisition of data image, then followed by image segmentation through the Otsu technique. Image segmentation is an essential step in digital image processing to partition the HAZ image into foreground and background regions. In general, there are three segmentation techniques that used in image segmentation which classified as edge-based segmentation, region-based segmentation, and special theory-based segmentation [37]. In our study, we used region-based segmentation using OTT.

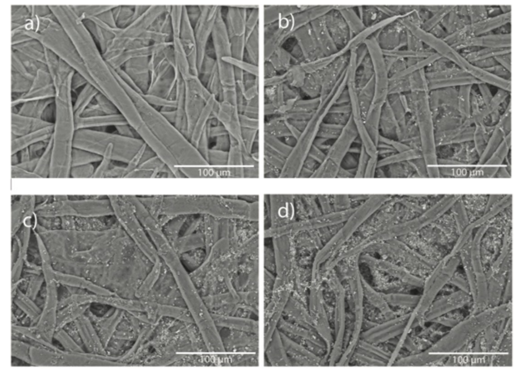


Fig. 3. The Micrographs of the Paper Surface with Different Conditions, (a) Sample A—Control (without Additives), (b) Sample B—10% of AH Loading, (c) Sample C—20% of AH Loading, (d) Sample D—40% of AH Loading.

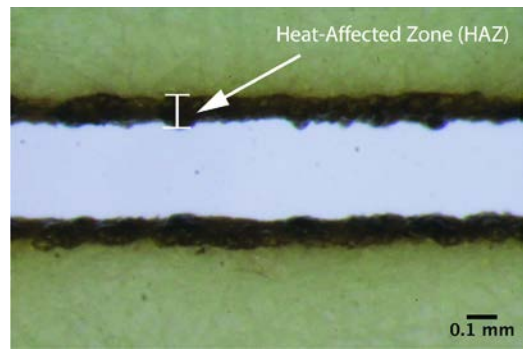


Fig. 4. HAZ Observed through an Optical Microscope.

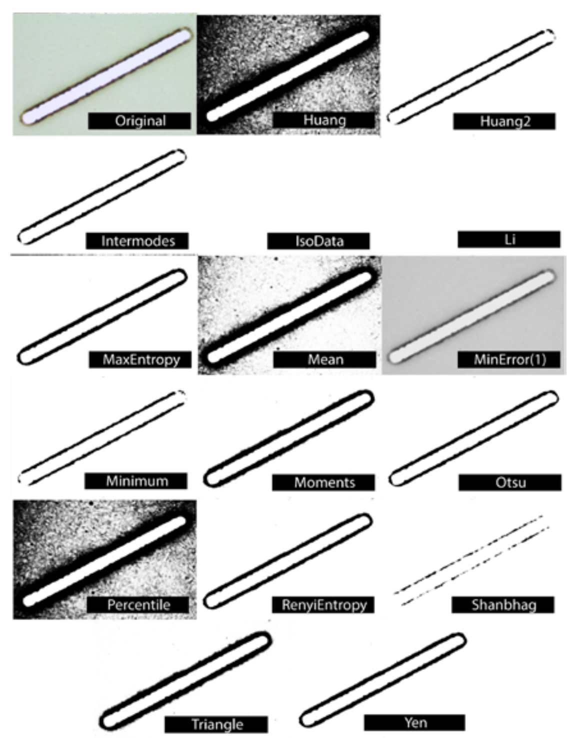


Fig. 5. Original Image and Threshold Images using 16 ImageJ Global Tresholding Techniques.

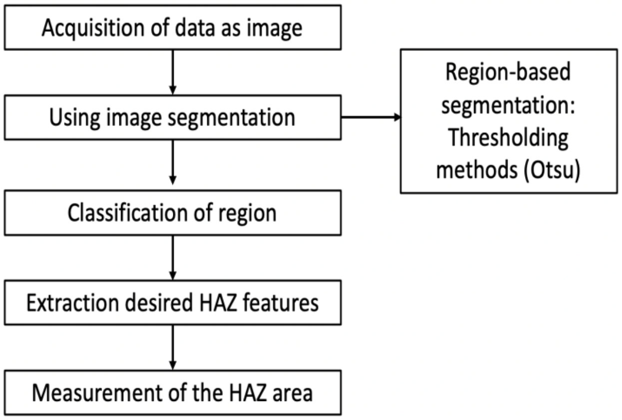


Fig. 6. Main Steps used in Image Analysis in Digital Image Processing.

Meanwhile, Fig. 7 shows some of the overlay images of threshold images on top of original images. The threshold images segmented successfully and unsuccessfully using various thresholding techniques. Fig. 7a shows the grayscale of the HAZ (ground-truth image) and red threshold refers to the detected threshold of HAZ using the mentioned techniques. Fig. 7b shows the successful segmentation of the HAZ using OTT. Fig. 7c and 7d show the unsuccessful segmentation using Shanbhag and Triangle thresholding technique. The criteria that we used to determine the OTT performance by observed the images of original grayscale and threshold images were overlaid on top of each other, and OTT was the best option and optimal thresholding based on its sensitivity in detecting the HAZ area. The OTT-treated image overlapped perfectly with the HAZ.

C. Criteria for Selection of Thresholds

The proposed method makes use some thresholds in HAZ detection for showing the comparison between Otsu and other two examples which are Shanbag and Triangle. As mentioned earlier in Fig. 5, all ImageJ global thresholding performed in identifying the optimal thresholding for the HAZ detection. Thus, in Fig. 7 we overlapped each detected threshold images on top of the original grayscale image. The result of Fig. 7b shown the OTT successfully overlapped accurately compared to the Fig. 7c and 7d which is represent the Shanbag and Triangle technique respectively. However, for another alternative, the detected threshold value of HAZ can be measured or assessed using P-Tile as suggested by Samopa, F. and Akira, A. (2009) and Neogi, N., Dusmanta, K.M., and Pranab K. Dutta (2017) [38][39]. However, the challenges are when it is not applicable when the object area ratio is unknown.

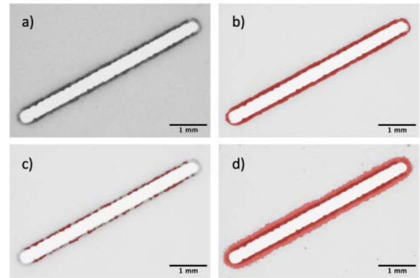


Fig. 7. Some Image Segmentation Results using Different Thresholding Techniques: (a) Original Grayscale Image, (b) Successful Segmentation using OTT, (c) Unsuccessful Segmentation using Shanbhag, and (d) Unsuccessful Segmentation using Triangle.

D. HAZ Analysis

Fig. 8 shows the HAZ area based on various cutting speeds and laser power ratios for Sample A. The low values of HAZ obtained through an optimal laser power ratio of 80% and cutting speed of 20 mm/s gave the optimum desired output of HAZ with the 0.705 mm^2. Secondly, the results of Sample B (Fig. 9) showed that the optimal process obtained the smallest HAZ (0.690 mm^2) through 60% and 100% with a cutting speed of 20 mm/s. However, by considering the power consumption, 60% was preferred compared to the 100% ratio. Thirdly, the results of Sample C (Fig. 10) demonstrated that the highest cutting speed was significant in the smallest area of HAZ. The smallest area value with 0.755 mm^2 was achieved through the highest cutting speed (20 mm/s) and combined with a 100% power ratio. Lastly, the results of Sample D (Fig. 11) showed that the combination of a 60% power ratio with a cutting speed of 20 mm/s successfully obtained the smallest HAZ with 0.629 mm^2. Based on the 100 laser cuts that had been performed, it demonstrated that the modified paper with a different composition of aluminum hydroxide showed resistance to the heat and produced a smaller area compared to the control sample. The smallest area of HAZ obtained through Sample D loaded with the largest amount of AH. The optimal laser process condition of each sample is highlighted in Table III.

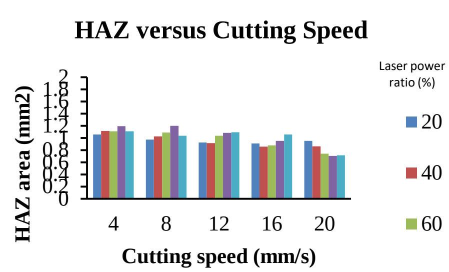


Fig. 8. HAZ Area Compared at Different Laser Power Ratios and Different Cutting Speeds for Laboratory Paper Sheets with no AH.

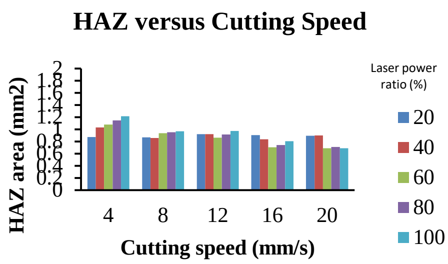


Fig. 9. HAZ Area Compared at Different Laser Power Ratios and Different Cutting Speeds for Laboratory Paper Sheets with 10% AH.

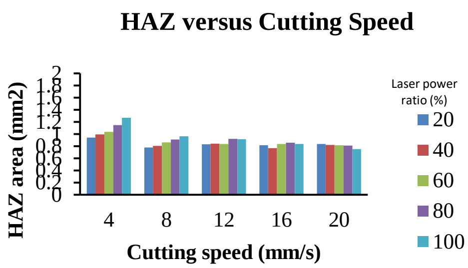


Fig. 10. HAZ Area Compared at Different Laser Power Ratios and Different Cutting Speeds for Laboratory Paper Sheets with 20% AH.

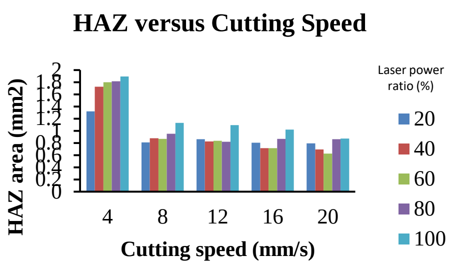


Fig. 11. HAZ Area Compared at Different Laser Power Ratios and Different Cutting Speeds for Laboratory Paper Sheets with 40% AH.

TABLE III. LABORATORY SHEET THICKNESS AND OPTIMAL LASER PROCESS CONDITION FOR EACH SAMPLE

Sample	**Thickness (μm)**	**HAZ area at minimum (mm^2)**	**Optimal condition**
A	140	0.705	80 % power ratio and 20 mm/s cutting speed
B	148	0.609	60 % power ratio and 20 mm/s cutting speed
C	150	0.755	100 % power ratio and 20 mm/s cutting speed
D	183	0.629	60% power ratio and 20 mm/s cutting speed

V. CONCLUSION

In the laser cutting process of paper materials for packaging applications, blackening generation of a HAZ by pulp fiber carbonization mars beauty. However, the characterization and evaluation of HAZ lead to an intriguing discussion on the possibility of reducing this defect zone. Thus, laboratory paper sheets loaded with aluminum hydroxide (AH) were prepared because we highlighted the effect of AH to stabilize heat resistance and carbonization prevention. Laboratory paper sheets with four different amounts of AH loaded were processed by laser cutting. All the HAZ were captured with an optical microscope. The HAZ area was measured by image processing through OTT available with ImageJ. The image segmentation based on OTT permitted to accurately measure the HAZ area. Also, an optimal combination of the laser power ratio and cutting speed providing a minimum of the HAZ area was pursued. We have found that the laboratory paper sheets with the largest amount of AH loaded produced the smallest area of HAZ by a combination of a laser power ratio of 60% and a laser cutting speed of 20 mm/s. Our future work will possibly focus on depth discussion on the other image processing techniques that can be applied in laser cutting image analysis such Gonzalez & Woods technique, and other ImageJ global thresholding techniques.

ACKNOWLEDGMENT

The authors gratefully acknowledge the Japanese Society of Printing Science and Technology (JSPST) for the printing technology research incentive. Also, special thanks to Open Facility, Research Facility Centre for Science and Technology, the University of Tsukuba for allowing the authors to use the equipment. Shalida Mohd Rosnan wants to acknowledge the Japanese Government Monbukagakusho (MEXT) Scholarship under the Special Program on Transworld Professional Human Resources Development Program on Food Security and Natural Resources Management (TPHRD).

REFERENCES

[1] R. C. Gonzalez, R. E. Woods, J. Partridge, and J. Bai, “Digital Image Processing,” 2018. [Online]. Available: www.pearsoned.com.

[2] S. Joseph, “Image processing techniques and its applications: an overview,” vol. 4, p. 2018, [Online]. Available: www.ijariie.com2168.

[3] J. Powell and A. Kaplan, “Laser cutting: From first principles to the state of the art,” in *PICALO 2004 - 1st Pacific International Conference on Applications of Laser and Optics, Conference Proceedings*, 2004, pp. 1–6. doi: 10.2351/1.5056075.

[4] O. Keles, U. Oner, O. Keles[1], and U. Oner[2], *A Study of the laser cutting process: Influence of laser power and cutting speed on cut quality laser cutting*, vol. 20.

[5] F. Masoud, S. M. Sapuan, M. K. A. Mohd Ariffin, Y. Nukman, and E. Bayraktar, “Cutting processes of natural fiber-reinforced polymer composites,” *Polymers*, vol. 12, no. 6. MDPI AG, Jun. 01, 2020. doi: 10.3390/POLYM12061332.

[6] I. Pinćjer, N. Miketić, I. Tomić, and S. Adamović, “Exploring the various parameters of CO2 laser in the cutting of paper,” in *Proceedings - The Tenth International Symposium GRID 2020*, Nov. 2020, pp. 261–267. doi: 10.24867/GRID-2020-p28.

[7] R. T. Mushtaq, Y. Wang, M. Rehman, A. M. Khan, and M. Mia, “State-of-the-art and trends in CO2 laser cutting of polymeric materials-A review,” *Materials*, vol. 13, no. 17. 2020. doi: 10.3390/ma13173839.

[8] R. Shalida and T. Enomae, “The art of gift-giving: Laser cutting as a value-added technology for fabricating customized biodegradable packaging,” in *International Conference on Digital Printing Technologies*, 2021, vol. 2021-Octob. doi: 10.2352/ISSN.2169-4451.2021.37.69.

[9] H. Piili, “A Theory of Interaction Mechanism between Laser Beam and Paper Material,” in *Physics Procedia*, 2015, vol. 78, pp. 163–173. doi: 10.1016/j.phpro.2015.11.029.

[10] C. Tan *et al.*, “Experimental and theoretical investigation of localized CO2 laser interaction with fused silica during the process of surface damage mitigation,” *Results in Physics*, vol. 16, 2020, doi: 10.1016/j.rinp.2020.102936.

[11] O. Anicic, S. Jović, H. Skrijelj, and B. Nedić, “Prediction of laser cutting heat affected zone by extreme learning machine,” *Optics and Lasers in Engineering*, vol. 88, 2017, doi: 10.1016/j.optlaseng.2016.07.005.

[12] Nobuyuki Otsu, “A threshold selection method from gray-level histograms,” 1979, doi: 10.1109/TSMC.1979.4310076.

[13] T. Y. Goh, S. N. Basah, H. Yazid, M. J. Aziz Safar, and F. S. Ahmad Saad, “Performance analysis of image thresholding: Otsu technique,” *Measurement: Journal of the International Measurement Confederation*, vol. 114, 2018, doi: 10.1016/j.measurement.2017.09.052.

[14] R. C. dos Santos, M. Galo, A. C. Carrilho, and G. G. Pessoa, “The use of Otsu algorithm and multi-temporal airborne LiDAR data to detect building changes in urban space,” *Applied Geomatics*, vol. 13, no. 4, pp. 499–513, Dec. 2021, doi: 10.1007/s12518-021-00371-6.

[15] M. J. Madić and M. R. Radovanović, “Analysis of the heat affected zone in CO2 laser cutting of stainless steel,” *Thermal Science*, vol. 16, no. SUPPL.2, 2013, doi: 10.2298/TSCI120424175M.

[16] F. Masoud, S. M. Sapuan, M. K. A. M. Ariffin, Y. Nukman, and E. Bayraktar, “Experimental analysis of heat-affected zone (Haz) in laser cutting of sugar palm fiber reinforced unsaturated polyester composites,” *Polymers (Basel)*, vol. 13, no. 5, pp. 1–12, Mar. 2021, doi: 10.3390/polym13050706.

[17] A. Stepanov, E. Saukkonen, and H. Piili, “Possibilities of laser processing of paper materials,” in *Physics Procedia*, 2015, vol. 78, pp. 138–146. doi: 10.1016/j.phpro.2015.11.026.

[18] A. Stepanov, H. Piili, and A. Salminen, “Color change in laser cutting of paper,” 2010. doi: https://doi.org/10.2351/1.5061990.

[19] K. Mugdha Bhat, J. Rajagopalan, R. Mallikarjunaiah, N. Nagaraj Rao, and A. Sharma, “Eco-Friendly and Biodegradable Green Composites,” 2021. doi: 10.5772/intechopen.98687.

[20] A. M. Youssef, O. A. Mohamed, M. El-Samahy, A. H. Abd-El Kader, and L. Pichtec, “Flame Retardant and Mechanical Properties of Modified Paper Using Inorganic Salts for Packaging Applications,” 2014.

[21] S. Soares, M. P. S. Ricardo, S. Jones, and F. Heatley, “High temperature thermal degradation of cellulose in air studied using FTIR and 1 H and 13 C solid-state NMR,” 2001.

[22] C. Hoffendahl, G. Fontaine, S. Duquesne, F. Taschner, M. Mezger, and S. Bourbigot, “The combination of aluminum trihydroxide (ATH) and melamine borate (MB) as fire retardant additives for elastomeric ethylene vinyl acetate (EVA),” *Polymer Degradation and Stability*, vol. 115, pp. 77–88, 2015, doi: 10.1016/j.polymdegradstab.2015.03.001.

[23] A. L. Mohamed and A. G. Hassabo, “Flame Retardant of Cellulosic Materials and Their Composites,” 2015, pp. 247–314. doi: 10.1007/978-3-319-03467-6_10.

[24] F. Yang, Y. Zhang, and Y. Feng, “Adding Aluminum Hydroxide to Plant Fibers Using In Situ Precipitation to Improve Heat Resistance,” 2017, Accessed: Jan. 20, 2022. [Online]. Available: https://bioresources.cnr.ncsu.edu/wp-content/uploads/2017/01/BioRes_12_1_1826_Yang_ZF_Aluminum_Trihydrate_Plant_Fibers_Improved_Heat_Resistance_10943.pdf.

[25] J. Powell, *CO2 Laser Cutting*. Springer London, 1998. doi: 10.1007/978-1-4471-1279-2.

[26] M. Madić, M. Radovanović, M. Manić, M. Trajanović, and M. Madić, “Optimization of CO2 laser cutting process using Taguchi and Dual Response Surface Methodology,” vol. 36, no. 3, pp. 236–243, 2014, [Online]. Available: www.tribology.fink.rs.

[27] S. Rao *et al.*, “Fibre laser cutting of CFRP composites and optimization of process parameters through response surface methodology”.

[28] M. D. Abramoff, P. J. Magalhaes, and S. J. Ram, “Image processing with ImageJ,” *Biophotonics International*, 2004.

[29] G. E. Sujji, Y. V. S. Lakshmi, and G. W. Jiji, “MRI brain image segmentation based on thresholding.”

[30] T. Y. Goh, S. N. Basah, H. Yazid, M. J. Aziz Safar, and F. S. Ahmad Saad, “Performance analysis of image thresholding: Otsu technique,” *Measurement: Journal of the International Measurement Confederation*, vol. 114, pp. 298–307, Jan. 2018, doi: 10.1016/j.measurement.2017.09.052.

[31] J. Brocher, “Qualitative and Quantitative Evaluation of Two New Histogram Limiting Binarization Algorithms,” 2014.

[32] S. Saleh Bafjaish, M. Sanusi Azmi, M. Nasser Al-Mhiqani, A. Ramzani Radzid, H. Mahdin, and U. Tun Hussein Onn Malaysia Johor, “Skew detection and correction of mushaf al-Quran script using Hough Transform,” 2018. [Online]. Available: www.ijacsa.thesai.org.

[33] B. D. Setiawan, A. N. Rusydi, and K. Pradityo, “Lake edge detection using Canny algorithm and Otsu thresholding,” 2017.

[34] S. Kalathiya and V. Patel, “Implementation of Otsu method with two different approaches,” *International Journal of Software & Hardware Research in Engineering*, vol. 2, no. 2, pp. 24–28, 2014.

[35] W. Gao, Z. Xiang, K. Chen, R. Yang, and F. Yang, “Effect of depth beating on the fiber properties and enzymatic saccharification efficiency of softwood kraft pulp,” *Carbohydrate Polymers*, vol. 127, 2015, doi: 10.1016/j.carbpol.2015.04.005.

[36] L. Nichele, V. Persichetti, M. Lucidi, and G. Cincotti, “Quantitative evaluation of ImageJ thresholding algorithms for microbial cell counting,” *OSA Continuum*, vol. 3, no. 6, p. 1417, Jun. 2020, doi: 10.1364/osac.393971.

[37] A. Singh Chaudhary and D. K. Chaturvedi, “Efficient Thermal Image Segmentation for Heat Visualization in Solar Panels and Batteries using Watershed Transform,” *International Journal of Image, Graphics and Signal Processing*, vol. 9, no. 11, pp. 10–17, Nov. 2017, doi: 10.5815/ijigsp.2017.11.02.

[38] F. Samopa and A. Asano, “Hybrid Image Thresholding Method using Edge Detection.”

[39] N. Neogi, D. K. Mohanta, and P. K. Dutta, “Defect Detection of Steel Surfaces with Global Adaptive Percentile Thresholding of Gradient Image,” *Journal of The Institution of Engineers (India): Series B*, vol. 98, no. 6, pp. 557–565, Dec. 2017, doi: 10.1007/s40031-017-0296-2.

Hadoop as a Service: Integration of a Company's Heterogeneous Data to a Remote Hadoop Infrastructure

Yordan Kalmukov, Milko Marinov
Department of Computer Systems and Technologies, University of Ruse, Ruse, Bulgaria

***Abstract*—Data analysis is very important for the development of any business today. It helps to identify organizational bottlenecks, optimize business processes, foresee customers' demands and behavior, and provides summarized data that could help reducing costs and increase profits. Having this information when designing new products or services highly increases the chances of their success, and thus provides an additional competitive advantage over other businesses. However, having a single data analyst with a computer is far from enough in the era of big data. There are powerful data analytical software tools, but they are either expensive or hard to deploy and require multiple high-performance servers to run. Buying expensive hardware and software, and hiring high-qualified IT experts, is not affordable for all companies, especially for smaller ones and start-ups. Therefore, this article proposes an architecture for integration of a company's heterogeneous data (stored within a database of any type, or in the file system) to a remote Hadoop cluster, providing powerful data analytical services on demand. This is an affordable and cost-effective cloud-based solution, suitable for a company of any size. Businesses are not required to by any hardware or software, but use the data analytical services on demand, paying a small processing fee per request or by subscription.**

***Keywords*—Hadoop integration; data analytical tools; heterogeneous data integration; Hadoop distributed file system (HDFS); HBase; hive**

I. Introduction

Data acquisition and analysis are very important for the development of any business today. They help to better understand the underlying business processes and allow their subsequent optimization and reorganization. The analysis of the collected data provides the company the ability to identify dependencies among processes, objects and subjects; to optimize resource planning; to analyze consumers' needs and behavior; to foreseen customers' actions and demand; and many others. All these analyses give a significant competitive advantage to the company over the other competitors, since it can easily detect trends or bottlenecks and identify ineffective processes to optimize.

For all this to happen, however, the company must have the necessary computer hardware, software and the human resources to do the analysis. In case of processing large volumes of data, having a computer and an expert in the field is far from enough. Manual analysis of big data by a single person is practically impossible. Instead, automated, computer-based tools should be applied in order to analyze big data efficiently and effectively. However, these tools are quite expensive and not every company can afford to buy them, especially start-ups. There are free tools either, some of them very powerful, but not that user-friendly, requiring highly-qualified programmers and IT administrators to deploy and configure them.

A set of powerful data analytical tools is provided by the Hadoop ecosystem [1],[2]. They are built on top of the Hadoop Distributed File System (HDFS). Most of them are distributed as open source applications developed and maintained by community of enthusiast professionals, under the auspices of the Apache Software Foundation. The tools are quite powerful, but as distributed computing applications they should be installed and configured on multiple servers. So, a company that wishes to build its own data analytics infrastructure should buy multiple high-performance servers and hire IT experts to deploy and configure the entire cluster.

Although there are third party installation and management wizards, for example Cloudera Manager, installation of Hadoop services is far from trivial and easy. There are many dependencies that should be considered in advance. For example, since all services run on top of HDFS, if the HDFS itself is not properly installed and configured, the installation of the data analytics tools may fail. The biggest problem is that in case of failure, a subsequent reinstallation is not an option at all, since there are many files left in the server's filesystem from the previous installation. A subsequent reinstallation will just make things messier. Installation and deployment should be done by really qualified and experienced IT experts. Their wages however could be even higher than the money spent for dozens of high-performance servers. As a result, installation and maintenance of own Hadoop infrastructure could be quite expensive and not affordable for small and medium size enterprises, especially for star-up companies.

Fortunately, there is a cost-effective, more flexible and easier way to use Hadoop ecosystem's data analytics tools than building a company's own Hadoop infrastructure. The cloud computing model can help a lot here. An experienced IT company, a service provider, could build the entire Hadoop infrastructure on its own servers, and then offers all the Hadoop's data analytics tools to other companies as services for rental. The idea could be named "Hadoop as a Service (HaaS)".

How could an external organization/company take advantage of the data analytics resources of the service provider? By integrating its data to the Hadoop cluster of the service provider. And this could be done at much lower cost for the external company than if it had to build its own infrastructure for data analysis. During integration however, it should be taken into account that both the Hadoop infrastructure and the company's database are already existing systems. Not just existing systems, but heterogeneous systems. The latter is very important and implies that some intermediate layer should be developed and deployed on top of the Hadoop cluster that will allow external applications/users to communicate with it regardless of their own architecture and implementation technologies. Furthermore, there is another key requirement as well – together with the data integration, the service provider should support data separation. Data separation means that data shared by one company should be invisible and inaccessible to all other companies working with the cluster. That is a key requirement from any business – to prevent illegal industrial espionage and stealing of sensitive data. In this sense, probably the most reliable, flexible and appropriate integration strategy is by implementing a service-oriented architecture (SOA). According to it, the service provides should deploy a wrapping web service (soap-based or rest-based) on top of the entire Hadoop cluster that will allow external authorized users (could be applications or people) to load and analyze data within the cluster.

The aim of this article is to propose an architecture for integration of a company's heterogeneous data (stored within a database of any type, or in the filesystem) to a remote Hadoop cluster, providing data analytical services on demand. Such integration will give the company (business organization) an affordable, cost-effective, access to powerful big data analytical tools as the ones included in the Hadoop Ecosystem. The paper is structured as follows: Section 2 reviews some previous work done by other researchers. Section 3 proposes architecture for integration of a company's data to a remote Hadoop cluster accessible as a service. Finally, Section 4 ends the article with a conclusion, outlining the usefulness of the proposed solution.

II. Related Work

Big data management has introduced some challenges (maintaining heterogeneous data, large data volumes and increased data throughput) for applications, related to data processing. Currently NoSQL databases are proposed to address these challenges by offering horizontal scalability, high availability and data storage without using fixed schemas. NoSQL databases, however, do not have a standard query language, which makes developers' life harder. On the other hand, the traditional relational databases and the SQL language are very popular for processing and storage critical data, but not suitable for large volumes of data.

As a result, multiple approaches have been proposed to define, process and store large volumes of relational data in NoSQL databases by using similar to SQL interfaces. They all focus on both scalability and availability. Schreiner et al. present a comparative analysis [3] of these approaches based on their architectural solutions. The authors motivate their research with the fact that the use of generalized architectural solutions is a proper strategy for relational-based applications, which could be used to move/migrate relational data to NoSQL databases. One of the approaches for that is to propose a way to access NoSQL databases by SQL-like commands. Suggested methods, however, convert the relational data model to a single NoSQL data model and sometimes directly to a specific NoSQL database management system (DBMS). Schreiner et. al. present a canonical approach, called SQLToKeyNoSQL [4], which converts relational schemas and SQL commands to equivalent schemas and data access methods for any NoSQL DBMS, based on the key-value data model, the document data model, and the column-oriented data model. The authors propose architecture of a layer, focusing on strategies for data transformation and mapping.

The analysis of the relational and the non-relational databases leads to the conclusion that these data storage and processing systems are to some extent complementary. Unfortunately, the complementary behavior has a negative effect on the integration possibilities, both in terms of data model and data processing. In terms of performance, it may be useful to rely on multi-lingualism, multi-model approach or multi-step modeling, or even to transform the SQL schema into a NoSQL model and then to migrate the data. Another option is to integrate relational and non-relational databases by the help of a third-party data model. This option is discussed by Pokorný in [5]. Together with that, the author presents some new methods for designing such integrated database architectures. He also discusses that it is not possible to simply apply traditional approaches for integration of relational and non-relational DBMSs, due to the complementary nature of these two types of databases. The author reviews multiple methods for integration of these heterogeneous databases. The heterogeneity itself leads to a new set of problems. NoSQL databases are flexible, but their data design also requires specific modeling solutions that affect their performance in the integrated architectures.

Integration of data stored in heterogeneous systems is a quite challenging task as well and very difficult to solve. Vathy-Fogarassy and Hugyák propose a new data integration methodology [6] in order to provide personalized data queries to multiple relational and NoSQL database management systems. Their solution does not support joins and aggregation of data from multiple sources, but it just collects and migrates data from/to the different individual sources. The method is based on a metamodel approach and covers the heterogeneity of the source systems in terms of their structure, semantic and syntactic features.

Due to the complex management of organizations, the corporate applications could contain a vast number of tables with multiple relationships and constraints among them. Organizations are currently still using relational databases, but the NoSQL systems are constantly increasing their market share due to their excellent performance and high availability. So, more and more migration tools will be needed to migrate relational to non-relational data models. The database schema is important not just for the relational databases, but for the

NoSQL systems as well, since it could provide better query execution. Although the NoSQL databases do not explicitly require database schemas, maintaining such information is important for activities such as data integration, data validation and operational compatibility. Frozza et. al. suggest in [7] a process to automatically extract schemas of NoSQL databases relying on the column-oriented data model. The authors use JSON as a canonical data representation format and test their approach by extracting schemas from HBase. Dai discusses a transformation technique [8] that is capable of transforming database schemas, and translate and optimize queries. Other approaches to design and transform schemas could be found in [9],[10].

Processing and storage of large data volumes by using conventional techniques is not possible. Nowadays, organizations should target their development to non-relational (NoSQL) systems since they support more flexible data models. It is a great challenge for business organizations to transform their existing relational data to NoSQL data models, especially when having in mind the heterogeneity and the complexity of their data. Furthermore, big data management faces an additional challenge – data cleaning and flushing. Ramzan et al. suggest a solution that supports two modules [11] – data transformation module and data cleaning module. The first phase transforms the relational database to a NoSQL database by applying an appropriate transformation model. Then the second phase provides data cleaning, aiming to improve data quality and prepare them for further (big data) analyses.

Most data model transformation methods target a single (specific) NoSQL model and provide a little or no support for transformation's personalization. Kuszera et al. present an approach [12] for transforming relational to NoSQL databases that supports the document-oriented and the column-oriented data models. Their method uses as an input a set of directed acyclic graphs (DAGs) that represent the targeted NoSQL model. DAGs are used to generate data transformation commands. The approach supports different relations' cardinality and the transformation commands could be personalized. The authors developed a tool that interprets the input DAGs and supports multiple transformation strategies.

Since big data processing applications, based on Cloud Computing become more and more popular, many existing systems will upscale their services in order to support the increasing volumes of data. Liao et al. propose a data transformation system [13] that provides hybrid architecture, including both relational and NoSQL databases. It supports simultaneous query processing and data transformation, and data synchronization. The authors focused their work on the transformation speed and the diversity of the data.

Modern applications that process different data formats often use several types of databases, and the need to migrate data between them is common. Although there are multiple methods to perform data model conversion, the process of choosing the ideal data structuring for specific application requirements is not a trivial task. Kuszera et al. describes an approach for converting relational databases to NoSQL ones, consisting of multiple steps [14] for defining, evaluating and comparing alternative NoSQL schemas (data structuring), before migrating the data.

III. Architecture for Integrating a Company's Heterogeneous Data to an Existing Remote Hadoop Infrastructure

The existing approaches, reviewed in the related work, offer an integration of a relational database management system to a non-relational one with a specific target data model. This is important for companies to move their data from an old relational database management system to a modern, flexible NoSQL one, supporting data models suitable for big data processing. However, this data model transformation does not provide any data analytics at all. It just transforms the existing data. In contrast, the architecture proposed here allows integration of any type of data to an existing Hadoop infrastructure, providing powerful data analytical tools.

In order to make the most of the data analytical capabilities of the Hadoop Ecosystem, the company should be able to share not only its database, but also files whose content could be analyzed as well. So, integration is needed on multiple levels:

- Integration with the Hadoop Distributed File System (HDFS).
- Integration with the real-time column-oriented database management system HBase.
- Integration with the Hive service and obtaining aggregated analyses of it.

Integration with HDFS allows the company to share data exported from any type of database managements systems (DBMSs) – relational or NoSQL. Data could be exported in the form of CSV (comma separated values) data files or in other format. Once loaded in HDFS, these files could be later imported in any Hadoop's data analytics service. This approach provides very high level of flexibility and abstraction from the specific DBMS, and allows integration of any type of data to the Hadoop cluster. Integration could be implemented by using HDFS's REST API, called WebHDFS. There is other alternative as well – through the HDFS client distributed together with Hadoop, but the WebHDFS provides higher flexibility.

If it is needed the company's data to be stored in real time within the Hadoop infrastructure, then an integration with the column-oriented database management system HBase [15],[16] is necessary. It could be done through:

- The HBase's REST API.
- Thrift Server.

There is a third way as well – by storing data in files and loading them in HDFS, but it is not suitable for real-time operations.

The HBase's REST API provides an easy access to HBase, and since communication is done over the HTTP protocol, the client could be implemented in any programming language.

The problem however is that both the RESTAPI and the Thrift Server do not provide any access control by default. If a user knows the namespaces and the tables' names of another user, then the first one could easily steal the data of the second user. A possible solution of that is to use the Kerberos authentication protocol [1]. An alternative solution is to develop a wrapping web service around the HBase's REST API (or the Thrift Server) that should handle user authentication and perform access control.

Another useful integration is with the Hive service [1],[2]. Hive is not a database, but just a query engine that allows searching and retrieving data from different data stores by using a language that is almost identical to SQL. It is called HQL – Hive query language. The key point here is that data could be retrieved and analyzed from different data stores. Actually, HQL runs over temporarily virtual tables that could be built from multiple sources – HBase, ordinary files from HDFS and other sources. This is the highest strength of Hive – it allows using SQL to search and process data from files. Usually, files which are mapped as virtual tables are csv and tab-delimited data files, but could be also MS Excel or other sources of structured or semi-structured data. So, Hive could be used to perform data aggregation and analyses of: real database data; system log files; application log files; accounting data files; csv and MS Excel spreadsheets; and many others sources of data. Within the real world, such data files are generated as often as data stored within a database. Furthermore, Hive supports the Hadoop's MapReduce framework, meaning that the HQL queries could be distributed among multiple servers and run in parallel, which allows analysis to be performed on very large amounts of data.

Fig. 1 presents a conceptual architecture for a company's data integration (database and filesystem) to an existing Hadoop cluster. It implies design and development of just two new modules (applications) – "the integration middleware / wrapping service" from the service provider's side and the "Customer's data integration tool" from the client's side. All other components (except DBeaver) are existing services provided by Hadoop or "built-in" applications in the Hadoop Ecosystem.

Modules marked in peach color are the main Hadoop services - HDFS, HBase, Hive and Hue [1],[2]. Modules marked in white within the cluster are accompanying application programming interfaces (APIs) that allows third party applications to interact with the Hadoop services (the modules in peach). These APIs are not directly built-in within the main Hadoop services, but implemented as separate (external) applications, which however are usually distributed together with the services. The easiest way to install all of them is by installing Cloudera Manager that deploys the Cloudera Distributed Hadoop (CDH).

Cloudera Manager offers a user-friendly, comprehensive interface for installation and management of the entire Hadoop Ecosystem. It could be used to start, configure and stop individual services in the cluster. However, if services are installed as stand-alone applications, then the APIs should be started from a command-line interface and configured through xml files.

DBeaver is a database client, in general, that could connect to multiple relational and non-relational DBMS, and to some services (including Hive) from the Hadoop Ecosystem as well. It provides a graphical user interface to execute SQL queries (and not only) and visualize the results. It can be used directly by end users, without any skills in system administration and programming.

The company's database could be of any type – relational, non-relational, or even file-based. Its data are retrieved based on the database type and then packed within the necessary data structures, which are sent to the "Integration Middleware" for further processing and storing within the cluster.

The two custom and important modules are "Integration Middleware / Wrapping Service" and "Customer's data integration tool". They are the ones, which are directly responsible for data integration. In contrast to the others, they do not exist in advanced, but should be designed and implemented entirely for the purpose of integration.

1) "*Customer's data integration tool*": provides a graphical user interface (GUI) that allows the company to manage the data it wants to integrate (including to choose what exactly to share) with the Hadoop cluster, and also to render/visualize the received results. This module exactly extracts the data from the local database and packs them in suitable data structures (usually JSON objects) that are later send to the "Integration Middleware / Wrapping Service". The module should work with any type of database management systems (relational or non-relational) and pack the data in similar manner, regardless of their type. The customer's data integration tool should allow browsing not only the local database, but the local filesystem as well. User should be able to browse, select and upload files to the remote HDFS filesystem for further processing and analysis by Hadoop's data analytics services.

2) "*Integration Middleware / Wrapping Service (API)*": implements data integration on the service provider's side. It works as a wrapping service around the entire Hadoop cluster. The customer's data integration tool works with this module only, not directly with the Hadoop services or APIs. In this sense "Integration Middleware / Wrapping Service" *serves as a mediator* between the company (the client) and the WebHDFS API, HBase API and the Thrift Interface. But why such a mediation service is needed? Because all the three interfaces do not provide any user authentication and access control. For HBase, for example, if one user knows the namespaces and table names of another user, then the first one could steal the data of the second one. Similar apply to the Thrift interface as well – all connections to HBase are done through the same socket without any authentication. All users have access to everything, by default. And that is a big problem. No company will share its data if they become publicly available and easily accessible to competitors. To solve this problem the Kerberos authentication service could be installed and configured, or to implement a wrapping service around the APIs to perform user authentication and access control. The latter could be easily integrated in this module and perform a flexible, custom access control.

The second important task of this module is to convert the data from their source model (relational or whatever) to column-oriented one, suitable for loading in HBase. This is the module that is solely responsible for data conversion. They arrive from the "Customer's data integration tool" packed as JSON objects, then they should be converted to suitable model and loaded to HBase through the API or the Thrift Interface.

3) Conversion of source data models to the column-oriented data model suitable for HBase: As already mentioned, conversion from a source data model to a target data model is important for data integration from heterogeneous systems. Research in data models transformation has intensified a lot during the last years – multiple approaches have been proposed, as seen from the related work section. Most of them apply the relational model as a source model, since there are many legacy data storage systems currently in use that relies mainly on the relational model.

In contrast to the traditional relational databases, which are strictly normalized and should meet the requirements of the 3-rd and the BCNF normal form, almost all NoSQL data stores are actually not normalized at all – data could be grouped and replicated in order to increase the read speed. Although the NoSQL databases also maintain some consistency, there may be inconsistencies in the data for a short period of time after performing an operation. As a result, the NoSQL databases are mainly suitable for OLAP systems, but not for OLTP systems, where support of the ACID properties is mandatory.

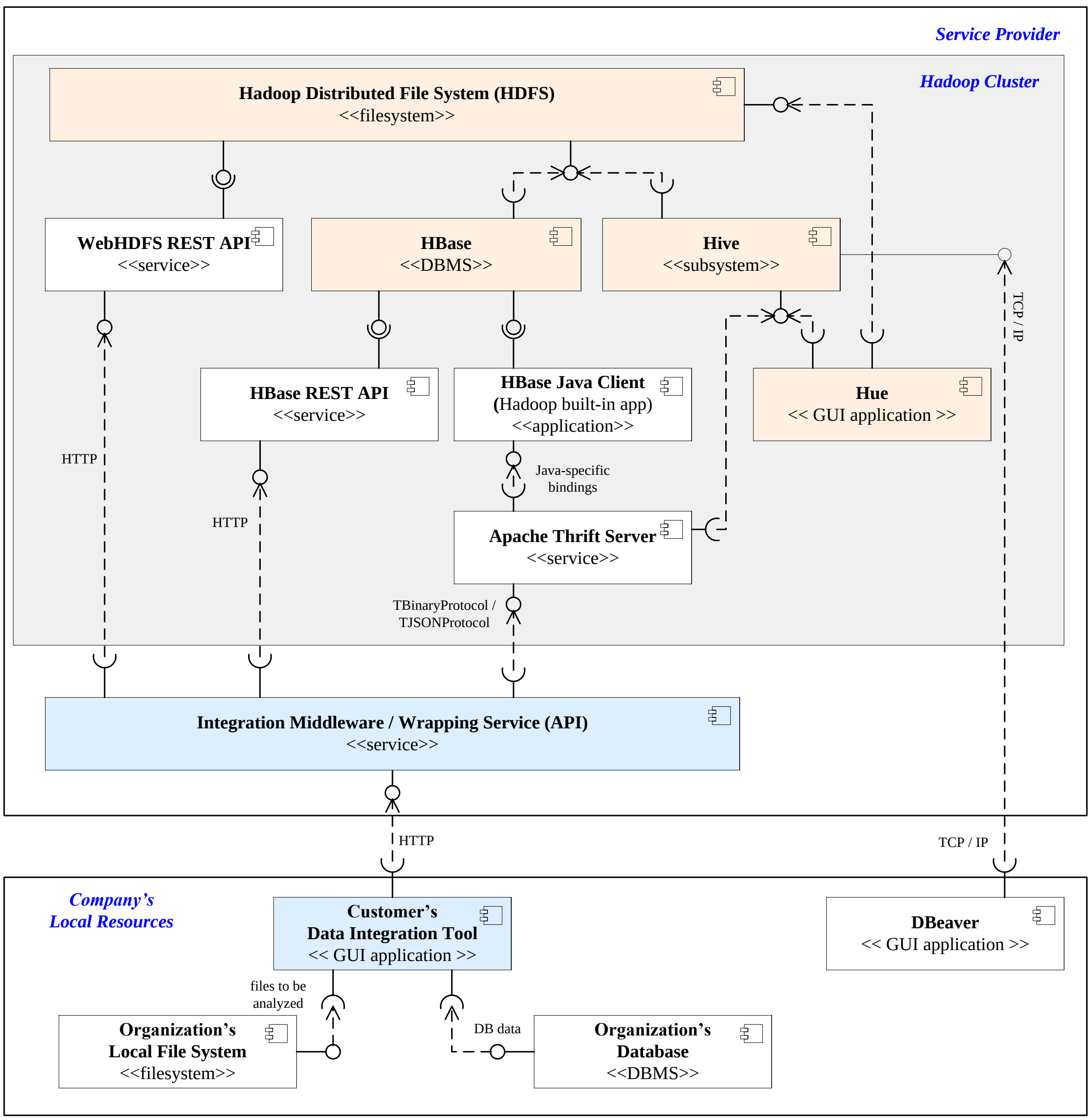


Fig. 1. Architecture for Integration of a Company's Database and Filesystem to Hadoop Infrastructure.

Any data model could be transformed to any non-relational model by applying the methods and algorithms described in [17],[18],[19],[20]. Since this article is related to data integration with an existing Hadoop cluster, the target model here is the column-oriented data model, used by Apache HBase [15],[16]. Here is an intuitive heuristic approach for transforming the relational data model to a column-oriented one:

1) Grouping all related data in a single column family. These are data that should be read/written from/into the database at once. These are the tables' records (rows) in the relational databases. Actually, the column-oriented schema could be easily extracted from the relational schema, since the latter provides more than enough information about attributes' relationships. Every table in the relational database becomes a table in the column-oriented database, and the primary keys of the former are used as row keys in the latter. All attributes within the relational table are group together in a single column family, in order to guarantee that they will be stored and read together.

2) Convert the relationships between tables in the relational schema to "relationships" in the column-oriented schema. There are three types (cardinalities) of relationships in the relational schema: one-to-one; one-to-many; and many-to-many. They will be reviewed separately since the conversion is done in different way:

a) One-to-one. This is intuitive – a column with a foreign key is treated is an ordinary column and could be grouped together with other columns, based on rule 1.

b) One-to-many. In the column-oriented data model, this relationship is actually implemented in the "opposite way" in comparison to the relational model. In the latter, the foreign keys (multiple values) are placed at the side of "many" of the relationship, since the database should meet 1-st and 2-nd normal form. However, in the column-oriented model, multiple values could be grouped together in a single column family. So, to create a one-to-many relationship, a new column family should be created at the side of "one", which could keep multiple values, pointing to the keys at the side of "many". In other words, in the column-oriented model, all the foreign keys are kept at the side of "one".

c) Many-to-many. Implementing this relationship is even easier in the column-oriented data model. In the relational model, many-to-many relationships are created by the help of a third (called "connection") table that keeps and combines the foreign keys of the other two tables. However, in the column-oriented model, a single column family can contain multiple values. So, to create a many-to-many relationship, a new column family needs to be created at each table, containing the values of the row keys of the other table. There is no need of a third table at all.

In the relational model, the relational database management system (RDBMS) guarantees data integrity and relationships integrity. However, in the column-oriented data model, this is not the case, so the application itself, which uses the data, should be designed to check, maintain and validate data integrity. This responsibility is moved from the DBMS to the software developer. As a result, it is possible to have data inconsistencies for a short period of time, until the next read.

IV. Conclusion

Data analysis is very important for the development of business organizations. It helps them to identify organizational bottlenecks, optimize business processes and increase efficiency and profits. In the era of big data, their analysis could not be done manually. There are powerful data analytical software tools, but they are either expensive or hard to deploy and requires multiple high-performance servers to run. Buying expensive hardware and software, and hiring high-qualified IT experts, is not affordable for all companies, especially for smaller ones and start-ups.

The architecture proposed in this article allows integration of a company's heterogeneous data (both database and filesystem) to a remote Hadoop infrastructure, providing powerful data analytical services on demand. This is an affordable and cost-effective cloud-based solution, suitable for any type company, from start-ups to large-size companies. Businesses are not required to buy any (expensive) hardware or software, but use the data analytical services on demand, paying a small processing fee per request or by subscription. The architecture could be easily implemented by developing just two data integration modules – an integration middleware or a wrapping service around the Hadoop data analytics services, at the service provider's side; and a graphical user interface (GUI) to work with the integration middleware, at the company's side.

The major difference between the proposed architecture and the existing approaches, reviewed in the related work, is that they offer an integration of a relational database management system to a non-relational one with a specific target data model. Instead, our architectural approach provides integration of data of any type (including relational, non-relational and even ordinary text files) to an existing Hadoop cluster. The purpose of this integration is to provide companies with cost-effective and easy to use access to the Hadoop's data analytical tools, not to permanently convert their data to other data models.

Since the proposed architecture will be very useful to the business organizations, providing them an easy access to powerful data analytical tools, and is fairly easy to deploy in the same time, our next goal is to implement it and prove its real usefulness and effectiveness. In general, it allows integration of any type of data to an existing Hadoop infrastructure. However, for this to happen, the source data model should be recognized and data should be parsed properly before sending them to the "Integration Middleware" module. For that reason, we plan to introduce data type documents (DTDs) that contain a formal definition of the source data models and the relevant parsers for these DTDs. Then the integration of a new, even custom, source data model will be quite easy – just by uploading the associated DTD. The use of DTDs highly increase flexibility of the proposed approach and allows integration of virtually any type of data.

ACKNOWLEDGMENT

This paper is supported by project 2022–EEA–01 "Analysis of big data processing algorithms and their application in multiple subject domains", funded by the Research Fund of the "Angel Kanchev" University of Ruse.

REFERENCES

[1] A. Holmes, Hadoop in practice, Manning Publications Co., 2015.

[2] B. Lublinsky, K. Smith, A. Yakubovich, Professional Hadoop solutions, John Wiley & Sons, Inc., 2015.

[3] G.A. Schreiner, D. Duarte and R.D. Mello, "When Relational-Based Applications Go to NoSQL Databases: A Survey", Information, 10, 241, 2019.

[4] G.A. Schreiner, D. Duarte and R.D. Mello, "Bringing SQL databases to key-based NoSQL databases: a canonical approach," Computing, vol. 102, pp. 221-246, 2019.

[5] J. Pokorný, "Integration of Relational and NoSQL Databases", Vietnam Journal of Computer Science, vol. 6, no. 4, pp. 389-405, 2019.

[6] A.Vathy-Fogarassy and T. Hugyák, "Uniform data access platform for SQL and NoSQL database systems", Information Systems, vol. 69, pp. 93-105, 2017.

[7] A.A. Frozza, E.D. Defreyn and R. dos Santos Mello, "An Approach for Schema Extraction of NoSQL Columnar Databases: the HBase Case Study", Journal of Information and Data Management, vol. 12, no. 5, pp. 384–395, 2021.

[8] J. Dai, "SQL to NoSQL: What to do and How", IOP Conference Series: Earth and Environmental Science, vol. 234, 2019.

[9] K. Herrmann, H. Voigt, T.B. Pedersen and W. Lehner, "Multi-schema-version data management: data independence in the twenty-first century," The VLDB Journal, vol. 27(4), pp. 547-571, 2018.

[10] M.J. Mior, K. Salem, A. Aboulnaga, R. Liu, "NoSE: Schema design for NoSQL applications," In Proc. of the 32nd International Conference on Data Engineering (ICDE), pp. 181–192, 2016.

[11] S, Ramzan, I.S. Bajwa, B. Ramzan, and W. Anwar, "Intelligent Data Engineering for Migration to NoSQL Based Secure Environments", IEEE Access, vol. 7, pp. 69042-69057, 2019.

[12] E.M. Kuszera, L.M. Peres and M.D. Fabro, "Toward RDB to NoSQL: transforming data with metamorfose framework", in Proc. of the 34th ACM/SIGAPP Symposium on Applied Computing, 2019.

[13] Y. Liao, J. Zhou, C. Lu, S. Chen, C. Hsu, W. Chen, M. Jiang and Y. Chung, "Data adapter for querying and transformation between SQL and NoSQL database", Future Generation Computer Systems, vol. 65, pp. 111-121, 2016.

[14] E.M. Kuszera, L.M. Peres, and M.D. Fabro, "Exploring data structure alternatives in the RDB to NoSQL document store conversion process", Information Systems, vol. 105, 2022.

[15] M. Kerzner, S. Maniyam, HBase Design Patterns, Packt Publishing Ltd., 2014.

[16] R. Choudhry, HBase High Performance Cookbook, Packt Publishing Ltd., 2017.

[17] L. Stanescu, M. Brezovan and D. Dan Burdescu, "Automatic mapping of MySQL databases to NoSQL MongoDB." In Proc. of 2016 Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 837-840, 2016.

[18] M.C.d. Freitas, D. Souza and A. Salgado, "Conceptual mappings to convert relational into NoSQL databases," in: Proc. of the 18th International Conference on Enterprise Information Systems, ICEIS, pp. 174–181, 2016.

[19] R. Ouanouki, A. April, A. Abran, A. Gomez and J.M. Desharnais, "Toward building RDB to HBase conversion rules," Journal of Big Data, vol. 4(1), pp. 1-21, 2017.

[20] J. Yoo, K. Lee and Y. Jeon, "Migration from RDBMS to NoSQL using column-level denormalization and atomic aggregates", Journal of Information Science & Engineering, vol. 34(1), pp. 243-259, 2018.

Method for Estimation of Oleic Acid Content in Soy Plants using Green Band Data of Sentinel-2/MSI

Kohei Arai[1]
Information Science Dept.
Saga University, Saga City, Japan

Yoshitomo Hideshima[2], Yuuhi Iwaki[3], Ryota Ito[4]
Saga Prefectural Agricultural Experiment Research Center
Planning and Coordination Department, Saga City, Japan

***Abstract*—A method for estimation of oleic acid content in soy plants using green band data of Sentinel-2/MSI: Multi Spectral Imager is proposed. Conventionally, vitality of agricultural plants is estimated with NDVI: Normalized Difference Vegetation Index. Spatial resolution of Near Infrared: NIR band of Sentinel-2/MSI for calculation of NDVI, however, is 20 m. Therefore, a method for estimation of vitality with only green band data of Sentinel-2/MSI is proposed here. Through regressive analysis with the satellite data as well as drone mounted NDVI camera data together with component analysis data by gas chromatography, it is found the correlation between NDVI and green band data of the optical sensor (MSI) onboard Sentinel-2 as well as the component analysis data. It is also found that the new variety of soy plant, Saga University brand: HO1 contains about 50% much oleic acid in comparison to the conventional variety of soy plant, Fukuyutaka.**

***Keywords*—Oleic acid; NDVI (normalized difference vegetation index); regressive analysis; soy plant; Multi Spectral Imager: MSI**

I. INTRODUCTION

Oleic acid is a type of unsaturated fatty acid and is a monovalent unsaturated fatty acid. It is abundant in animal fats, vegetable oils, and nuts and seeds. Since it reduces only bad cholesterol without reducing good cholesterol, it helps prevent lifestyle-related diseases such as arteriosclerosis and heart disease. In addition, it is resistant to heat and is difficult to oxidize, suppresses fatty acids peroxide that may cause carcinogenesis, regulates gastric acid secretion, and activates intestinal motility.

In addition to amino acids, meat contains various health ingredients. Examples include "heme iron," which has an anemia-preventing effect, and "conjugated linoleic acid," which has anti-cancer effects, body fat reduction, and arteriosclerosis-preventing effects. Among them, "oleic acid", which is one of the fatty acids, is expected to have a preventive effect on lifestyle-related diseases.

Oleic acid, which is abundant in beef and pork, has the function of maintaining proper cholesterol in the blood. Cholesterol includes bad (LDL) cholesterol and good (HDL) cholesterol. Bad cholesterol causes arteriosclerosis, heart disease, and high blood pressure, but good cholesterol has a function to prevent arteriosclerosis. Oleic acid is said to have the effect of reducing only bad cholesterol without reducing this good cholesterol.

Another characteristic of oleic acid is that it is not easily oxidized. In general, when fat is oxidized, it binds to active oxygen in the body and damages DNA, causing cancer, arteriosclerosis, heart disease, brain disease, diabetes, and so on. Oleic acid has the effect of preventing cancer and lifestyle-related diseases.

The main varieties of Saga Prefecture are cultivated as standard, and the crops of the year are estimated by analyzing the growth reaction in climate change, and they are used as materials for soybean production guidance.

The new varieties were cultivated by Saga University, a national university corporation, and the application for cultivar registration was announced in 2019. The variety name is "Saga University HO1", which has a high oleic acid content, and "Fukuyutaka" has a high content of 80%, while 20% of the total fatty acid. In recent years, there have been requests from actual consumers and producers to expand cultivation in the midst of regional specialization, and there are voices in the Takeo / Kishima area requesting the establishment of technology for stable production. Therefore, the purpose is to obtain knowledge about the basic cultivation characteristics of this variety.

"High oleic acid soybean (Saga University HO1)" has a flowering period of 2 to 3 days and a maturity period of about 1 day earlier than "Fukuyutaka". The length of the main stem is a little short, and the height of the lowermost pod node is about 3 cm lower. The number of main stem nodes and the number of branches is about the same. There are many pods. Yieldability is considered to be about the same. Grains tend to be slightly flat and small in grain shape.

Relating to soy plant monitoring and content estimation in soy, there are the following papers.

Prediction of Isoflavone Content in Soybeans with Sentinel-2 Optical Sensor Data by Means of Regressive Analysis [1], and Soybean Quality Estimation with Sentinel-2 of Optical Sensor Data [2]. On the other hand, Sentinel-2 data analysis method related papers, there are the following papers.

Flooding and oil spill disaster relief using Sentinel of remote sensing satellite data [3], Sea breeze damage estimation method using Sentinel of remote sensing satellite data [4], ACIX-Aqua: A global assessment of atmospheric correction methods for Landsat-8 and Sentinel-2 over lakes, rivers, and coastal waters [5], Method for Most Appropriate Plucking Date Determination based on the Elapsed Days after Sprouting with NIR Reflection from Sentinel-2 Data [6], Prediction of

Isoflavone Content in beans with Sentinel-2 Optical Sensor Data by Means of Regressive Analysis [7].

Furthermore, agricultural plantation related research works, there are the following papers.

Regressive analysis on leaf nitrogen content and near infrared reflectance and its application to agricultural farm monitoring with helicopter mounted near infrared camera [8], Effect of sensitivity improvement of visible to near infrared digital cameras on NDVI measurement in particular for agricultural field monitoring [9], Smartphone image based agricultural product quality and harvest amount prediction method [10], A computer aided system for tropical leaf medicinal plant identification [11], Product amount and quality monitoring in agricultural fields with remote sensing satellite and radio-control helicopter [12], Intelligent System for Agricultural Field Monitoring [13], Multi-level observation system for agricultural field monitoring [14], Multi-Layer Observation for Agricultural (Tea and Rice) Field Monitoring [15], Bigdata Platform for agricultural field monitoring and environmental monitoring [16].

Prediction of isoflavone content in beans with Sentinel-2 optical sensor data by means of regressive analysis is made [17]. Meanwhile, method for most appropriate plucking date determination based on the elapsed days after sprouting with NIR reflection from Sentinel-2 data is proposed and validated [18].

The purpose of this study is to establish a method for estimation of oleic acid content in soy plants using NDVI derived from optical sensor (MSI) onboard Sentinel-2 satellite. Traditionally, it is used to use drone mounted NDVI data for the estimation. It does cost. It cannot be used for the windy condition. Turns out, Sentinel-2/MSI data can be accessible from the ESA web site freely even for windy condition, it cannot be used for cloudy and rainy conditions, though. In terms of spatial resolution of drone mounted NDVI camera, it is fine about a couple of cm depending on altitude of the drone. On the other hand, spatial resolution of Sentinel-2/MSI is 10 m for visible bands and is 20 m for Near Infrared: NIR band. Therefore, spatial resolution of NDVI which is derived with visible and NIR is 20 m. It is not good enough for Japanese soy plantation fields. In order to estimate NDVI with 10 m of spatial resolution using Sentinal-2/MSI optical sensor data, green band data derived from visible band of Sentinel-2/MSI is used in the proposed method. Through regressive analysis with the satellite data as well as drone mounted NDVI camera data together with component analysis data by gas chromatography, it is found the correlation between NDVI and green band data of the optical sensor (MSI) onboard Sentinel-2 as well as the component analysis data. It is also found that the new variety of soy plant, Saga University brand: HO1 contains about 50% much oleic acid in comparison to the conventional variety of soy plant, Fukuyutaka.

The next section describes the proposed method. Then experiments with drone mounted NDVI data and Sentinal-2/MSI of green band data together with component analysis data by gas chromatography followed by concluding remarks and future research work.

II. Proposed Method

A method for estimation of oleic acid content in soy plants using NDVI derived from optical sensor onboard Sentinel-2 satellite is proposed. Table I shows Sentinel-2 fast track service compliance to land user requirements. Every 10 days, Sentinel-2 observes the same location of Earth’s surface.

Major specification of Multispectral Imager: MSI is as follows, MSI covering 13 spectral bands (443–2190 nm), with a swath width of 290 km and a spatial resolution of 10 m (four visible and near-infrared bands), 20 m (six red edge and shortwave infrared bands) and 60 m (three atmospheric correction bands).

MSI spatial resolution versus wavelength: Sentinel-2’s span of 13 spectral bands, from the visible and the near-infrared to the shortwave infrared at different spatial resolutions ranging from 10 to 60 m on the ground, takes land monitoring to an unprecedented level. Table II shows MSI spectral band specification. Green band in the MSI spectral bands can be used for estimation of oleic acid containing soybeans in living soy plantations.

The spatial resolution of Sentinel-2 is 10 m for visible bands and is 20 m for Near Infrared: NIR band (865 nm). Therefore, spatial resolution of NDVI which is derived with visible and NIR is 20 m. It is not good enough for Japanese soy plantation fields. To estimate NDVI with 10 m of spatial resolution using Sentinal-2 optical sensor data, green band data derived from visible band of Sentinel-2 is used in the proposed method.

TABLE I. SENTINEL-2 FAST TRACK SERVICE COMPLIANCE TO LAND USER REQUIREMENTS

Fast Track Service (Land Monitoring Core Services)	Compliance of the Sentinel-2 system
Geographic coverage	All land areas/islands covered (except Antarctica)
Geometrical revisit	5 days revisit cloud free fully in line with vegetation changes
Spectral sampling	Unique set of measurement/calibration bands
Service continuity	Sentinel-2A launch in 2014: the mission complements the SPOT and Landsat missions.
Spatial resolution	< 1 ha MMU (Minimum Mapping Unit) fully achievable with 10 m
Acquisition strategy	Systematic push-broom acquisitions, plus lateral mode capability for emergency events monitoring
Fast Track Service (Emergency Response Core Service)	Compliance of the Sentinel-2 system
Spatial resolution down to 5 m	Reference/damage assessment maps limited to the 10m SSD (Spatial Sampling Distance)
Accessibility/timeliness down to 6 hrs offline & 24hrs in NRT	Fully compliant (retrieval of already archived reference data in < 6 hrs, and delivery of data after request in NRT in 3 hrs for L1c)

TABLE II. MSI SPECTRAL BAND SPECIFICATION

Spectral bands (center wavelength in nm/SSD in m)	Mission objective	Measurement or calibration
B1 (443/20/60), B2 (490/65/10) & B12 (2190/180/20)	Aerosol correction	Calibration bands
B8 (842/115/10), B8a (865/20/20), B9 (940/20/60)	Water vapor correction	
B10 (1375/20/60)	Circus detection	
B2 (490/65/10), B3 (560/35/10), B4 (665/30/10), B5 (705/15/20), B6 (740/15/20), B7 (775/20/20), B8 (842/115/10), B8a (865/20/20), B11 (1610/90/20), B12(2190/180/20)	Land cover classification, Leaf chlorophyll content, leaf water content, LAI, fAPAR, snow/ice/cloud, mineral detection.	Land measurement bands

The online services provided by Amazon for web operators and developers are called AWS (Amazon Web Services). Among them, a large amount of data taken by Sentinel-2 is stored in Amazon's cloud storage called Amazon S3 (Amazon Simple Storage Service), and his AWS service that provides the data taken by Sentiel-2 is Sentinel- It's called 2 on AWS. Users can now download satellite imagery data via Amazon S3 storage. Amazon publishes a huge amount of geospatial data in a project called Earth on AWS, including not only satellite images, but also terrain tiles, SpaceNet machine learning training data, next-generation weather radar (NEXRAD) weather data, and more. It also provides a variety of geospatial data and is accessible to anyone. With a paid service, you can analyze this huge amount of data on the cloud, so there is no need to download data for analysis or perform troublesome preprocessing, and it is a larger project than before. It can be realized in a shorter period of time.

Download satellite images of where you need them. Access the Sentinel Image browser. The search screen is displayed. Specify the period to search for the image and the amount of clouds, display the area where you want to download the image, and click the [Submit] button to display the detailed information and preview of the image, the amount of clouds, the date and time when the image was acquired, the product level, etc. can be confirmed. Select the image you want to use from the search result field and click the link button. Then it can be gotten a link to the image, click AWS path (link above). After that, Launch ArcGIS Pro (or ArcMap) and add the downloaded data from the Project window (in the case of ArcMap, add it from the Catalog window). This time, we will synthesize a multi-spectral band, so add B02.jp2 to B04.jp2. Since it is added for each band, it does not look natural as it is. Next, composite each band to synthesize the true color. Click the Raster Function button on the Analysis tab to bring up the Raster Function window. Expand Data Management and click the Composite icon. Click the Raster pull-down button under Composite Properties to add each band. Add all the bands, sort them by B04, B03, B02 from the top, and click the Create New Layer button. A "Composite" layer has been added to the Content window to composite the true color multiband image.

III. EXPERIMENT

A. Intensive Study Area

Intensive study area is situated at the Saga Prefectural Agriculture Research Institute: SPARI which is situated at 33°13'11.5" North, 130°18'39.6" East as shown in Fig. 1. Fig. 1(a) shows the intensive study area in Japan while Fig. 1(b) shows the soy plantation areas for the new variety of soy (Saga University originated HO1) and for the conventional variety of soy (Fukuyutake). Both soy plantation areas are indicated in red and white rectangles in Fig. 1.

(a) Intensive Study site in Japan (Red Circle).

(b) Red Rectangle: Conventional Fukuyutaka, White Rectangle: New Variety of HO1.

Fig. 1. Intensive Study Area.

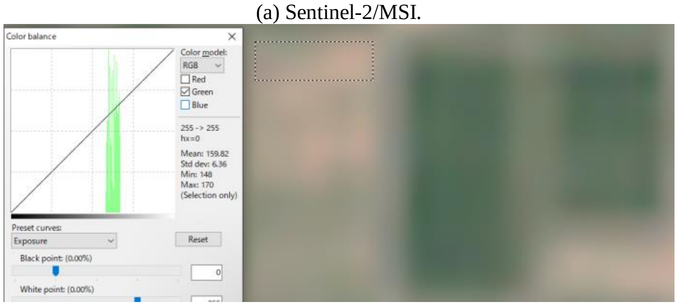
B. Acquired Drone Mounted NDVI Camera Data and MSI Imagery Data of Sentinel-2

Examples of the drone mounted NDVI camera data and Sentinel-2/MSI imagery data which are acquired on 29 July 2021 and on 30 July 2021, respectively are shown in Fig. 2. Fig. 2(a) shows Sentinel-2/MSI image which is acquired on 30 July 2021 while Fig. 2(b) shows the histogram of the green band data of the Fukuyutaka field of the Sentinel-2/MSI image. On the other hand, Fig. 2(c) shows the histogram of the green band data of the HO1 field of the Sentinel-2/MSI image. Meanwhile, Fig. 2(d) shows the natural color image of drone mounted NDVI camera data which is acquired on 29 July 2021 while Fig. 2(e) shows the histogram of the NDVI image of drone mounted NDVI camera data of the Fukuyutaka. On the other hand, Fig. 2(f) shows the histogram of the NDVI image of drone mounted NDVI camera data of the HO1.

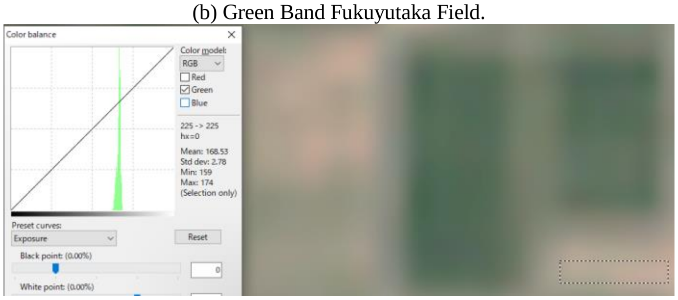
C. Relation between Sentinel-2/MSI and Drone Mounted NDVI Camera Data

The trend of the means of the Sentinel-2/MSI data of the Fukuyutaka and the Saga University originated HO1 of new variety of soy fields is shown in Fig. 3(a) while Fig. 3(b) shows both trends of Sentinel-2/MSI and drone mounted NDVI camera data of the same soy fields. Although mean and standard deviation are different each other, trends are almost same. Therefore, the green band data of the Sentinel-2/MSI can be used for growing process analysis of soy fields.

(a) Sentinel-2/MSI.

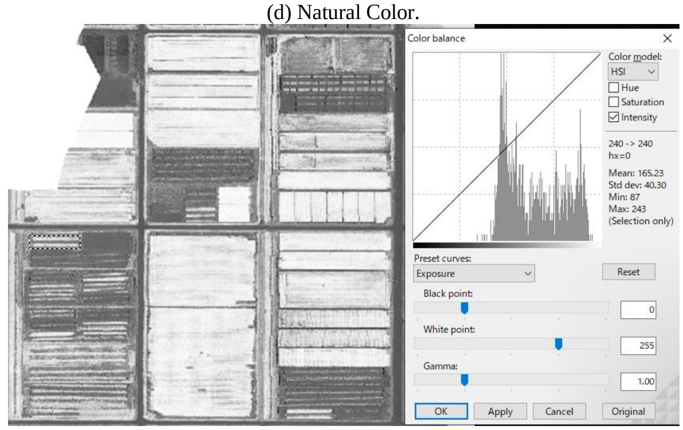


(b) Green Band Fukuyutaka Field.

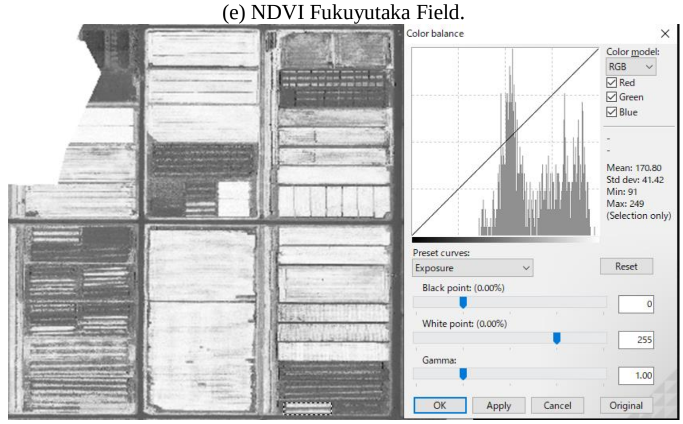


(c) Green Band HO1 Field.

(d) Natural Color.

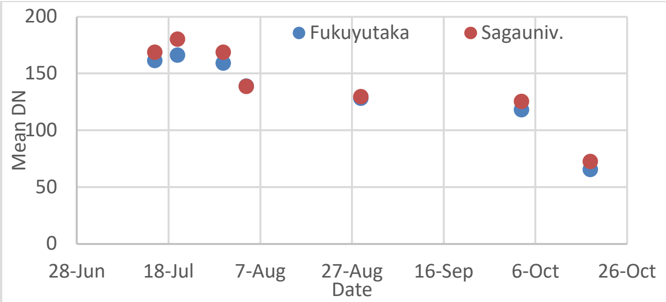


(e) NDVI Fukuyutaka Field.

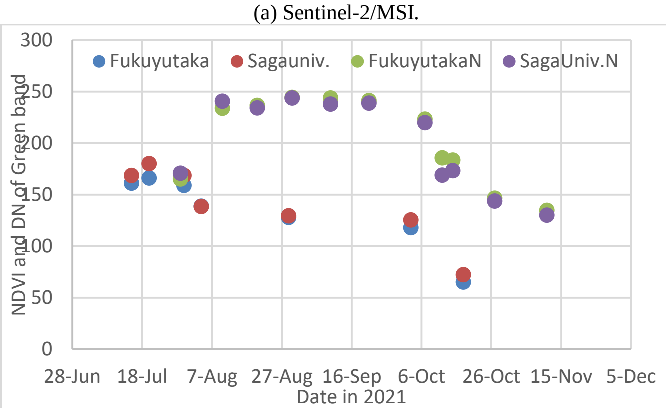


(f) NDVI HO1.

Fig. 2. Examples of the Drone Mounted NDVI Camera Data and Sentinel-2/MSI Imagery Data which are Acquired on 29 July 2021 and on 30 July 2021, respectively.



(a) Sentinel-2/MSI.



(b) Both Sentinel-2/MSI and Drone Mounted NDVI Camera.

Fig. 3. Relation between Sentinel-2/MSI and Drone Mounted NDVI Camera Data: The Trend of the means of the Sentinel-2/MSI Data of the Fukuyutaka and the Saga University Originated HO1 of New Variety of Soy Fields.

D. Experiment Procedure

Field experiments are conducted as follows:

Tillage overview:

1) Field: C4-10 (previous crop: wheat, previous summer crop: paddy rice).

2) Test variety: High oleic acid soybean "Saga University HO1" (Control: "Fukuyutaka")

3) Sowing method: After plowing and leveling, hand-sown 4 seeds per plant → 2 tailoring (Hirataka ridge)

4) Weeding: After sowing and before budding) Laxar granules 4kg / 10a (Growth period) Weeding as appropriate

5) Fertilization: No fertilization

Composition of test plot is as follows:

Ward sowing period Planting style:

1) Early sowing practice zone Late June 75 cm between rows * 20 cm between stocks * 2 (13.3 / ㎡)

2) Early sowing sparsely planted area Late June Joma 75 cm * Strain 30 cm * 2 (8.9 shares / ㎡)

3) Conventional sowing area Mid-July 75 cm between rows * 20 cm between stocks * 2 (13.3 / ㎡)

4) Control area (Fukuyutaka) Mid-July 75 cm between rows * 20 cm between stocks * 2 (13.3 / ㎡)

Survey items are as follows:

Understanding the growth stage (budding stage, flowering stage, pod setting stage, grain enlargement stage, yellow leaf stage, leaf fall stage, maturity stage).

Seedling survey is as follows:

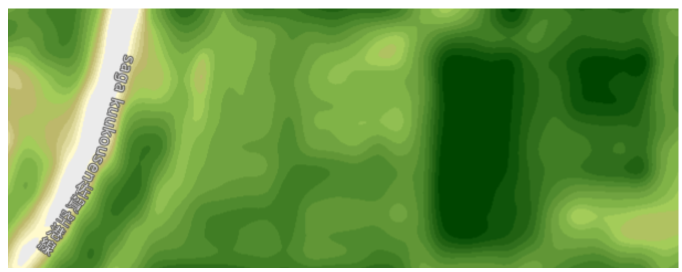
1) Yield survey (grain weight, stem weight, 100 grain weight, particle size composition).

2) Decomposition survey (main stem length, number of main stem nodes, bottom pod height, number of branches, number of pods, number of grains).

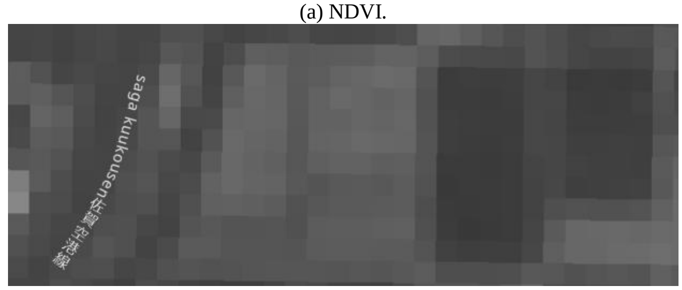
3) Quality survey (inspection grade, protein content).

E. Relation between Sentinel-2/MSI Derived NDVI and Green Band Data

Relation between Sentinel-2/MSI derived NDVI and green band data is investigated. If the relation between Sentinel-2/MSI derived NDVI and the green band data is high, then vegetation vitality can be estimated with green band only. Fig. 4(a) shows Sentinel-2/MSI derived NDVI while Fig. 4(b) shows green band data of the intensive study area acquired on July 30 2021.



(a) NDVI.



(b) Green Band.

Fig. 4. Comparison between Sentinel-2/MSI Derived NDVI and Green Band Data of the Intensive Study Area Acquired on July 30 2021.

The correlation coefficient between both is more than 0.95. Therefore, it is concluded that vegetation vitality can be estimated with green band only.

F. Comparisons among the Several Band Combinations of Sentinel-2/MSI Data

Fig. 5(a) to (d) show the following several band combinations of Sentinel-2/MSI imagery data.

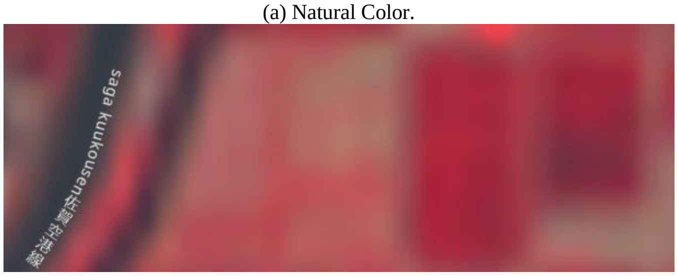
a) Natural Color.

b) False Color.

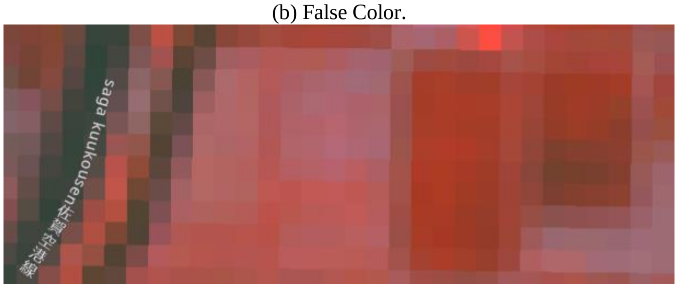
c) R: Band 8, G: Band 3, B: Band 4 and.

d) NDVI.

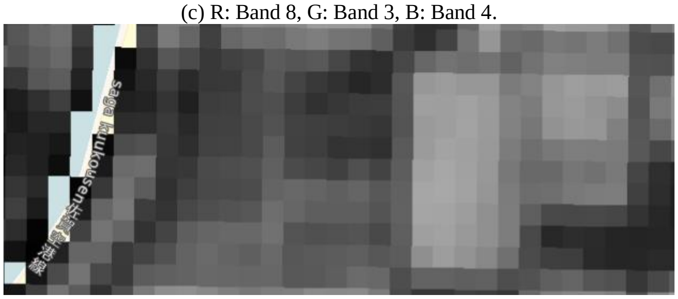
(a) Natural Color.



(b) False Color.



(c) R: Band 8, G: Band 3, B: Band 4.



(d) NDVI.

Fig. 5. Several Band Combinations of Sentinel-2/MSI Imagery Data.

Fig.6 shows the relation between the NDVI (Fig. 5(d)) and the green band data (Fig. 4(b)). Correlation coefficient between both is 0.977 and the NDVI is estimated with the green band data with regressive equation in Fig. 6 with R^2 value of 0.954.

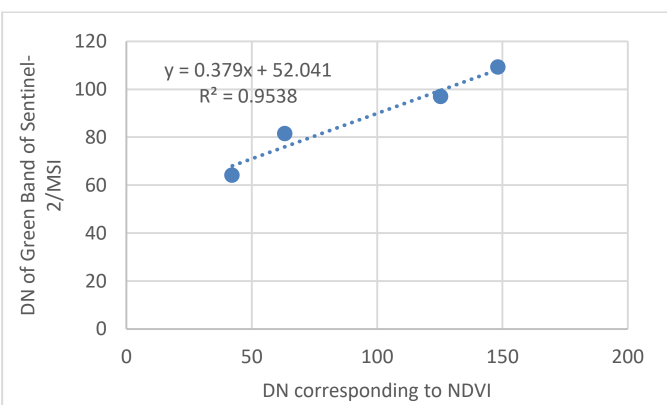


Fig. 6. Correlation between NDVI and Green Band Data.

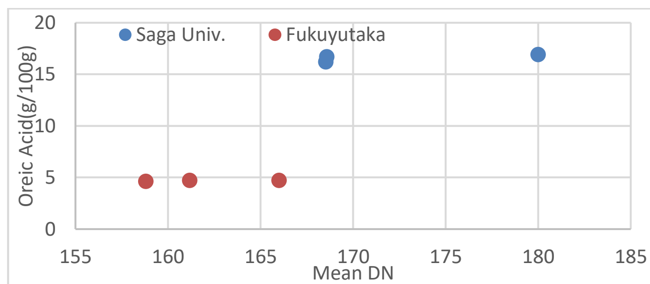
Through a comparison between Fig. 4(b) and Fig. 5(d), it is confirmed that vegetation vitality can be estimated with green band only.

G. *Experimental Results*

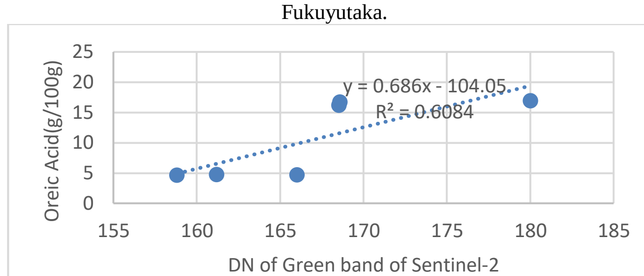
Oleic acid content in the harvested soybeans is measured with gas chromatography. Fig. 7(a) shows the results from the gas chromatography measurements for Saga University originated HO1 and conventional Fukuyutaka against mean of digital Number: DN of the Sentinel-2/MSI of the green band data for each test site. On the other hand, Fig. 7(b) shows the relation between oleic acid content and mean of the green band data derived from Sentinel-2/MSI. It is found that the R^2 value of the relation between oleic acid content and mean of the green band data derived from Sentinel-2/MSI is much more than 0.6. Therefore, it can be said that it is possible to estimate oleic acid content by using the mean of the green band data derived from Sentinel-2/MSI.

The relationship between sowing time, number of repetitions and amount of oleic acid) for Fukuyutaka is shown in Fig. 8 while effect of n-piece stands on Saga University originated HO1 variety (sown on June 28) is shown in Fig. 9, respectively. In this connection, the number of iterations is defined as follows:

For those with a small number of iterations, the number of iterations had to be reduced because the field was small. The difference in the number of iterations is meaningless.



(a) Relation between Oleic Acid Content and mean od DN of the Green Band Data of Sentinel-2/MSI for the Saga University Originated HO1 and Fukuyutaka.



(b) Relation between Oleic Acid Content and Mean of the Green Band Data Derived from Sentinel-2/MSI.

Fig. 7. Relation between Oleic Acid Content and Mean of the Green Band Data Derived from Sentinel-2/MSI.

On the other hand, the number of tailoring is defined as follows:

To be exact, it was "difference between strains (difference in planting density)" rather than the number of tailored plants. The correct treatment for "Saga University originated HO1" is that the seeding on June 30 is 30 cm (13.3 seeds / ㎡) and the seeds are 20 cm (8.9 seeds / ㎡), and the seeding on July 13 is 20 cm between the stocks. Regarding the planting density, we have set two styles that can be handled by mechanical sowing at the site in order to examine the appropriate planting density that enables stable production.

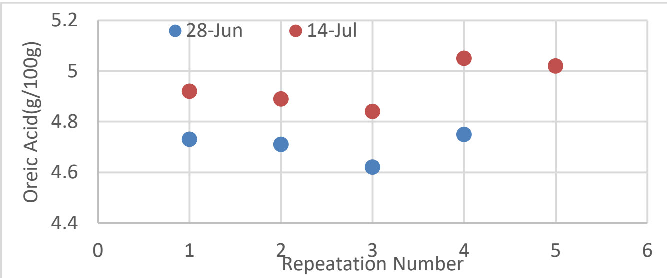


Fig. 8. Relationship between Sowing Time, Number of Repetitions and Amount of Oleic Acid) for Fukuyutaka.

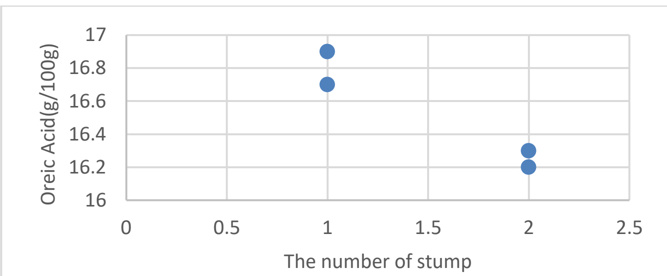


Fig. 9. Effect of n-piece Stands on Saga University Originated HO1 Variety (Sown on June 28).

Meanwhile, regressive analysis for estimation of harvest number of soybeans is made. Fig. 10 shows the result. It is found that harvest amount can be predicted with DN of green band data of Sentinel-2/MSI data about three months before the harvest with R^2 value of 0.965.

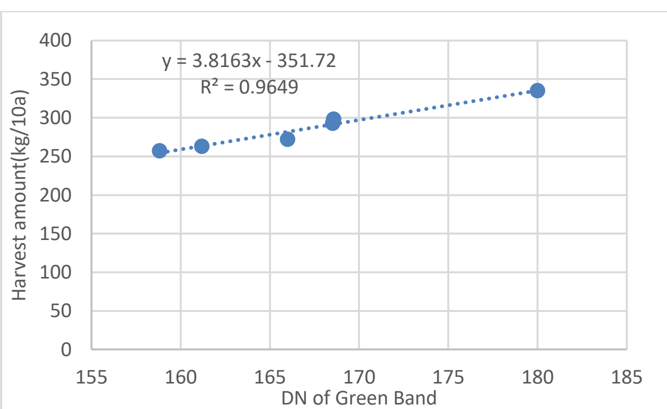


Fig. 10. Harvest Amount Prediction with Green Band Data of Sentinel-2/MSI.

IV. Conclusion

A method for estimation of oleic acid content in soy plants using NDVI derived from optical sensor onboard Sentinel-2 satellite is proposed. Through regressive analysis with the satellite data as well as drone mounted NDVI camera data together with component analysis data by gas chromatography, it is found the correlation between NDVI and green band data of the optical sensor (MSI) onboard Sentinel-2 as well as the component analysis data. It is also found that the new variety of soy plant, Saga University brand: HO1 contains about 50% much oleic acid in comparison to the conventional variety of soy plant, Fukuyutaka.

It is found that the R^2 value of the relation between oleic acid content and mean of the green band data derived from Sentinel-2/MSI is much more than 0.6. Therefore, it can be said that it is possible to estimate oleic acid content by using the mean of the green band data derived from Sentinel-2/MSI.

The limitation of the proposed method is caused by the observation frequency of fine Sentinel-2/MSI data. The revisit cycle of Sentinel-2 satellite is 10 days. It is not possible to get fine data when the weather condition is not fine (cloudy and rainy).

V. Future Reaesrch Work

Next thing we would like to do is estimation of most appropriate harvest date with Sentinel-2/MSI data. Also, deep learning is applied to these estimations.

Acknowledgment

The authors would like to thank to Professor Dr. Hiroshi Okumura and Professor Dr. Osamu Fukuda for their valuable discussions.

References

[1] Kohei Arai1, Osamu Shigetomi, Hideki Ohtsubo, Eri Ohya, Prediction of Isoflavone Content in Soybeans with Sentinel-2 Optical Sensor Data by Means of Regressive Analysis, Proceedings of the IntelliSys Conference 2021, 2021.

[2] Kohei Arai, Soybean Quality Estimation with Sentinel-2 of Optical Sensor Data, General Lecture of Postgraduate Program on Computer Science IPB "Satellite Image Processing for Smart Agriculture and Forestry" Wednesday 20 October 2021, Time: 10.00 – 12.00 WIB (Jakarta time).

[3] Kohei Arai, Flooding and oil spill disaster relief using Sentinel of remote sensing satellite data, International Journal of Advanced Computer Science and Applications IJACSA, 10, 12, 290-297, 2019.

[4] Kohei Arai, Sea breeze damage estimation method using Sentinel of remote sensing satellite data, International Journal of Advanced Computer Science and Applications IJACSA, 11, 2, 40-47, 2020.

[5] Nima Pahlevan, Antoine Mangin; Sunadarabalan Balasubramanian; Brandon Smith; Krista Alikas; Claudio Barbosa; Simon Bélanger; Caren Binding; Mariano Bresciani; Claudia Giardino; Daniela Gurlin; Yongzhen Fan; Tristan Harmel; Moritz K Lehmann; Ronghua Ma; Leif Olmanson; Natascha Oppelt; Steef Peters; Nathalie Reynaud; Lino A Sander de Carvalho; Evangelos Spyrakos; François Steinmetz; Kerstin Stelzer; Sindy Sterckx; Thierry Tormos; Quinten Vanhellemont; Mark Warren; Martin Ligi; Kohei Arai; Peter Hunter; Joji Ishikaza; Susanne Kratzer; Andrew Tyler; Yanqun Pan, ACIX-Aqua: A global assessment of atmospheric correction methods for Landsat-8 and Sentinel-2 over lakes, rivers, and coastal waters, Journal: Remote Sensing of Environment, 258 (2021) 112366, 1-22, RSE-D-20-01482R1, 2021.

[6] Kohei Arai, Yoshiko Hokazono, Method for Most Appropriate Plucking Date Determination based on the Elapsed Days after Sprouting with NIR Reflection from Sentinel-2 Data, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 12, No. 4, 22-29, 2021.

[7] Kohei Arai, Prediction of Isoflavone Content in beans with Sentinel-2 Optical Sensor Data by Means of Regressive Analysis, Proceedings of SAI Intelligent Systems Conference, IntelliSys 2021: Intelligent Systems and Applications pp 856-865, 2021.

[8] Kohei Arai, Sadayuki Akaishi, Hideo Miyazaki, Yasuyuki Watabe, Takaki Yoshida, Regressive analysis on leaf nitrogen content and near

infrared reflectance and its application to agricultural farm monitoring with helicopter mounted near infrared camera, International Journal of Advanced Research in Artificial Intelligence, 2, 3, 38-43, 2013.

[9] Kohei Arai, Takuji Maekawa, Toshihisa Maeda, Hiroshi Sekiguchi, Noriyuki Masago, Effect of sensitivity improvement of visible to near infrared digital cameras on NDVI measurement in particular for agricultural field monitoring, International Journal of Advanced Research on Artificial Intelligence, 4, 12, 1-8, 2015.

[10] Kohei Arai, Osamu Shigetomi, Yuko Miura, Satoshi Yatsuda, Smartphone image based agricultural product quality and harvest amount prediction method, International Journal of Advanced Computer Science and Applications IJACSA, 10, 9, 24-29, 2019.

[11] Yeni Herdiyeni, Kohei Arai, Ellyn Danayanti, Ervizal A.M. Zaherdi, A computer aided system for tropical leaf medicinal plant identification (2013), Proceedings of the 1st International Conference on Sustainable Agriculture, Food and Energy, SAFE 2013.

[12] Kohei Arai, Product amount and quality monitoring in agricultural fields with remote sensing satellite and radio-control helicopter, Proc. of the 40th Science Assembly of the COSPAR (Solicited), Advanced RS Methods and Technics, ID12566, A3.1-0041,2014.

[13] Kohei Arai, Intelligent System for Agricultural Field Monitoring, Proceedings of the SAI Intelligent Systems Conference 2016.

[14] Kohei Arai, Multi-level observation system for agricultural field monitoring, Proceedings of the 1st International Conference on Sustainable Information, Engineering and Technology, Keynote Speech, 2016.

[15] Kohei Arai, Multi-Layer Observation for Agricultural (Tea and Rice) Field Monitoring, Proceedings of the Seminar at Bogor Agriculture University, Keynote Speech, 2016.

[16] Kohei Arai, Bigdata Platform for agricultural field monitoring and environmental monitoring, Proceedings of the 4th LISAT Symposium (Invited Speech), p.37, 2017.

[17] Kohei Arai, Prediction of Isoflavone Content in beans with Sentinel-2 Optical Sensor Data by Means of Regressive Analysis, Proceedings of SAI Intelligent Systems Conference, IntelliSys 2021: Intelligent Systems and Applications pp 856-865, 2021.

[18] Kohei Arai, Yoshiko Hokazono, Method for Most Appropriate Plucking Date Determination based on the Elapsed Days after Sprouting with NIR Reflection from Sentinel-2 Data, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 12, No. 4, 22-29, 2021.

Authors' Profile

Kohei Arai, He received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science and Technology of the University of Tokyo from April 1974 to December 1978 also was with National Space Development Agency of Japan from January, 1979 to March, 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post Doctoral Fellow of National Science and Engineering Research Council of Canada. He moved to Saga University as a Professor in Department of Information Science on April 1990. He was a councilor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was a councilor of Saga University for 2002 and 2003. He also was an executive councilor for the Remote Sensing Society of Japan for 2003 to 2005. He is a Science Council of Japan Special Member since 2012. He is an Adjunct Professor of University of Arizona, USA since 1998. He also is Vice Chairman of the Science Commission "A" of ICSU/COSPAR since 2008 then he is now award committee member of ICSU/COSPAR. He wrote 55 books and published 620 journal papers as well as 450 conference papers. He received 66 of awards including ICSU/COSPAR Vikram Sarabhai Medal in 2016, and Science award of Ministry of Mister of Education of Japan in 2015. He is now Editor-in-Chief of IJACSA and IJISA. http://teagis.ip.is.saga-u.ac.jp/index.html

Portable ECG Monitoring System

Zhadyra N. Alimbayeva[1], Chingiz A. Alimbayev[2], Nurlan A. Bayanbay[3]
Kassymbek A. Ozhikenov[4], Oleg N. Bodin[5], Yerkat B. Mukazhanov[6]
Satbayev University, Almaty, The Republic of Kazakhstan[1, 3, 4]
U.Joldasbekov Institute of Mechanics and Engineering, Almaty, The Republic of Kazakhstan[2]
Penza State Technological University, Penza, Russia[5]
Zhetysu University, Taldykorgan, The Republic of Kazakhstan[6]

***Abstract*—The number of patients with cardiovascular diseases (CVD) is rapidly increasing in the world. Many CVDs are likely to manifest their symptoms some time prior to the onset of any adverse or catastrophic events, and early detection of cardiac abnormalities is incredibly important. To reduce the risks of life-threatening arrhythmia, it is necessary to develop and introduce portable systems for monitoring the state of the heart in conditions of free activity. This paper presents the second generation (prototype) of a portable cardiac analyzer and the developed system for non-invasive cardiac diagnostics. The portable cardiac analyzer mainly consists of an ADC for taking an electrocardiosignal (ECS) and an STM32L151xD microcontroller. To record operational data on current ECS, a block of non-volatile high-speed memory MRAM is connected to the microcontroller. A communication unit is based on the universal combo module SIM868 from SIMCOM, which supports data exchange in GSM/GPRS networks. The developed ECG monitoring system allows making decisions at different levels (cardiac analyzer, server, doctor), as well as exchanging information necessary to ensure an effective diagnostic and treatment process. We evaluated the performances of the developed system. The signal-to-noise ratio of the output signal is favorable, and all the features needed for a clinical evaluation (P waves, QRS complexes and T waves) are clearly readable.**

Keywords—Electrocardiography; portable ECG device; ECG monitoring systems; cardiovascular diseases; mobile healthcare

I. Introduction

Over the past ten years, the incidence rate of cardiovascular diseases in Kazakhstan has increased by 1.7 times. The data collected by scientists shows a four-fold increase in hospitalization due to chronic heart failure (CHF) compared with a period of 20 years ago. Of the 13 million adults in Kazakhstan 350 000 citizens are diagnosed with CHF [1]. The problem of combating cardiovascular diseases (CVD) among the population is gaining national importance due to high morbidity, a high level of disability and mortality from CVD, long-term, often lifelong, drug treatment, as well as its high cost, dictates the need to pay increasing attention to early primary prevention these diseases.

One of the areas of CVD diagnostics, which has become relevant due to rapid technological progress, is monitoring the state of the heart in conditions of free activity. Modern technologies have made it possible to develop miniature wearable devices for recording the functional parameters of a person operating in free activity conditions. The use of such devices required a deeper and more detailed study of the means and algorithms for processing the electrocardiosignal.

This work is a continuation of our previous work [2], in which we are provided the initial experimental results of the device for diagnosing heart conditions. During the initial layout of the first test version, a microcontroller in a TQFP-100 package and an 8-channel ADC were selected. However, after manufacturing the first version of the mobile ECG device and optimizing the circuit diagram, it became clear that a microcontroller in a TQFP-64 package and four ADC input channels were enough for the device. Therefore, the version of the microcontroller in the TQFP-64 package and the 4-channel version of the ADC were used in the development of the second generation of the ECG device.

In this prototype of the device, a proprietary communication unit was developed; it was implemented on the SIM868 universal combo module from SIMCOM, which supports data exchange in GSM / GPRS cellular networks, reception of location data and exact time signals from navigation satellites (GNSS).

The advantage of the developed hardware platform of a portable ECG device over existing similar solutions is that the platform included many technological solutions that allow development and modernization of the software part of the device without changing the hardware of the platform. A photo of the manufactured sample of the device is shown in Fig. 1.

Fig. 1. Photo of a Prototype Portable ECG Device.

Software for a mobile application has been developed to receive a signal from the recording device, performing its preliminary processing, diagnosing life-threatening cardiac arrhythmias, notifying the patient about the diagnostic results and transmitting data to the application server.

The main features of the system in comparison with existing are:

- work in conditions of free activity;
- performing a preliminary ECG analysis;
- automatic diagnosis of life-threatening arrhythmias;
- call an ambulance;
- storing ECS entries locally and on a remote server; and
- ability to integrate with the medical information system.

Our paper is organized as follows. Section 2 presents the state of the art. Section 3 describes of development of circuitry and printed circuit board of the ECG device. Section 4 details the software implementation of the system. Section 5 yields the discussion and future work directions. Finally, Section 6 outlines the main conclusions.

II. Literature Review

We reviewed portable systems and wearable devices designed to detect or predict cardiovascular diseases over the past 2-3 years, since the achievements in this area until 2019 were given in our previous publication [2].

The paper [3] presents a portable, wearable, low-power non-contact ECG monitoring device that helps in the early detection of cardiovascular diseases. The device is placed in a shirt pocket, from where it will transmit data via Bluetooth Low Energy (BLE) to the user's mobile phone. The device mainly consists of three non-contact electrodes for sensing the cardio signal, an AD8233 AFE chip for extracting the ECG signal, and a CC2650 microcontroller for reading, filtering and transmitting them. A study [4] described and evaluated a new transducer design and system for the "invisible" ECG. Designed in the shape of a toilet seat, it provides device-free femoral ECG data, bringing a new approach to automated comprehensive health monitoring systems. To record ECG signals on the toilet seat, a special sensor was developed, as well as polymer dry electrodes with different textures. A new wearable ECG measurement system [5] based on smart clothes consists of three subsystems, including smart clothes, a smartphone and a PC terminal. Three textile ECG electrodes are woven into the fabric of the smart clothes, and the smart clothes can transmit the received ECG signals to a smartphone via Bluetooth. The ECG signals are then sent by the smartphone to the PC terminal via WiFi, cellular network or the Internet. Also, the smart bracelet [6] VITAL-ECG, developed by the Neuronica Lab of the Polytechnic University of Turin, allows to record basic vital parameters such as heart rate and ECG, SpO 2, skin temperature and humidity. The developed mobile system [7] made to improve the ability to manage patients' cardiovascular diseases, as well as reduce the workload of doctors, includes both hardware and cloud software devices based on the latest advances in Internet of Things (IoT) and artificial intelligence (AI) technologies. A small hardware device has been developed to collect high quality electrocardiogram (ECG) data from the human body. A new cloud service based on deep learning will be deployed to automatically detect cardiovascular diseases. Twenty types of diagnostic items are supported, including sinus rhythm, tachycardia, and bradycardia. A study [8] explored the feasibility of multi-lead ECG recording using dry capacitive electrodes with a specially designed portable device to explore the feasibility of a wearable ECG device with electrodes embedded in clothing.

The main disadvantage of these devices from ECG devices with conventional wet electrodes is low signal accuracy.

The study [9] investigates the AliveCor portable device for recording and measuring the QTc interval on a 6-lead ECG. Automated QTc data from the 12-lead ECG for each patient ($n = 13$) were compared with the mean QTc value calculated from the corresponding AliveCor record of each patient. AliveCor underestimates QTc - 92% of the time, AliveCor calculated QTc as lower than their respective 12-lead QTc readings. AMAZFIT ® [10], a new wearable electrocardiogram (ECG) recording system, is used to measure, collect and store adult single-lead heart curves. The aim of the study was to evaluate the accuracy of AMAZFIT ® for diagnosing arrhythmia in elderly patients. The study has some limitations. First, the sample size was relatively small, especially the cohort of patients with arrhythmias without arrhythmias. Secondly, we only measured heart rate in very stable patients when immobile. Indonesian scientists have developed a portable and inexpensive Holter system [11] for recording an ECG signal during the day. The motherboard consists of a preamplifier, a bandpass filter, a notch filter, a summing amplifier, an Arduino microcontroller, an SD memory card, and a Bluetooth transmitter. The ECG signal is taken from the body based on a standard LEAD II measurement. At the stage of the laboratory sample and is large in size.

Software [12] was developed using MATLAB for QRS detection based on a combination of simple unweighted moving averages. In particular, the main elements of ECG signal processing are moving average cascades (MAC). Our algorithm improves the QRS complex by selecting a MAC from a set of MAC derivatives that are inherently noise-tolerant. Adaptive digital filtering is applied [13] to eliminate any interfering noise that may occur in a typical home environment while minimizing processing time. The accuracy of ECG and EMG signal coverage is evaluated using Bland-Altman analysis by comparison with a reference instrument for collecting physiological signals. The method proposed in [14] first extracts sequences of first-order differential RR intervals (Delta RRI) from segmented ECG data, and then performs a polar coordinate transformation on a Delta RRI Poincaré plot to obtain a phase distribution. Two features, distribution width Dw and mean distribution height Dh, are extracted from the phase distribution to classify episodes of arrhythmia and episodes without arrhythmia.

The model proposed in [15] is presented as a three-level model for analyzing ECG signals, which can potentially be adopted in portable and wearable real-time monitoring devices and it is designed, implemented and modeled the proposed CNN network using Matlab. The author in [16] proposed a classification model with low computational cost to reliably detect arrhythmia episodes in ECG signals by using signal RR intervals and injecting them into an artificial neural network (ANN) for classification to compensate for the lack of computational complexity in traditional wearable ECG monitoring devices. In [17] there is presented a deep learning algorithm for determining high-quality intervals on single-lead ECG recordings obtained from patients with paroxysmal arrhythmia. The study [18] uses a new implementation of a 1D Convolutional Neural Network (CNN) integrated with a validation model to reduce false positives. This CNN architecture consists of an encoder block and a corresponding decoder block, followed by a sampling classification layer to construct a one-dimensional R-peak segmentation map from the input ECG signal. All experiments in these papers are based on simulation software. To increase the practical value of the proposed algorithms, it is necessary to implement monitoring systems in a real platform.

The review shows that devices and systems that allow detecting dangerous cardiac arrhythmias in real time and preventing fatal outcomes have not yet been developed. Improving our non-invasive heart monitoring system remains an urgent scientific and technical task.

III. Development of the Hardware of a Portable Cardioanalyzer

A. *Development of a Functional Diagram of a Portable Mobile Cardioanalyzer*

Fig. 2 shows a detailed functional block diagram of the developed portable mobile ECG device.

The main element of a portable ECG device is a microcontroller. The ADC for ECS recording is connected to the microcontroller via a serial data link. The ADC must be independently powered to minimize noise passing through the power lines from the microcontroller and other digital components of the system.

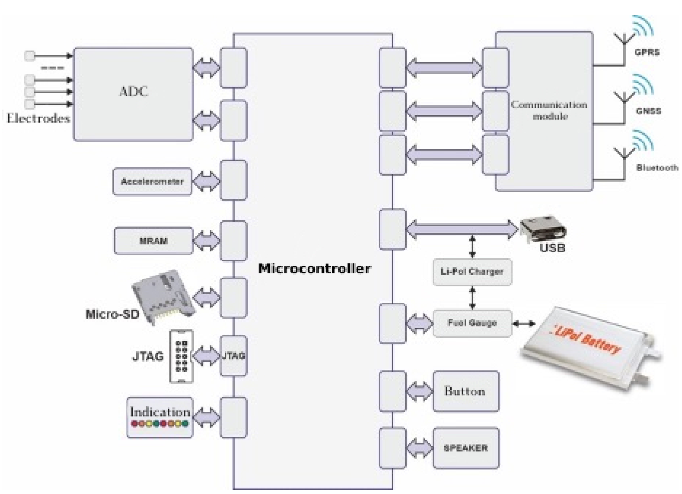


Fig. 2. Functional Block Diagram of the Cardioanalyzer.

To record operational data on current ECS, a block of non-volatile high-speed memory MRAM is connected to the microcontroller. For further accumulation and long-term storage of ECG data, an external drive is connected to the microcontroller in the form of a MicroSD Flash memory reader, into which a Flash card with non-volatile memory up to 32 GB can be inserted.

The ECG device is equipped with a MEMS accelerometer that detects falls, and immobility of patient.

The portable ECG device is powered from an external power source. The charger ensures that the lithium polymer battery is properly charged and then disconnected in accordance with battery regulations. In accordance with the requirements of portable stand-alone devices, the portable ECG device is equipped with a power button controller, which provides manual control of power supply. Also, the power supply system of the portable ECG device is equipped with a device that determines the remaining battery charge, with the ability to transfer and record data on the current value of the battery charge to the microcontroller.

For communication with the ambulance service, as well as for the rapid transmission of ECS data to the server of the medical information system (MIS), the portable ECG device is equipped with a GPRS communication module. In the case of ECG data transmission via an external mobile phone, the portable cardiac analyzer is equipped with a Bluetooth module. The heart analyzer is also equipped with a GNSS satellite navigation module to determine the exact location of the patient.

The system of visual control of the operation of a portable mobile ECG device is made on two three-color and one single-color LEDs. One LED provides visualization of the charger's operation and indicates the "External power connection", "Battery charging" and "Battery charging complete" modes. The second three-color LED is connected directly to the microcontroller and is controlled by software. A single-color LED is connected to the GPRS module and signals the Internet connection modes and data exchange with the server.

B. *Development of an Electrical Circuit Diagram of a Portable Mobile Cardioanalyzer*

The development of a portable mobile ECG device begins with the development of a block diagram of the device, which is followed by the development of a circuit diagram. The block diagram of the device was worked out to the level of interaction of individual blocks. To form a common circuit diagram, it is necessary to work out each individual block, develop a circuit diagram for each block and combine them on a common diagram.

1) Processor module: The processor module is built on a modern low-power STM32L151xD microcontroller. To ensure normal operation of the microcontroller, an external 20 MHz crystal is required; a power supervisor chip to provide a RESET signal at initial power-up, a clock crystal at 32.768 kHz, an SWD connector for programming, and the provision of decoupling capacitors on all power lines in accordance with the technical documentation at this time. Microcontroller when developing a mobile cardiac analyzer, a microcontroller in a TQFP-64 package was used (Fig. 3).

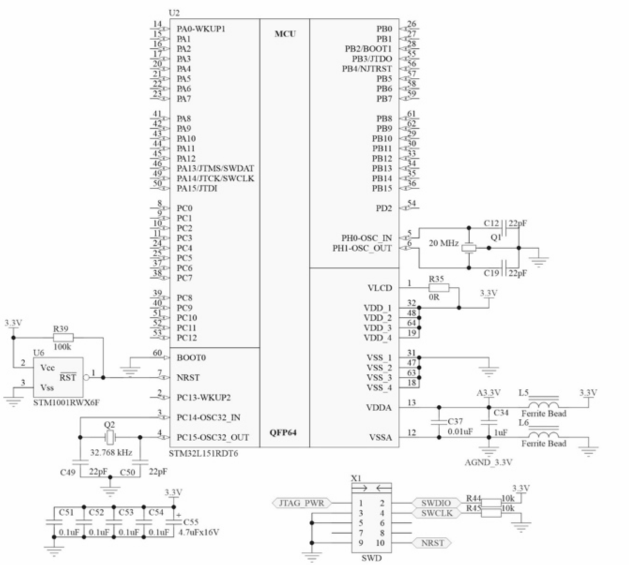


Fig. 3. Schematic Diagram of the Processor Module.

2) ADC module: The ADC module is based on a 4-channel 24-bit Front-End module (FEM) ADS1298 (manufactured by Texas Instruments). The module is functionally complete and is connected to the processor module via a serial synchronous SPI interface. An external 2.048 MHz reference oscillator is required to ensure normal operation of the ADC. This frequency is generated by the microcontroller and fed to the ADS1298 via the MCO line. On each input channel of the ADC (4 channels), protection lines are formed, consisting of two low-frequency RC filters and a diode assembly. Protection lines are used to filter high-frequency signals, eliminate reverse polarity and input overvoltage. This ADC construction option is shown in Fig. 4.

3) RAM block: The RAM block is built on non-volatile MRAM memory MR25H40 with a memory organization of 512Kx8. Access to non-volatile memory is carried out via a serial synchronous SPI interface. The MRAM connection option is shown in Fig. 5.

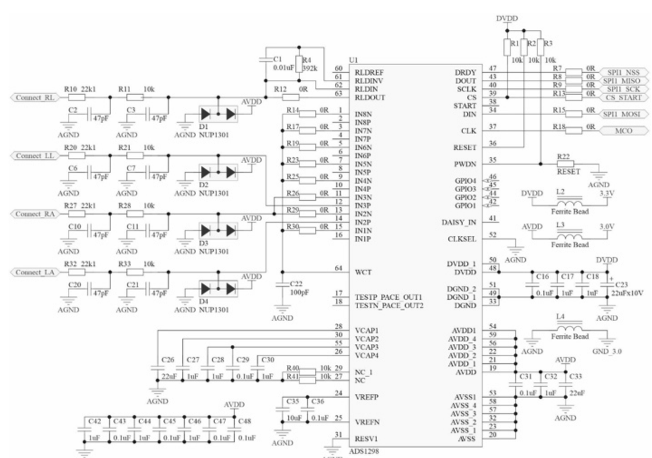


Fig. 4. Schematic Diagram of the ADC Module.

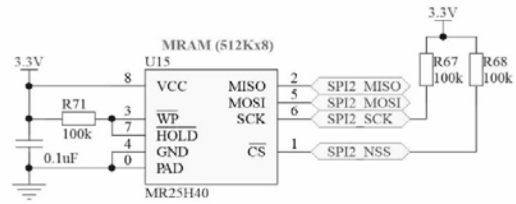


Fig. 5. Schematic Diagram of the RAM Block.

4) Persistent memory block: To ensure non-volatile storage of large amounts of data, a replaceable microSD solid-state drive with a memory capacity of up to 32 gigabytes is used. To use an external drive in a portable ECG device, a microSD cardholder is used, which is connected to the microcontroller via a special SDIO bus. The MRAM connection option is shown in Fig. 6.

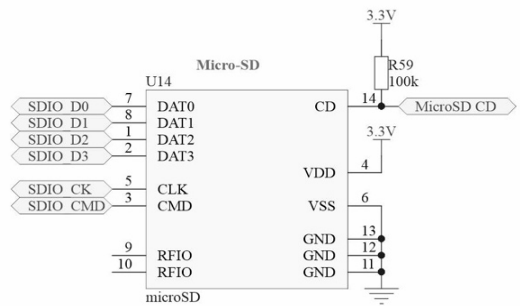


Fig. 6. Schematic Diagram of a Block of Permanent Memory.

5) Block for determining accelerations: The acceleration detection unit is built on a MEMS accelerometer that registers shocks, falls and a stationary state. As an accelerometer, a LIS3DH 3-axis MEMS accelerometer chip in an LGA-16 package from ST Microelectronics was used. The MEMS accelerometer is connected to the microcontroller via the I2C serial interface. Also, the accelerometer has 2 interrupt outputs INT1 and INT2, through which it is possible to wake up and wake up the microcontroller from sleep in case of sharp shocks and falls. The accelerometer connection option is shown in Fig. 7.

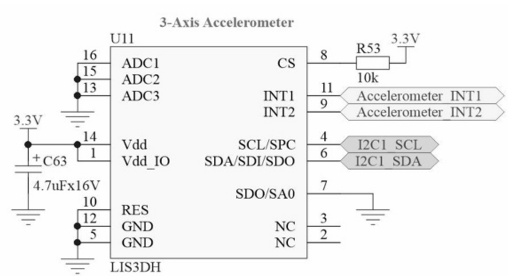


Fig. 7. Schematic Diagram of the Block for Determining Accelerations.

6) Communication block: A communication unit was developed based on the universal combo module SIM868 from SIMCOM, which supports data exchange in GSM / GPRS cellular networks, reception of location data and exact time signals from navigation satellites (GNSS). The SIM868 uses a Serial Port Interface (SPI) to communicate with the microcontroller. Because since the SIM868 module and the microcontroller have different supply voltages, a signal level converter on the ADG3308 chip from Analog Devices was used to ensure compatibility. To connect a telephone microSIM card module, we used a cardholder model MOLEX 503960-0696, which is additionally protected from static electricity by a quad suppressor SMF05C for the lines going to the SIM868 module. 3 antennas are connected to the SIM868 module: GPRS, GNSS and Bluetooth. Additional matching of these antennas with the module is performed on RLC components, which is shown in Fig. 8.

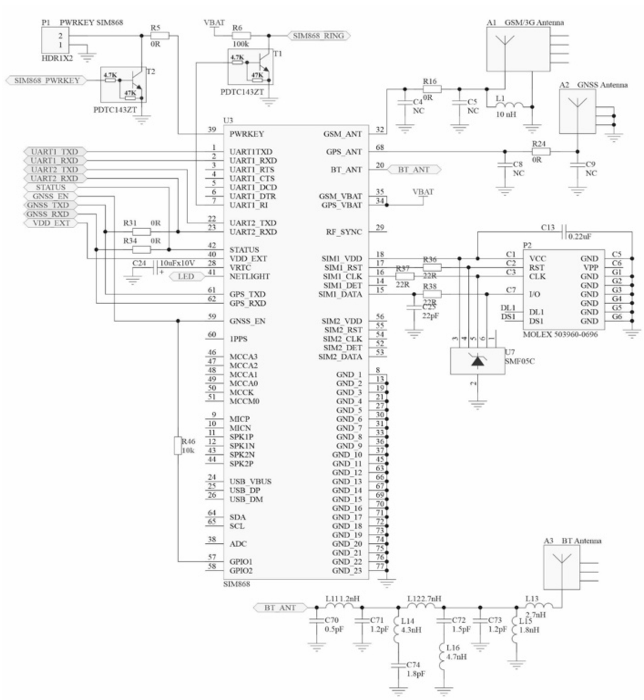


Fig. 8. Schematic Diagram of the Communication Block.

7) Battery charge control unit: A battery is required for the portable mobile cardiac analyzer to operate in standalone mode. When choosing a battery, the main criteria were:

- large capacity.
- minimum thickness.
- dimensions are as close as possible to the size of the printed circuit board.

A Li-POL battery in the form of a prism with a size of 90x53x4.4mm fully meets these criteria. Since such a battery requires a charging block, the BQ24070 chip from Texas Instruments was used in the design of this block. This chip is a specialized module for charging a Li-POL battery in 1S format, i.e. with a maximum voltage when charging no more than 4.3V. The BQ24070 can support up to 2A lithium polymer battery charging while supporting external power supply up to 16V. The microcircuit ensures the correct process of charging the battery in accordance with the technical documentation for lithium polymer batteries. The microcircuit ensures the transfer of data on the status of the charge / discharge process to the microcontroller and in parallel this status is displayed on the LED indicator. The circuitry of the developed battery charge control unit is shown in Fig. 9.

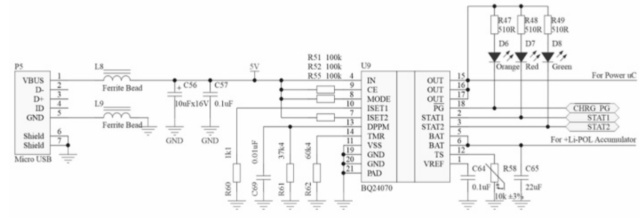


Fig. 9. Schematic Diagram of the Battery Charge Control Unit.

8) Block for determining the residual battery charge: To determine the exact value of the residual battery charge, circuitry was developed based on the MAX17043 chip from MAXIM. This microcircuit determines the level of residual charge and transmits data via a two-wire I2C interface. The circuitry of the developed battery charge control unit is shown in Fig. 10.

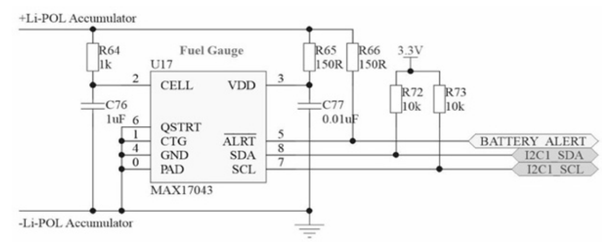


Fig. 10. Schematic Diagram of the Block for Determining the Residual Battery Charge.

9) Overcharge/overdischarge protection block: When operating a Li-POL battery, strict observance of the charge / discharge modes is required. It is necessary to limit the current and voltage at the end of the charge cycle to prevent overcharging and failure of the battery. It is also required to completely disconnect the load from the battery when the battery is discharged and the voltage drops below the minimum for Li-POL batteries. These modes are provided by the DW01A chip manufactured by H&M Semiconductor. The microcircuit protects the battery from overcharging / overdischarging by controlling the load with power N-MOSFET transistors in the lower arm. The developed scheme of the battery protection unit is shown in Fig. 11.

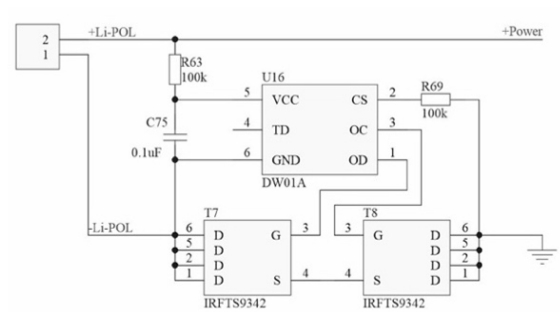


Fig. 11. Schematic Diagram of the Battery Protection Unit.

10)Indication and notification block: The indication of a portable mobile ECG device is represented by three LEDs that determine the charge mode in the battery charge control unit; one LED providing indication of the operation of the radio channel in the communication module; three LEDs connected to the microcontroller and providing indication for various program modes. Also, the microcontroller, using a transistor key, controls the warning system built on a piezoceramic sound emitter. This circuitry is shown in Fig. 12.

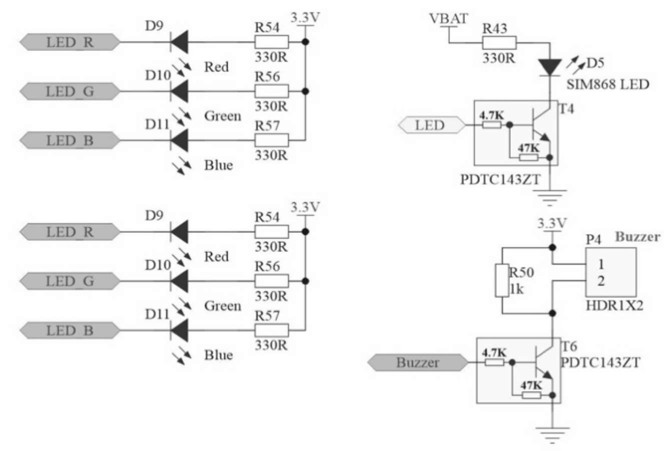


Fig. 12. Schematic Diagram of the Indication and Notification Block.

C. *Development of the Printed Circuit Board (PCB) of a Portable Mobile Cardio Analyzer*

When developing a portable ECG device, a list of 215 components in 85 positions (nominal values) was used. The design complexity of the device is quite high, because a large number of components are used.

It is determined that the printed circuit board will be multilayer. Based on the design criteria for multilayer printed circuit boards, the number of individual layers allocated for power supply and the complexity of the circuitry, the number of layers in the designed PCB was determined - 6 layers. SMT components were mounted on one side of the PCB.

When arranging the components, the total area of all components and the estimated overall dimensions of the PCB were determined. After the final arrangement of the components and the multilayer tracing, the final dimensions of the PCB were determined. The size of the software was 120 x 60mm. For visual representation of the PCB and for the subsequent design of the hull solution, 3D models of the PCB with installed components were generated on an accurate scale. See Fig. 13.

1) Defining a conductor as a transmission line: When tracing the printed circuit board, it was revealed that it was necessary to connect the chip antennas to the radio transmitting module. Since the operating frequencies of the antennas lie within 1.8 ... 2.48 GHz, it is necessary to analyze and determine whether the strip line is an inhomogeneous transmission line, since when designing high-speed devices, various kinds of reflections and signal distortions occur. In order for the signal to retain its integrity, it is necessary to develop transmission lines with a given wave impedance in the form of conductors of a certain width on a printed circuit board. Signal transmission without distortion is possible along a transmission line with a given wave impedance from the source to the signal receiver.

The electrical length of the printed conductor depends on the minimum wavelength of the transmitted signal and, accordingly, on the rise time of this signal. If the copper conductor on the PCB is too long in relation to the rising edge, then the conductor should be designed as a transmission line to prevent distortion. A copper conductor on a PCB is a transmission line if its electrical length on the PCB is more than 1/3 of the rise time. That is, the conductor is electrically long. Signal propagation velocity V_p is the speed at which an electrical signal travels along a copper conductor on a printed circuit board. The speed is calculated by the formula.

$$V_P = \frac{c}{\sqrt{\varepsilon_r}} \tag{1}$$

where c - is the speed of light,

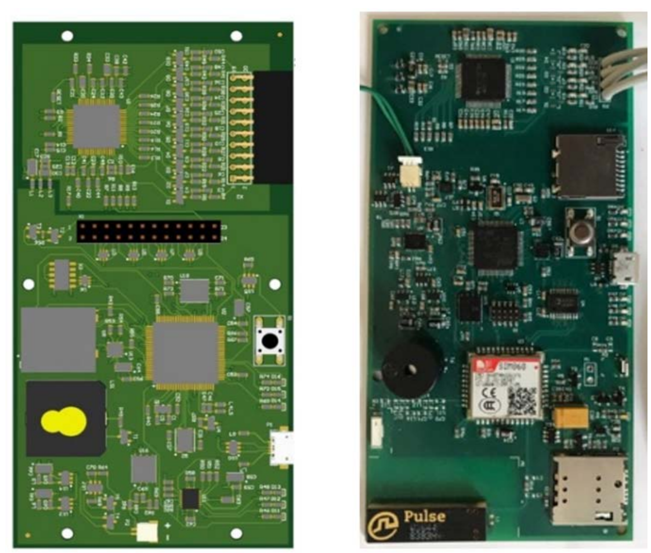
εr - is the dielectric constant of the printed circuit board material.

Since we are using the widespread FR-4 material with a dielectric constant of 4.0 in the project, the signal propagation speed will be equal to:

$$V_P = \frac{c}{\sqrt{\varepsilon_r}} = \frac{299{,}79258}{\sqrt{4}} mm/ns = 149.89 mm/ns,$$

The length of the printed conductor L_r is calculated according to the rule of 1/3 signal rise time:

$$L_r \geq \left(\frac{T_r}{3}\right) * V_p, \tag{2}$$



(a) (b)

Fig. 13. 3D Model (a) and Fabricated Sample (b) PCB with Installed Components.

where L_r – conductor length,

T_r - signal rise time, *ns*.

Thus, the printed conductor should be considered as a transmission line when we use FR-4 material, with the following length:

$$L_r \geq T_r * 49.96mm, \quad (3)$$

Thus, if T_r=1 *ns*, then:

$$L_r \approx 50\ mm,$$

That is, it turns out that under the above conditions for frequencies of 2.48 GHz, a printed conductor having a length of 20 mm or more must be considered as a transmission line, which requires additional matching.

Calculation of the wave impedance of a microstrip transmission line.

The total impedance (impedance) of the transmission line from the radio module to the antenna must be 50 ohms, because the GSM/GNSS module has antenna terminals designed to work with antennas with a characteristic impedance of 50 ohms.

The transmission line that connects the module and the antenna must be matched, i.e. should have a wave impedance of 50 ohms.

If the transmission line has a different impedance, then the electromagnetic wave propagating along the copper conductor of the printed circuit board will be partially reflected at the boundary of media with different wave impedances (if the transmission line is homogeneous, reflection will occur at the junctions with the module and with the antenna). This can lead, at best, to a decrease in the sensitivity and output power of the device, and at worst, to failure of the output stage of the module.

When designing a printed circuit board, we define and design the transmission line between the module and the antenna as a microstrip line. The impedance of a microstrip transmission line depends on the ratio of the dielectric thickness between the microstrip and the return conductor, as well as the width of the signal conductor. As long as this ratio is constant, the wave impedance of the microstrip line will also be constant. With a proportional change in these parameters, the wave impedance of the transmission line remains unchanged.

The microstrip line configuration is shown in Fig. 14.

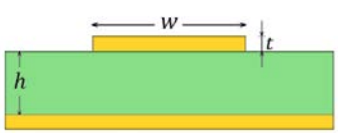


Fig. 14. Cross Section of a Microstrip Line.

The microstrip is a non-uniform transmission line because not all of the field lines between the strip conductor and the ground plane pass through the substrate. Therefore, a wave propagating along a microstrip conductor is a "quasi - T-wave". The effective permittivity εef is less than the permittivity of the substrate, since the field outside the substrate of the return current is also taken into account.

IV. Development of the Software Part of the Cardiodiagnostic System

Currently, wired connections between individual devices in medical systems are being replaced by wireless technologies for information transfer, and a number of factors must be taken into account: the distance between devices and the ability to move them, the transfer rate and amount of information transmitted, power consumption, the possibility of encryption, the level of interference, etc. Consider the communication protocols between the components of the heart monitoring system.

The portable ECG device contains its own communication unit based on the universal combo module SIM868 from SIMCOM, which supports data exchange in GSM / GPRS cellular networks, reception of location data and exact time signals from navigation satellites (GNSS). The SIM868 uses a Serial Port Interface (SPI) to communicate with the microcontroller. Because since the SIM868 module and the microcontroller have different supply voltages, a signal level converter on the ADG3308 chip from Analog Devices was used to ensure compatibility.

In accordance with the functions of the entire system and its individual parts, the structure of the software for the non-invasive heart monitoring system was developed (Fig. 15).

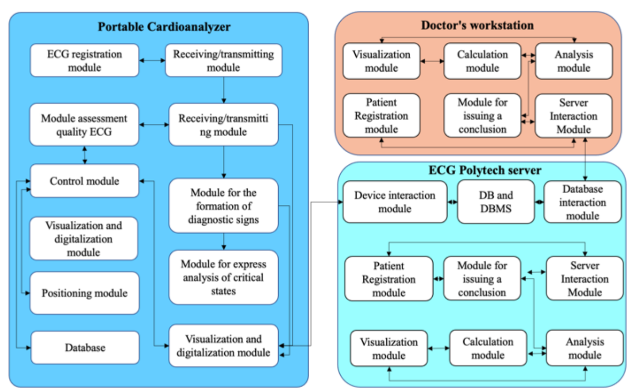


Fig. 15. Structure of the Cardiodiagnostic System.

The software of the non-invasive heart monitoring system consists of the software of a portable ECG device, a server, and a doctor's workstation. The server part (highlighted in blue in Fig. 14) includes a device interaction module, a database (DB) and a database management system (DBMS), a database interaction module.

In accordance with the three-level model, the following operating modes are possible during the operation of the heart monitoring system: "autonomous" - all tasks for processing and analyzing ECG are performed on a portable ECG device; “device -server” – ECG processing and analysis tasks are distributed between the device and the server; "cardiac analyzer-server-doctor" - all tasks for processing heart monitoring information are assigned to the doctor's workstation, and the server is used only for data storage. The levels of the decision hierarchy, the corresponding modes of operation, analysis options and decisions made at various levels are shown in Table I.

TABLE I. LEVELS OF THE DECISION HIERARCHY OF THE CARDIODIAGNOSTIC SYSTEM

Hierarchy levels	Operating modes	Analysis option	Decisions made
Lower	Autonomous	Automatic express analysis of ECG	Suitability of the ECG for analysis, the presence of a critical situation
Average	Cardioanalyzer-server	Automatic differentiated analysis of a critical situation	The need for a detailed medical analysis, the presence of a connection with the doctor's workstation.
Highest	Cardioanalyzer-server-doctor	Detailed medical analysis by a cardiologist	Calling an ambulance, making recommendations.

Actions and decisions that are not highlighted in color are implemented by a portable cardiac analyzer, those highlighted in yellow are implemented by the server, highlighted in blue are the doctor's workstation.

Thus, the developed system structure, software structure, decision-making model, as well as methods, algorithms and structural solutions make it possible to implement a system of non-invasive heart monitoring in accordance with the formulated requirements.

A. User Model, Linking a Specific Portable Cardiac Analyzer to a Specific user

Django comes with a default user model with fields like username, passwords and email, however in some cases these fields may not be enough for us to extend the "User" model or create our own user model. In this case, we will be extending the user model because we need a way to differentiate users.

To determine the device, we generated a unique "id" for each device and associated it with a specific user, so when transferring data, we also pass the "id" of the device. Fig. 16 shows a list of potential patients linked to the system via ID codes.

On the script (Appendix 1, A), using the Socket library, we create a socket to receive data from the device. To do this, you need to specify the server address and the number of maximum handles that can be opened at a time. Since this code was used for testing, the maximum number of descriptors is 30.

Next, we have to write a listener function and a signal handler, or as you can also call it, data packets (Appendix 1, B). The listener first of all looks at the address of the transmitted data, if they are addressed to our IP, we receive the data. Now we have to transfer our data further for processing and, if necessary, for storage in our database under a specific user.

If we run our socket script, at the output we will see data already converted to numbers in the server console. Using the hexlify() function, we converted the bytes first to hex, then by decoding with the decode() function to the String type (as shown in Fig. 17 and 18).

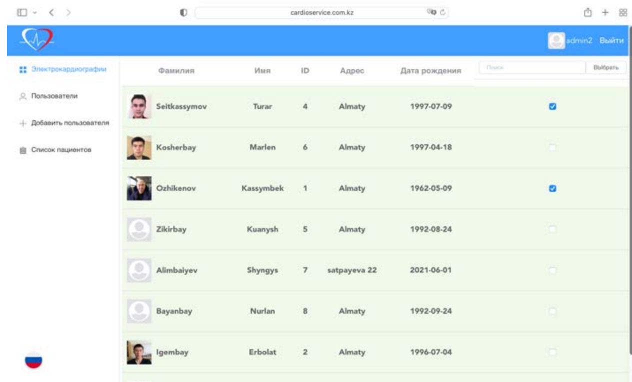


Fig. 16. User Registration in the System.

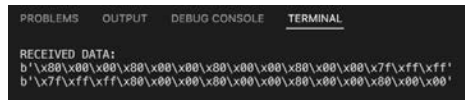


Fig. 17. Initial Data Type (Bytes).

b'\x80\x00\x00\x80\x00\x00\x80\x00\x00\x80\x00\x00\x7f\xff\xff\x7f\xff\xff\x80\x00\x00\x80\x00\x00\x80\x00\x00\x80\x00\x00' – here we assigned the first 8 bytes to identify the device itself, as well as to transfer the user id, date and time of data transfer.

Fig. 18. After Conversion (Type – List<String>).

But it is still impossible to use this data, because here we have a lot of data distorted during transmission or processing, noise and duplicated data. At the stage of processing the received data, we used the POST method. This method ensures secure data transfer. Next, using cropping, select the data area in which the transmission data and id are recorded for identification. A further goal is to filter the data in the packet body. The principle of noise trimming is used here, we determine the upper and lower threshold of the received data and trim the data accordingly. That is, all noise exceeding this threshold will be suppressed. After filtering, having determined the user id, we sort the data by the date and time of receipt and send it to the database for storage.

In our project, to receive and transmit data, we use a programming interface - a socket, which allows us to always accurately determine the state of a person and ensures immediate data exchange. In vue.js, sockets are quite mature and are constantly being improved. To use sockets, it is enough to refer to the web socket initialization function WebSocket(). The first attribute is the address directed to our server. It is worth paying attention to the explicit indication of the id of the patient whose data we want to see.

When running this script, we can observe the calculations aimed at the "@arction/lcjs" library, which is designed to visualize graphs. When drawing graphs, this library uses graph theory, which allows us to effectively arrange the received data in the form of integers. Thanks to this, the user can visually observe the pacemaker in real time (Fig. 19).

Fig. 19. Real-time ECG Visualization.

An important point in our system is the definition of dangerous cardiac arrhythmias. The key value in determining dangerous cardiac arrhythmias is - heart rate. To calculate the heart rate value, it is necessary that our device receives an ECG within 60 seconds, this will be enough. Knowing the formula for determining heart rate in advance, we rewrote the formula into a similar function, the syntax of which is written below:

$$heart\ rate = \frac{60}{R-R}, \tag{4}$$

where 60 is the number of seconds in a minute, R-R is the duration of the interval, expressed in seconds.

Formula for calculating heart rate written in the JavaScript programming language is given below:

```
if (self.seconds >= 60) {

self.chss = parseInt(len / 360);

self.rr = 60 / parseInt(len / 360);
```

It is also important to calculate the RR value (the distance between two signal amplitudes, that is, heartbeats). Having obtained these values, we can easily determine two extremely dangerous diagnoses: “Sinus bradycardia” and “Sinus tachycardia”. To be more precise, with a very rare heart rate, less than 45 beats/sec, “Sinus bradycardia” is detected, and if more than 100 beats/sec, then “Sinus tachycardia”. Code to visualize the conclusion is given in Appendix 1, C.

When we finally calculated the values needed for the conclusion, the user can see the final diagnosis of his condition (Fig. 20).

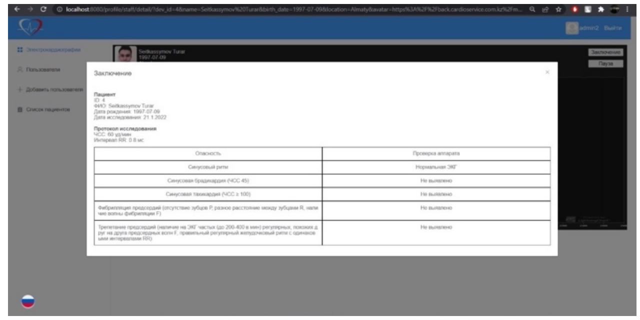


Fig. 20. Visualization of Diagnostic Result.

V. Discussion

The ECG monitoring system is an example of the practical implementation of the developed portable device and the method of noise-resistant ECG processing for patients with free activity. The basis for the creation of ECG monitoring system is modern information and communication technologies and patent-protected methods developed by the authors, algorithms, structural, circuit and software solutions.

The main features of the device in comparison with existing are:

- work in conditions of free activity;
- performing a preliminary ECG analysis;
- automatic diagnosis of life-threatening arrhythmias;
- call an ambulance;
- storing ECS entries locally and on a remote server;
- ability to integrate with the medical information system

The introduction of a ECG monitoring system in medical practice requires additional clinical trials and can serve as the basis for successful competition with leading foreign companies working in this field.

Currently, on the basis of the Satbayev University, ten sets of a portable ECG device have been designed and preclinical tests of the system are being conducted.

The developed system does not allow diagnosing more complex heart diseases, such as myocardial infarction. Currently, the authors are trying to develop a method of neural network analysis of ECG for diagnosing myocardial infarction and improving the accuracy of determining the location of myocardial damage, in particular, identifying the stages of myocardial infarction and the depth of myocardial damage.

VI. Conclusion

A mobile intelligent heart monitoring system has been developed that provides a three-level ECG analysis: automatic express ECG analysis (mode: "Autonomous"), automatic differentiated ECG analysis ("device -server" mode), detailed medical analysis using a cardiologist's workstation ("device -server" mode). doctor server). Based on the formulated requirements, a detailed functional block diagram of the device was developed and built. According to the developed functional block diagram, an analysis was carried out, criteria for selecting components were developed, and an element base was selected for designing a hardware platform for a portable mobile ECG device. After selecting the element base, an electrical circuit diagram of a portable mobile ECG device was designed. Particular attention was paid to the functional composition of the device and its reduced power consumption. Based on the developed electrical circuit diagram, the printed circuit board of the device was designed.

Software for a mobile application has been developed for receiving a signal from an ECG recording device, performing its preliminary processing, diagnosing life-threatening cardiac arrhythmias, notifying the patient about the diagnostic results, and transmitting data to the application server.

We evaluated the performances of the developed system. The signal-to-noise ratio of the output signal is favorable, and all the features needed for a clinical evaluation (P waves, QRS complexes and T waves) are clearly readable.

The current model of a portable information-measuring system for monitoring the heart in conditions of free activity is being tested in the city hospital of JSC "Central Clinical Hospital" in Almaty.

Acknowledgment

The current model of a portable monitoring system was developed on the basis of Satbayev University under the grant program of the Science Fund of the Republic of Kazakhstan. Project №0281-18-"Portable ECG device".

References

[1] Statistical data of Medinform LLP, Kazakhstan, http://www.medinfo.kz/#/stats.

[2] Development of measuring system for determining life-threatening cardiac arrhythmias in a patient's free activity, Alimbayev Ch., Alimbayeva Zh., Ozhikenov K., Bodin O., Mukazhanov Y., Eastern-european journal of enterprise technologies, Vol 1, No 9 (103) (2020), DOI: https://doi.org/10.15587/1729-4061.2020.197079.

[3] A Non-Contact Compact Portable ECG Monitoring System, Chen Q., Kastratovic S., Eid M., Ha S., Electronics, Volume 10, Issue 18, Article Number 2279, DOI 10.3390/electronics10182279, Published SEP 2021.

[4] Design and evaluation of a novel approach to invisible electrocardiography (ECG) in sanitary facilities using polymeric electrodes, Silva A., Almeida H., da Silva H., Oliveira A., Scientific reports, Volume 11, Issue 1, Article number 6222, DOI 10.1038/s41598-021-85697-2, Published on MAR 18 2021.

[5] Wearable Measurement of ECG Signals Based on Smart Clothing, Li M., Xiong W., Li Y., International journal of telemedicine and applications, Volume 2020, Article number 6329360, DOI 10.1155/2020/6329360, Published on JAN 18 2020.

[6] A Wearable Smart Device to Monitor Multiple Vital Parameters-VITAL ECG, Randazzo V., Ferretti J., Pasero, E., Electronics, Volume 9, Issue 2, Article number 300, DOI 10.3390/electronics9020300, Published FEB 2020.

[7] Artificial-Intelligence-Enhanced Mobile System for Cardiovascular Health Management, Fu Z., Hong S., Zhang R., Du S., SENSORS, Volume 21, Issue 3, Article number 773, DOI 10.3390/s21030773, Published FEB 2021.

[8] Development and Test of a Portable ECG Device with Dry Capacitive Electrodes and Driven Right Leg Circuit, Zompanti A., Sabatini A., Grasso S., Pennazza G., Ferri G., Barile G., Chello M., Lusini M., Santonico M., Sensors, Volume 21, Issue 8, Article Number 2777, DOI 10.3390/s21082777, Published APR 2021.

[9] Monitoring prolongation of QT interval in patients with multidrug-resistant tuberculosis and non-tuberculous mycobacterium using mobile health device AliveCor, Puranik S., Harlow C., Martin L., Coleman M., Russell G., Park M., Kon O., Journal of clinical tuberculosis and other mycobacterial diseases., Volume 26, Article Number 100293, DOI 10.1016/j.jctube.2021.100293, Published FEB 2022.

[10] The Auxiliary Diagnostic Value of a Novel Wearable Electrocardiogram-Recording System for Arrhythmia Detection: Diagnostic Trial, Zhang S., Xian H., Chen Y., Liao Y., Zhang N., Guo X., Yang M., Wu J., Frontiers in medicine, Volume 8, Article number 685999, DOI 10.3389/fmed.2021.685999, Published on JUN 24 2021.

[11] A low-cost Holter monitor design equipped with external memory and Bluetooth connection, Nofitasari D., Wisana I., Triwiyanto, T., Setioningsih E., Mak'ruf M., Nugraha P., International symposium on materials and electrical engineering, Book Series IOP Conference Series-Materials Science and Engineering, Volume 850, Article Number 012020, DOI 10.1088/1757-899X/850/1/012020, Published 2020.

[12] QRS Detection Based on Medical Knowledge and Cascades of Moving Average Filters, Bachi L., Billeci L., Varanini M., Applied sciences-basel, Volume 11, Issue 15, Article Number 6995, DOI 10.3390/app11156995, Published AUG 2021.

[13] A Real-Time Wearable Physiological Monitoring System for Home-Based Healthcare Applications, Jeong J., Lee W., Kim Y., Sensors, Volume 22, Issue 1, Article number 104, DOI 10.3390/s22010104, Published JAN 2022.

[14] An improved Poincare plot-based method to detect atrial fibrillation from short single-lead ECG, Luo C., Rao H., Huang X., Jiang H., Rao N., Biomedical signal processing and control, Volume 64, Article Number 102264, DOI 10.1016/j.bspc.2020.102264, Published FEB 2021.

[15] Multiclass ECG Signal Analysis Using Global Average-Based 2-D Convolutional Neural Network Modeling, Wasimuddin M., Elleithy K., Abuzneid A., Faezipour M., Abuzaghleh O., Electronics, Volume 10, Issue 2, Article number 170, DOI 10.3390/electronics10020170, Published JAN 2021.

[16] Automated Classification of Atrial Fibrillation Using Artificial Neural Network for Wearable Devices, Ma F. Zhang J., Liang W., Xue J., Mathematical problems in engineering, Volume 2020, Article number 9159158, DOI 10.1155/2020/9159158, Published APR 25 2020.

[17] Application of Deep Learning for Quality Assessment of Atrial Fibrillation ECG Recordings, Huerta A., Martinez-Rodrigo A., Arias M., Langley P., Rieta J., Alcaraz R., Computing in cardiology, Computing in Cardiology Conference Book Series, Article Number 367, DOI 10.22489/CinC.2020.367, Published 2020.

[18] Robust R-Peak Detection in Low-Quality Holter ECGs Using 1D Convolutional Neural Network, Zahid, M., Kiranyaz S., Ince T., Devecioglu O., Chowdhury M., Khandakar A., Tahir A., Gabbouj M., IEEE Transactions on biomedical engineering, Volume 69, Issue 1, Pages 119-128, DOI 10.1109/TBME.2021.3088218, Published JAN 2022.

Authors' Contribution

Zhadyra N. Alimbayeva conducted the main research throughout the paper and contributed to the development of software part of the system in section IV A; Kassymbek A. Ozhikenov is the project supervisor and participated in all parts of the study; Chingiz A. Alimbayev contributed to the development of the hardware of a portable ECG device (part III, A, B) and wrote the paper; Oleg N. Bodin developed a three-level model of the heart monitoring system (part IV); Yerkat B. Mukazhanov wrote literature review and problem statement (part III); all authors had approved the final version; Nurlan A. Bayanbay contributed to the development of the printed circuit board (PCB) of a portable mobile cardio analyzer (part III, C); all authors had approved the final version.

APPENDIX 1

A. Scripts for Receiving Data from the Device

```
import socket
import select
import requsts
import binascii

SERVER_ADDRESS=('***.***.**.**',"****")

dd=""
# Говорит о том, сколько дескрипторов единовременно могут быть открыты
MAX_CONNECTIONS=30

# Откуда и куда записывать информацию
INPUTS=list90
outputs=LIST()

def get_non_blocking_server_socket():
 #Создаем сокет, который работает без блокирования основного потока
 server= socket.socket(socket.AF_INET, socket.SOCK_STREAM)
 server.setblocking(0)

 # Биндим сервер на нужный адрес и порт
 server.bind(SERVER_ADDRESS)

 # Установка максимального количество подключений
 server.listen(MAX_CONNECTIONS)

 return server
```

B. Handling the Occurrence of Events at the Inputs

```
def handle_readables(readables,server);
      #Обработка появления событий на входах
global dd
for resource in readables:
            # Если событие исходит от серверного сокета, то мы получаемновое подключение
            if resource is server:
                        connection, client_address=resource.accept()
                        connection.setblocking(0)
                        INPUTS.append(connection)
      #Если событие исходит не от серверного сокета, то мы получаем новое подключение
else:
 data=" "
 try:
 data = resource.recv(1024)
# Если сокет был закрыт на другой стороне
Except ConnectionResetError:
 pass

if data:
 data=binascii.hexlify(data).decode()
 # Выход полученных данных на консоль
            print("Received data:{data}".format(data=str(data)))
            # Первоначальный фильтр ддя поучения данных, исключая все микрозапросы
 if len(data) > 18:
 response=requests.post('https://back.cardioservice.com.kz/api/setByte/', data={'byte':str(data)})
            print(response)
 # Говорим о том, что мы будем еще и писать данный сокет
            if resource not in OUTPUTS:
 OUTPUTS.append(resource)
 # Если данных нет, но событие сработало, то ОС отправляет флаг о полном прочтении ресурса и о его закрытии
 else:
      # Очищаем данные о ресурсе и закрываем дескриптор
clear_resource(resource)
```

C. Code to Visualize the Conclusion

```
<div>
      <b>{{ $t("protocol") }}</b>
</div>
<div>{{ $t("hs") }}: {{ chss }} {{ $t("bl_min") }}</div>
<div></div>
<div class="mb10">{{ $t("interval") }} RR: {{ rr }} Mc/div
<div class="table-conclusion">
      <div> {{ $t("danger") }}</div>
      <div>{{ $t("device_check") }}</div>
</div>
<div class="table-conclusion">
      <div>{{ $t("sinus_rhythm") }}</div>
      <div> {{ $t("normal_ecg") }}</div>
</div>
<div class="table-conclusion">
<div>{{ $t("sinus_bradycardia") }} ({{ $t("hs") }} 45)</div>
      <div v-if="chss > 45">{{ $t("not_found") }}</div>
      <div v-else>{{ $t("found") }}</div>
</div>
<div class="table-conclusion">
<div> {{ $t("sinus_tachycardia") }} ({{ $t("hs") }} >= 100)</div>
      <div v-if="chss < 100">{{ $t("not_found") }}</div>
      <div v-else>{{ $t("found") }}</div>
</div>
```

D. Processing of Received Data on the Server and Assignment to the User.

```
class SetBytesView(APIView):
 permission_classes = (permissions.AllowAny,)
 def get(self, request):
 profiles = Profile.objects.all()
 serializer = ProfileSerializer(profiles, many=True)
 return Response(serializer.data[0])

 def post(self, request):
 group_name = "user"
 channel = get_channel_layer()

 try:
 byte = request.POST.get("byte")
 print(byte)
 p = None
 array = byte[2:len(byte)-1]
 array = byte
 lenght = len(array)
 byte_array = []
 if lenght > 18:
 s = 0
 for i in range(12, len(array), 6):
 bytes = ''
 bytes += array[i:i+6]
 if len(bytes) == 6:
 point = int(bytes, 16)
 if point < 10:
 continue
 if point>10 and point < 12400000:
 point = 12400000
 byte_array.append(point)
 wid = int(array[:12], 16)
```

```
p = Profile.objects.filter(device_id=wid)
if p.exists():
p = p[0]
else:
p = Profile.objects.create(device_id=wid)
byte_array.insert(0, wid)
group_name = "room_"+str(wid)
async_to_sync(channel.group_send) (
group_name,
{
'type': 'send_point',
'content': {
'pointers': byte_array,
}
}
)
today_d = datetime.now()
today = f'{today_d.year}-{today_d.month}-{today_d.day}'
pd = ProfileData.objects.filter(profile = p, date=today)
if pd.exists():
pd = pd[0]
pd.data = pd.data + byte_array[1:]
if len(pd.data) > 50000:
pd.data = byte_array[1:]
pd.save()
else:
ProfileData.objects.create(date=today, data = byte_array[1:], profile=p)
return JsonResponse({'status': 'ok'})
except ValueError as e:
return JsonResponse(e.args[0], status.HTTP_404_NOT_FOUND)
```

E. Front-end.

```
mounted() {
const lcjs = require("@arction/lcjs");
const { AxisScrollStrategies, emptyLine } = lcjs;
this.graf(this.data);
this.socket = new WebSocket(
"wss://back.cardioservice.com.kz/api/setByte/?wid=" + this.did
);
let self = this;
this.timer = setInterval(function () {
self.ss += 1;
if (self.ss >= 61) {
self.ssCheck = true;
self.ss = 0;
}
}, 1000);
let period = [];
let oldK = 0;
this.socket.onopen = function (e) {
console.log("open");
};
let len = 0;
this.socket.onmessage = function (event) {
let d = JSON.parse(event.data)["content"]["pointers"]["content"][
"pointers"
];
len += d.slice(1).length;
if (self.seconds >= 60) {
self.chss = parseInt(len / 360);
self.rr = 60 / parseInt(len / 360);
}
for (let i = 1; i < d.length; i++) {
if (d[i] > 10) {
self.k += 10;
let mmax = Math.max(...period)
let mmin = Math.min(...period)
self.series.add({ x: self.k, y: d[i] });
let mmax = self.series.getYMax() + 100000
let mmin = self.series.getYMin() - 100000
let mmax = d[i] + 55000;
let mmin = d[i] - 55000;
self.chart
.getDefaultAxisY()
.setTickStrategy("Empty")
.setStrokeStyle(emptyLine)
.setInterval(mmin, mmax, false, true)
.setScrollStrategy(AxisScrollStrategies.progressive);
self.data.push({x: self.k, y: d[i]})
}
}
};
this.socket.onerror = function (error) {
console.log(error);
};
},
methods: {
pause() {
this.socket.close();
},
graf(p) {
const lcjs = require("@arction/lcjs");
const {
lightningChart,
DataPatterns,
AxisScrollStrategies,
SolidLine,
SolidFill,
ColorHEX,
AutoCursorModes,
Themes,
emptyLine,
emptyTick,
} = lcjs;

this.chart = lightningChart()
.ChartXY({
})
.setTitle("");
// Add line series to visualize the data received
this.series = this.chart.addLineSeries({
dataPattern: DataPatterns.horizontalProgressive,
});
// Style the series
this.series.setStrokeStyle(
new SolidLine({
thickness: 3,
fillStyle: new SolidFill({ color: ColorHEX("#5aafc7") }),
})
);
this.chart.setAutoCursorMode(AutoCursorModes.disabled);
// Setup view nicely.
this.chart
.getDefaultAxisY()
.setTickStrategy("Empty")
.setStrokeStyle(emptyLine);

this.chart
.getDefaultAxisX()
.setInterval(0, 3000)
.setScrollStrategy(AxisScrollStrategies.progressive);

let old = p[0]
for (let i in p){
this.k+=3
// if (Math.abs(p[i]-old) < 2000000){
this.series.add({x: this.k, y: p[i]})
// }
}
let lcjss = document.querySelector("#lcjs-auto-flexbox");
let section = document.querySelector(".section");
lcjss.style.height = "100%";
```

```
lcjss.style.marginTop = "40px";
section.appendChild(lcjss);
lcjss.querySelector("canvas").style.zIndex = "99";
},
conclusion() {
this.dialogVisible = true;
},
},
beforeDestroy() {
this.socket.close();
},
```

Development of an IoT Device for Measurement of Respiratory Rate in COVID-19 Patients

Jean Pierre Tincopa[1], Paulo Vela-Anton[2], Cender U. Quispe-Juli[3], Anthony Arostegui[4]
Departamento de Ingeniería, Laboratorios de Investigación y Desarrollo, Facultad de Ciencias y Filosofía
Universidad Peruana Cayetano Heredia, Lima, Perú[1, 2, 3]
Centro de Investigación en Transformación Digital, Universidad Privada Norbert Wiener, Lima, Perú[1]
Dirección de Investigación en Salud, Instituto de Evaluación y Tecnologías en Salud e Investigación, Lima, Perú[2]
Departamento de Ingeniería Industrial, Universidad de Ingeniería y Tecnología, Lima, Perú[4]

Abstract*—During the COVID-19 pandemic, patients who required face-to-face attention and tested positive, even showing signs of high risk, were forced to isolate themselves in their own homes immediately without adequate medical monitoring. Continuous remote monitoring of their vital signs would have helped to avoid subsequent hospitalization caused by the progression of the virus. Using deterministic design methods, a system to measure respiratory rate through impedance pneumography was proposed, amplifying microvolt signals to read and process data with a microcontroller. An embedded algorithm was designed to measure inspiration and expiration time. The values captured were sent via WiFi to a server, for posterior evaluation by the clinician. The key findings of this study are as follows (1) a respiratory-rate remote monitoring system was developed, displaying values calculated from impedance pneumography signals, (2) the correlation of the respiratory rate values from a patient during exercising and resting time, measured by a physician and by the device, was 0.96. (3) when analyzing separately the data obtained by the resting test and the exercise test, the performance of the device presented an average error percentage of -5.36% and +1.97%, respectively. As a conclusion, this device has practical applications for acute and chronic respiratory diseases, where respiratory rate is an indicator of the progression of these conditions.

Keywords*—*Covid-19; respiration rate; internet of things; vital signs; hardware

I. Introduction

At the beginning of the COVID-19 pandemic, 80% of patients were mild cases, meaning they required outpatient care [1]. However, unmonitored patients quickly developed acute respiratory reactions often overlooked by physicians [2]. Constant follow-up on a high number of patients is a challenging task for public healthcare due to the shortage of human resources, as well as the lack of equipment or technology intended for this purpose, especially in low- and middle-income countries.

The traditional way of measuring vital signs requires face-to-face evaluation at a distance of less than one meter, contrary to the preventive measures established to reduce the spread of COVID-19. Due to this, healthcare professionals on the frontline were forced to use personal protective equipment (PPE), while still being exposed to a high rate of contagion, which according to initial estimates represented 10-20% of all diagnoses from each nation [3]. Moreover, the direct consequences of the frequent use of PPE -such as eczema and blisters-, as well as the social stigma of healthcare professionals, considered “agents of contagion”, should not be ignored. This scenario was seen in cities like Italy, Brazil, and Peru, where healthcare systems collapsed due to the high demand for outpatient care for cases of COVID-19 [4,5].

To address some of these challenges, the utilization of the Internet of Things (IoT) remains a promising solution. The Internet of Things (IoT) is based on establishing a connection between the Internet and electronic devices to achieve certain goals. IoT applications are increasingly implemented in different areas such as medical care settings as are the case with wearables, agricultural monitoring, or even transportation to control traffic [6].

Following on these applications, the Internet of Medical Things (IoMT) has gained wide adoption lately due to its multiple advantageous features, such as facilitating disease and drug management, improving treatment methods and patient experience, and reducing clinical expenses of various kinds. These systems require adequate attributes for effective performance: security, adaptability, real-time measurement, intelligent decision-making technologies, and compliance with information management regulations [7].

Respiratory rate is a prognostic factor that could be crucial for early respiratory risk detection [2]. Its consistent and reliable monitoring would allow for a more effective control of the patient's progress for timely care. The proposed solution is a hardware-based system using Internet of Medical Things (IoMT) technology, capable of capturing respiratory rate measurements through impedance pneumography, which is a technique based on the direct relation between the volume variations in the lungs and the electrical impedance from the electrodes located in separated areas of the chest.

II. Literature Review

In 2018, thirteen medical subfields have benefited from the development of IoMT applications, with neurological, cardiovascular and mental diseases having the greatest presence of this type of solution. Chronic diseases impose enormous costs on the healthcare system. In that manner, IoMT applications for patients suffering these conditions are expected to increase, particularly for those in need of rehabilitation. 22% of IoMT applications are commonly used to monitor, manage, and treat cardiovascular disease [8]. The findings indicates that IoMT applications have hardly been developed and implemented for diseases with a high mortality rate, such as chronic respiratory infections [8]. One example of this is Whoop [9], a wrist fitness tracker that collects patient data for an analysis of the change in the patient's respiratory system to prevent respiratory episodes.

However, some devices have been utilized in other medical fields. For instance, Sarmah et. al. [10] developed a system which consists of monitoring patients with IoT hardware in conjunction with a deep neural network to aid in the diagnosis of heart disease. Additionally, Dese et al. [11] developed LVital, a device for monitoring vital signs of pregnant women including blood pressure, heart rate, and body temperature. The device has the notable benefit of monitoring maternal vital signs and is appropriate for those located in low-resource environments.

In the context of the pandemic, IoMT initiatives emerged to combat COVID-19. Ashraf et. al. [12] reported the monitoring of a set of physiological parameters (temperature, heart rate and respiratory rate) embedded in a wearable to map people with this infection. Similarly, Amachi-Choqque et. al. [13] developed an IoT system that measures heart rate, blood oxygen saturation and body temperature of patients with preliminary diagnosis of coronavirus, and then transmits data via WiFi to an IoT platform. The technology establishes the different levels of severity of the signs using a web-based system, in which the physician visualizes statistics and graphs of the different signals, sending alerts to an email or a smartphone.

The background above indicates the extensive interest in monitoring physiological parameters remotely, even more so with the endemization of COVID-19 in the following years. In this study, a device focused on respiratory rate, a vital sign of great relevance to the progression of the coronavirus infection, is proposed. This parameter is paramount to provide early warning of any respiratory episode arising, in the course from moderate to severe phase, which can result in dire consequences for the patient when not treated at an initial stage. This solution helps to ease the workflow in healthcare systems, aiming to reduce the rate of hospitalization in the Intensive Care Units (ICU) due to acute respiratory complications.

III. Design and Development

For the development of this electronic device, the project was divided into the following stages using the deterministic design method:

A. Establishment of Functional and Non-Functional Requirements

Functional requirements are understood as those operational aspects that the solution must perform throughout its execution. Meanwhile, Non-functional requirements are those qualities of the device that are observable at the time of its execution. The established requirements are:

Functional requirements (FR):

- The device must be able to read the electric potential difference through electrodes placed on the chest.
- The device must amplify analog signals to the microvolt level.
- The device must have the capacity to convert analog values read into digital ones.
- The device must transmit the processed data to a server.

Non-Functional requirements (NFR):

- The device must have a shape that allows its placement in the patient's body.
- The device must have an on and off switch.
- The device must have indicators of operating modes.

B. Selection of Electronic Components

Once the functionalities that the device will have been established, the electronic components were determined, being the main ones those mentioned below:

1) ADS1292R: It is a 24-bit multi-channel analog-to-digital converter (ADC) with a built-in programmable gain amplifier (PGA) that incorporates all the features commonly required in low power [14] portable medical devices. For this work, the ProtoCentral module that contains this integrated circuit was used (Fig. 1).

Fig. 1. Protocentral ADS1292R Module.

2) ESP32-WROOM-32: It is a powerful module containing a microcontroller with Wi-Fi and Bluetooth communication capabilities that are used in a wide variety of applications, ranging from low-power sensor networks to speech encoding. [15].

3) XC6203: A highly accurate, low-power, 3-terminal positive voltage regulator, providing up to 400 mA current with a significantly small dropout voltage [16]. In this design, the 3.3 V output version was used.

4) Li-Po 522040: A rechargeable lithium polymer battery with a capacity of 360 mAh (Fig. 2).

Fig. 2. 3.7 V LiPo Battery.

5) RGB Led: A LED with the ability to change color based on the 3 channels (red, blue and green) that will allow us to indicate the different phases of operation of the device.

C. Circuit Schematic Design

For the power stage of the circuit, the XC6203 voltage regulator was used, which converts the LiPo battery voltage from 3.7 volts (Vin) to 3.3 volts (V33) (Fig. 3). The datasheet of the regulator also indicates that both the input and the output must have 1 uF ceramic coupling capacitors. In addition, the output must have a 10 uF electrolytic capacitor. In this case, a tantalum capacitor is being used.

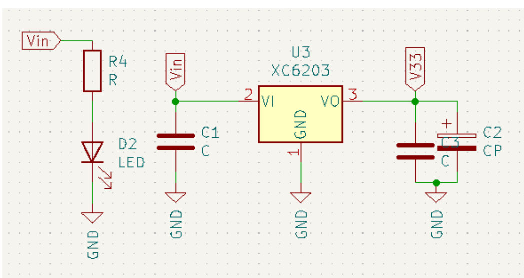


Fig. 3. Circuit Power Stage with Voltage Regulator XC6203.

Once a voltage of 3.3 V is obtained, this will be the stable level for all other active components. The communication between the ESP32 and the ADS1292R module is through the SPI protocol that uses the following pins: MISO, MOSI, SCK, CS, VCC and GND. Additionally, they have control pins (DRDY, START and PWDRST). 3 GPIOS were used to control the colors displayed by the RGB LED (Fig. 4).

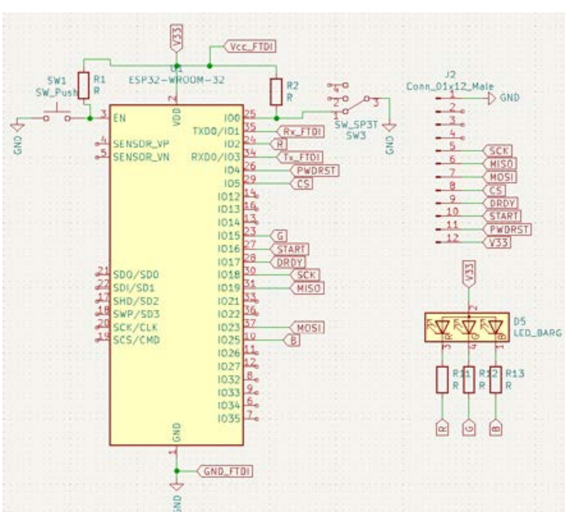


Fig. 4. Connection between ESP32 and ADS1292R Module.

Seeking to optimize the physical size of the device, a custom printed circuit board (PCB) was designed with the previously described circuits including a port to program the microcontroller with an FTDI module. On and off buttons were added as well (Fig. 5).

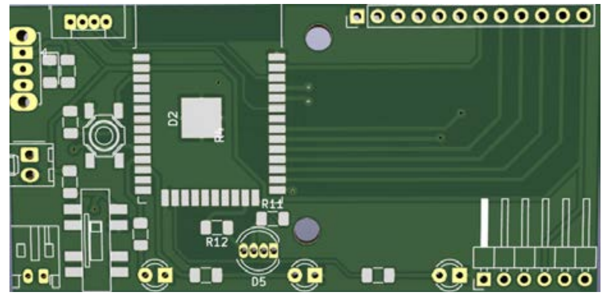


Fig. 5. Custom Printed Circuit Board (PCB).

To give the device wearable characteristics that make it comfortable on the body, an enclosure was designed to be 3D-printed using PLA-type filament. The bracelet-shaped device can be worn on the arm with an elastic band with Velcro patches (Fig. 6).

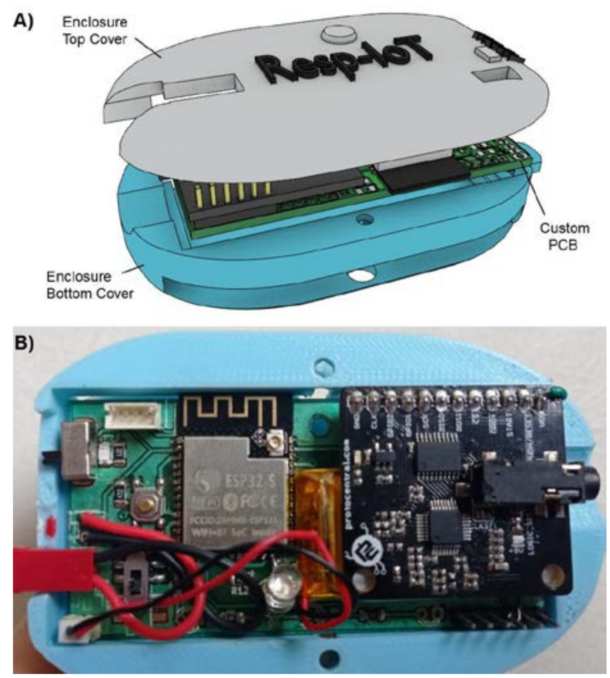


Fig. 6. Enclosure to Protect the Circuit. (A) 3D Design Rendering of the Enclosure. (B) PCB Placed Inside the 3D Printed Enclosure.

D. Respiratory Rate Measurement Algorithm

The first step in the script is to initialize the data acquisition and communication modules when the value obtained from the electric potential difference of the electrodes (previously amplified by the ADS1292R) is obtained. This signal presented high levels of noise; therefore, it was essential to use a filter to stabilize it and prepare it to be processed. The value obtained is stored in a memory space and an average of the last 10 values is made with a separation of 10 ms. This amount of time between each reading was chosen because 100 ms is less than the duration that an inspiration or expiration phase could have, so it does not affect the real-time reading of the sensor. This average value will allow the detection of increasing or decreasing values. The range of values indicates whether it is in the patient's inspiration or expiration phase (Fig. 7).

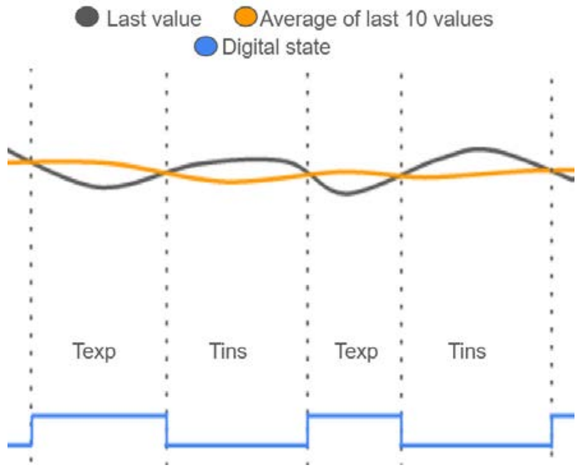


Fig. 7. Inspiratory and Expiratory Phases of the Patient.

The human respiratory cycle is defined as the time interval between the start of inspiration (Tins) and the end of expiration time (Texp). To calculate the duration of the breathing cycle (Tc), the inspiration and expiration phases are timed. The duration time of both phases are added to obtain the cycle time as follows:

Tc = Tins + Texp (seconds)

Once the Tc is calculated, in order to know how many breaths per minute (BpM) the patient performs, the conversion is calculated with the following formula:

BpM =60/Tc (breaths/min)

When the breaths per minute is obtained, the data is verified, as it has limit values that establish that it cannot be greater than 30 cycles/min and not less than 5 cycles/min. When the value obtained is not within that range, the algorithm redirects to the reading of the electrical potential obtained by the electrodes to start the whole process once again. If the value is valid, it is stored and sent via WiFi to the server. The flow of the algorithm can be seen in Fig. 8.

For each of the stages of the algorithm (Fig. 9), the RGB LED lights up a different color indicating the user if the reading is being captured correctly. The color legend is as follows.

E. Experimental Setup

To evaluate the performance of the device, multiple tests were carried out on real patients, comparing the value obtained by the device with values measured by a doctor using a stethoscope. For this purpose, the diaphragm of the device was placed on the chest wall, so that it does not rest on any structure. Then, the number of breaths in 60 seconds was quantified. For this pilot, five healthy patients were recruited, taking measurements in two different scenarios: total rest and after physical effort.

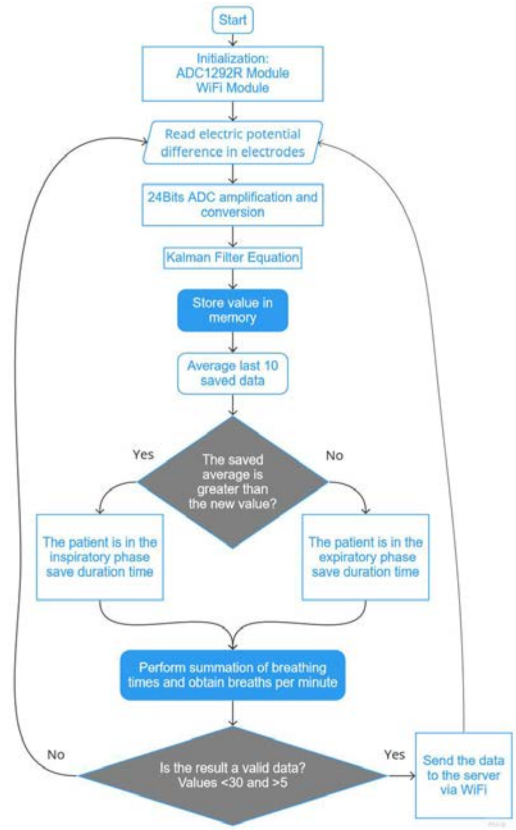


Fig. 8. Flowchart of the Algorithm for the Calculation of Breaths per Minute.

Fig. 9. Color Legend According to Script Execution Stage.

The device is placed on each patient's arm, secured with Velcro to reduce the risk of falls or impacts due to detachment (Fig. 10).

Fig. 10. Device Placed on the Patient.

Fig. 11. Electrodes Placed on the Patient.

Once the device was fixed, the electrodes were placed on the upper part of the patient's chest (Fig. 11). For the tests at rest, measurements were made for a total of two minutes, both by the doctor and the device. For the next test, the patient was instructed to do exercises such as planks or jogging for a total of five minutes with the device on, immediately after which the device was turned on and the doctor began with the quantification of breaths.

IV. Results and Discussion

The results of the tests can be seen in Table I. When separately analyzing the data obtained by the resting test and the exercise test, the performance of the device presents an average error percentage of -5.36% and +1.97%, respectively. For small values, it shows a tendency to overestimate the number of breaths per minute. Meanwhile for high respiratory rate, the measurement tends to be slightly below the real one. When considering both tests, it has an average error value of -1.70%, meaning that the accuracy of the device is approximately 98.3%.

A comparison was performed using the values obtained by the device and by the doctor from both contexts (resting and exercising) as shown in Fig. 12. A correlation coefficient of 0.96 was obtained.

Numerous initiatives have sought to measure the respiratory rate through electronic devices. Adapting sensors to masks could be one strategy, as was demonstrated by Huang et. al. [17]. By using a temperature sensor, the device presented an insignificant error according to its results. However, its precision was evaluated by contrasting its performance with an uncalibrated belt-shaped device that measures thoracic diameter variation.

A similar design is presented by Güder et. al. [18], who introduced a carbon-electrode based paper sensor that varies its conductivity along with the different respiratory phases. To validate its performance, an analysis of the breathing cycles per minute between a light and vigorous exercise session was conducted without selecting a gold standard. Moreover, the sensor electrodes were prone to breaking if the paper was folded, which complicated the process of putting on and taking off the mask.

Different projects have tried to measure respiration rate through chemical sensors that can be mounted on the skin [19] and flexible electronic skins [20], showing promising results. However, it presented an uncomfortable user experience, as the sensors should be placed on the space near the mouth and under the nose. Min et. al. [21] reported a device for respiratory rate measurement analogous to the proposed design. Despite calculating the lung's volume variation, it used a textile-based capacitive respiration sensor, comparing its performance to a nasal thermocouple. The strap surrounding the entire torso demonstrated a precision comparable to the proposed device ($R^2 = 0.98$).

Finally, Pegan et. al. [22] used a wrinkled platinum strain sensor, placed on a part of the chest with a flexible PCB, obtaining a similarity of 95% when compared to a spirometer, which is reliable up to 1000 cycles. Among the advantages of this device is the ability to measure respiratory rate based on the volume of the rib cage, without wrapping the entire torso or causing discomfort near the mouth or nose. One of the disadvantages is the short electrode life cycle, needing to change the electrode each five measurements, and adding conductive gel to improve the quality of the signal.

TABLE I. Comparison of Values Obtained by the Device and by Physician during the Respiratory Rate Test

Test	Patient ID	Device	Physician	%Error
Rest	1	9.49	10	5,10
	2	11,43	11,5	0,61
	3	9,4	9	-4,44
	4	8,7	8	-8,75
	5	12,53	10,5	-19,33
Exercise	1	17,4	20	13,00
	2	23,59	24	1,71
	3	19,72	20,5	3,80
	4	26,1	25	-4,40
	5	21,9	21	-4,29
			Average % Error	**-1,70**

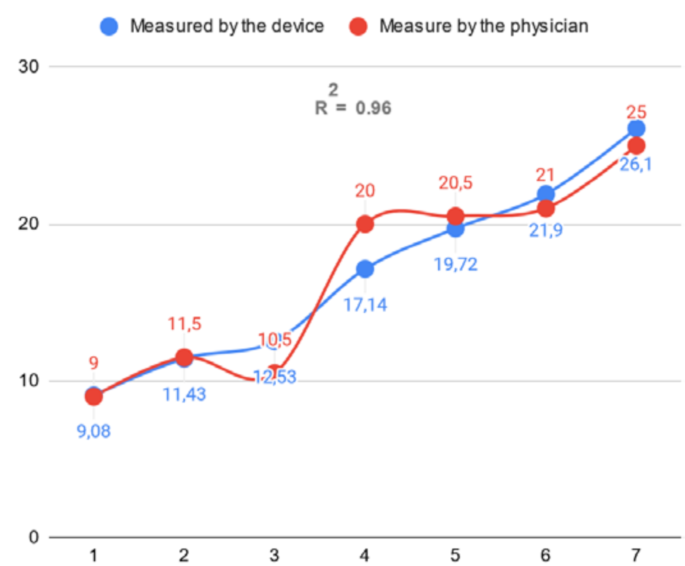


Fig. 12. Relationship between BpM Measured by the Device and by the Physician.

Among the factors that can generate an error in the reading of the proposed device, if the person using it constantly moves or continues to speak, it can generate misreading. It is recommended that patients remain in a state of rest for at least three minutes.

V. Conclusion and Future Work

In this work, an IoT device was developed and tested for remote and continuous monitoring of vital signs related to COVID-19, able to display respiratory rate measurements by calculating the electric potential difference from ECG electrodes located on the chest. The device was calibrated against a visual count performed by a trained healthcare professional in a patient while resting and exercising, demonstrating an accuracy of 96%.

As future work, this device will be tested on a wider population, and at extreme values, through randomized clinical trials with ICU patients and distinct respiratory conditions. It also needs to be calibrated by a metrology laboratory, following ISO standards for measurement management, before being tested in healthcare facilities. As COVID-19 becomes an endemic disease, this device will fill the need for monitoring from conditions like COPD or asthma.

The creation of an open database of respiratory rate values related to mild and severe COVID-19 patients would be of great impact for further research about this disease. Additionally, it will allow the development of machine learning models that anticipate the appearance of a chronic respiratory episode that may aggravate the patient's condition.

Acknowledgment

This project was developed with funding provided by the Royal Academy of Engineering (United Kingdom) through the "Engineering X Pandemic Preparedness" scheme.

References

[1] "Información básica sobre la COVID-19." https://www.who.int/es/news-room/questions-and-answers/item/coronavirus-disease-covid-19 (accessed Mar. 25, 2022).

[2] M. A. Cretikos, R. Bellomo, K. Hillman, J. Chen, S. Finfer, and A. Flabouris, "Respiratory rate: the neglected vital sign," vol. 188, no. 11, p. 3, 2008.

[3] L. H. Nguyen *et al.*, "Risk of COVID-19 among front-line health-care workers and the general community: a prospective cohort study," *Lancet Public Health*, vol. 5, no. 9, pp. e475–e483, Sep. 2020, doi: 10.1016/S2468-2667(20)30164-X.

[4] B. Armocida, B. Formenti, S. Ussai, F. Palestra, and E. Missoni, "The Italian health system and the COVID-19 challenge," *Lancet Public Health*, vol. 5, no. 5, p. e253, May 2020, doi: 10.1016/S2468-2667(20)30074-8.

[5] D. R. Q. Lemos *et al.*, "Health system collapse 45 days after the detection of COVID-19 in Ceará, Northeast Brazil: a preliminary analysis," *Rev. Soc. Bras. Med. Trop.*, vol. 53, Jul. 2020, Accessed: Mar. 25, 2022. [Online]. Available: http://www.scielo.br/j/rsbmt/a/98LMbshKyrXVc7sC4nZwkSG/?lang=en

[6] I. S. Udoh and G. Kotonya, "Developing IoT applications: challenges and frameworks," *IET Cyber-Phys. Syst. Theory Appl.*, vol. 3, no. 2, pp. 65–72, 2018, doi: 10.1049/iet-cps.2017.0068.

[7] F. Alsubaei, A. Abuhussein, and S. Shiva, "Ontology-Based Security Recommendation for the Internet of Medical Things," *IEEE Access*, vol. 7, pp. 48948–48960, 2019, doi: 10.1109/ACCESS.2019.2910087.

[8] F. Sadoughi, A. Behmanesh, and N. Sayfouri, "Internet of things in medicine: A systematic mapping study," *J. Biomed. Inform.*, vol. 103, p. 103383, Mar. 2020, doi: 10.1016/j.jbi.2020.103383.

[9] "Researchers to study if startup's wrist-worn wearable can detect early COVID-19 respiratory issues | TechCrunch." https://techcrunch.com/2020/04/01/researchers-to-study-if-startups-wrist-worn-wearable-can-detect-early-covid-19-respiratory-issues/ (accessed Mar. 25, 2022).

[10] S. S. Sarmah, "An Efficient IoT-Based Patient Monitoring and Heart Disease Prediction System Using Deep Learning Modified Neural Network," *IEEE Access*, vol. 8, pp. 135784–135797, 2020, doi: 10.1109/ACCESS.2020.3007561.

[11] K. Dese, G. Ayana, and G. Lamesgin Simegn, "Low cost, non-invasive, and continuous vital signs monitoring device for pregnant women in low resource settings (Lvital device)," *HardwareX*, vol. 11, p. e00276, Apr. 2022, doi: 10.1016/j.ohx.2022.e00276.

[12] M. U. Ashraf, A. Hannan, S. M. Cheema, Z. Ali, K. m. Jambi, and A. Alofi, "Detection and Tracking Contagion using IoT-Edge Technologies: Confronting COVID-19 Pandemic," in *2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, Jun. 2020, pp. 1–6. doi: 10.1109/ICECCE49384.2020.9179284.

[13] J. Amachi-Choqque and M. Cabanillas-Carbonell, "IoT System for Vital Signs Monitoring in Suspicious Cases of Covid-19," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 2, 2021, doi: 10.14569/IJACSA.2021.0120223.

[14] "ADS1292R data sheet, product information and support | TI.com." https://www.ti.com/product/ADS1292R (accessed Mar. 25, 2022).

[15] "ESP32 Wi-Fi & Bluetooth Modules I Espressif." https://www.espressif.com/en/products/modules/esp32 (accessed Mar. 25, 2022).

[16] "XC6203 Series | Your analog power IC and the best power management, TOREX." https://www.torexsemi.com/products/single-type-voltage-regulators/series/?name=xc6203 (accessed Mar. 25, 2022).

[17] Y. P. Huang, M. S. Young, and K. N. Huang, "Respiratory Rate Monitoring Gauze Mask System Based on a Pyroelectric Transducer," in *2008 2nd International Conference on Bioinformatics and Biomedical Engineering*, May 2008, pp. 1648–1649. doi: 10.1109/ICBBE.2008.741.

[18] F. Güder *et al.*, "Paper-Based Electrical Respiration Sensor," *Angew. Chem. Int. Ed.*, vol. 55, no. 19, pp.

[19] H. Guo, C. Lan, Z. Zhou, P. Sun, D. Wei, and C. Li, "Transparent, flexible, and stretchable WS2 based humidity sensors for electronic skin," *Nanoscale*, vol. 9, no. 19, pp. 6246–6253, May 2017, doi: 10.1039/C7NR01016H.

[20] J. Park *et al.*, "Giant Tunneling Piezoresistance of Composite Elastomers with Interlocked Microdome Arrays for Ultrasensitive and Multimodal Electronic Skins," *ACS Nano*, vol. 8, no. 5, pp. 4689–4697, May 2014, doi: 10.1021/nn500441k.

[21] S. D. Min, Y. Yun, and H. Shin, "Simplified Structural Textile Respiration Sensor Based on Capacitive Pressure Sensing Method," *IEEE Sens. J.*, vol. 14, no. 9, pp. 3245–3251, Sep. 2014, doi: 10.1109/JSEN.2014.2327991.

[22] J. D. Pegan *et al.*, "Skin-mountable stretch sensor for wearable health monitoring," *Nanoscale*, vol. 8, no. 39, pp. 17295–17303, Oct. 2016, doi: 10.1039/C6NR04467K..

Design and Implementation of Teaching Assistant System for Mechanical Course based on Mobile AR Technology

Jinglei Qu[1], Lulu Zheng[3], Yuhui Kang[4]
School of Mechanical Engineering
Henan Institute of Technology
Xinxiang, China

Bingxin Ma[2]
School of Materials Science and Engineering
Henan Institute of Technology
Xinxiang, China

***Abstract*—Augmented reality technology has become a hot spot in many fields because of its unique real-time interaction and the ability to add virtual objects in 3D video space. To enable students better understand and master traditional mechanical courses and solve the limitations of the courseware teaching method, a teaching assistant system based on Vuforia platform is designed and developed by combining augmented reality technology with existing teaching methods. On the smartphone application side, the 3D model of parts can be displayed by scanning mechanical engineering drawings, which can rotate, zoom, and section view the model. It realizes 3D visualization and interactive operation of 2D drawings in textbooks and achieves the purpose of image, diversity and efficient teaching. The results show that this method can not only promote the cognition of spatial relations based on visualization, but also create a situational learning environment based on experience, which can effectively improve the learning effect of courses and enhance students' interest and enthusiasm in learning.**

***Keywords*—Augmented reality; teaching assistance; mechanical teaching; visual operation**

I. Introduction

The rapid development of the new generation information technology has promoted the diversification of teaching resources. Improving teaching quality and learning efficiency is the basic requirement of modern education [1]. As an important factor to improve teaching quality and learning efficiency, teaching resources have attracted more and more attention [2-3]. The traditional teaching methods of the plane textbooks have obvious limitations. At present, some innovations have been made to the teaching media, such as using 2D codes and other methods to link to assistant resources on web pages. However, this form is generally not well accepted by students, and it is necessary to find a way to solve this problem that is more attuned to the needs and habits of today's students. The courses of mechanical specialty mainly study mechanical parts, mechanisms, mechanical transmission and other related contents, which have the characteristics of many knowledge points, many concepts and difficulty in understanding [4-5]. However, students seldom have the opportunity to contact related equipment and parts during their school years, which brings certain difficulties to students' learning. Even if students search through the Internet, the existing teaching resources do not have many 3D solid models related to textbooks, so students need to have a strong sense of space.

Augmented Reality (AR) register 3D information in a real environment [6-7]. It superimposes computer-generated virtual information into the real world so that virtual enhanced information and real scene information coexist, complement, and superimpose with each other, thus strengthening users' senses and cognition of reality [8]. With the development of mobile technology, the popularity of smart phones and tablet computers has provided a solid material foundation for the wide application of mobile AR technology in many fields, showing great potential in military, medical visualization, education, construction machinery, book design and so on [9-11]. Based on the teaching needs of mechanical discipline and the purpose of improving the current teaching methods, this paper designs a teaching assisted learning system based on mobile AR technology.

In order to explore the influence of mobile AR technology on mechanical teaching effect and to expand and supplement digital teaching means, in this study, an AR assisted teaching system is designed first, and a teaching experiment is designed around AR system, which is verified and analyzed from three aspects: software function, teaching effect and influence on teaching. The results of this study can be used as a reference for teaching researchers using augmented reality.

II. Literature Review

A. Augmented Reality

AR is developed from virtual reality [12]. It realizes the effective combination of the real world and computer-generated virtual objects. By superimposing the computer-generated virtual objects on the real scene in real time, users can feel the real and virtual scene information at the same time, so as to realize the combination of virtual and real. AR technology has been applied in many fields, such as aerospace, military, education, medical treatment, video games, manufacturing and so on.

AR technology has three main characteristics: it combines the real and the virtual, it interacts in real-time, and can be executed in three dimensions [13]. The core of AR technology is the rapid recognition and spatial positioning technology. At present, the mainstream AR recognition applications are mainly Google's ARCore, Apple's ARKit and Qualcomm's Vuforia. The comparison of the three development engines is shown in Table I.

TABLE I. COMPARISON OF COMMON AR DEVELOPMENT ENGINES

Development engine	Functions	Operation platform
ARCore	Motion tracking, environment understanding, light perception	Android
ARKit	Multiple face tracking, motion capture, and real-time augmented reality	IOS
Vuforia	Image recognition based on natural feature points, extended tracking, model target	IOS & Android

Mobile AR is a branch of AR technology. Traditional AR systems mostly take wearable external devices as the carrier, which is difficult to carry and difficult to maintain. Due to the continuous development and update of mobile terminal devices, mobile terminals have powerful processors, high-definition cameras and convenient touch screen operation [14]. These functions provide a convenient platform for the development of AR systems for mobile devices. On the basis of inheriting the key technologies and core problems of traditional augmented display technology, mobile AR technology combines the portability and mobility of mobile devices, greatly promotes the development of mobile AR applications, and makes the application field broader and the interaction means more humanized.

B. Augmented Reality in Education

This interactive means based on the real world and enhanced by virtual data provides a new teaching method for educators, which has a strong guiding significance for the abstract content teaching. Through AR technology, learners can enter an environment that is highly simulated with reality and dynamically learn in vision and interaction.

At present, the application of AR technology in the field of education is gradually expanding. Colleges and educational institutions are committed to promoting the deep integration of information technology with higher education and vocational education. AR technology will bring a positive impact on teaching content, models and learning methods [15].

AR technology can create an intelligent learning environment to adapt the educational content to students' personalized learning. To improve students' practical skills, based on emphasizing the theoretical knowledge in textbooks, AR technology is used to show the relevant knowledge in an information-based 3D way. Literature [16] research shows that adopting AR technology assisted teaching can improve students' learning efficiency and information understanding ability. Literature [17] uses AR technology in creative design courses to improve the learning motivation and creativity of students. Literature [18] developed an AR mobile application for engineering students, and the statistical data indicate that the application positively affects learning. An interactive teaching system based on mobile AR technology is designed and implemented, the case study shows that stimulate students' spatial imagination and learning interest [19]. [20-23] show that AR helps students improve their skills and knowledge in physics, biology, mathematics, art and architectural topics. A large number of empirical studies show that augmented reality technology has great potential and application prospects in learning support and teaching.

C. Advantages of using Augmented Reality in Education

1) Rich interactive means to create mobile learning environment: AR emphasizes the combination of virtual and real, superimposing virtual objects on the real world, which is a supplement to the real world. It allows users to see the real world. For some abstract cognitive knowledge, AR can transform it into specific situations. Different knowledge can shape different situations, so that students can obtain learning knowledge more directly.

Traditional teaching methods are generally divided into two links: classroom teaching and after-school practice. When students encounter difficulties, they often have to understand and digest them by themselves which is also difficult to solve problems in time. Moreover, AR can realize mobile learning and break the restrictions of teaching conditions. Even if not in class, students can immerse themselves in the learning situation again through mobile AR technology. It solves the blind area of understanding caused by the unclear expression of some teachers [24]. Through electronic devices and books, students can learn anytime and anywhere, realizing seamless and continuous mobile learning.

2) Stimulate learning interest and reduce cognitive difficulty: In many university basic courses, there are some knowledge that is difficult to describe in words. For example, engineering drawing of mechanical major and some courses of other majors that require students to carry out spatial imagination. For some students whose space imagination is not strong enough, this undoubtedly hinders their learning. For some students with strong space imagination, even if they can imagine most of them, some details are still unimaginable, which will make students feel frustrated without timely answers. AR technology can directly and vividly express the abstract things in three-dimensional form. Through the combination of virtual and real, it is more conducive to students' understanding and memory, improves students' autonomy and reduces the requirements for learners' cognitive transfer ability. This interactive way greatly improves students' learning participation, meets students' curiosity, and makes learning simple and natural.

3) Deepen immersive learning and strengthen knowledge cognition: In the classroom, teachers should reflect the application scene of knowledge as realistically as possible and be related to the social practice scene of human beings. Mobile AR technology organically combines virtual and real, which makes the cognitive scene closer to the effect carried out outdoors. Compared with traditional teaching resources, the biggest advantage of AR teaching is reflected in the constructed immersive learning environment [25-26]. The three-dimensional immersion teaching created by AR enhances learners' sense of experience, enables students to learn more quickly, further deepen their cognition of knowledge and achieve in-depth learning.

In the existing literature, some researchers have noted the limitations of AR in the field of education. Traditional AR systems are mostly based on wearable external devices, which are difficult to carry and maintain. Without a well-designed student interface and guidance, AR technology can be too complex to use. The various devices that provide AR applications can cause other technical problems. In this paper, mobile AR is used as the technical support to develop the teaching assistance system for engineering drawing courses. Mobile devices such as mobile phones and tablet computers are used to design and develop the teaching assistance system. Teachers can quickly deploy AR content, which is simple, stable and has high platform support.

III. Proposed Teaching Assistant system

A. Overall Design

The teaching assistant system based on mobile AR technology is mainly used to assist teaching. Its purpose is to strengthen students' imagination of 3D graphics and improve students' thinking conversion ability between 3D space and 2D plane. The system consists of three modules: information input module, fusion display module and information interaction module, as shown in Fig. 1.

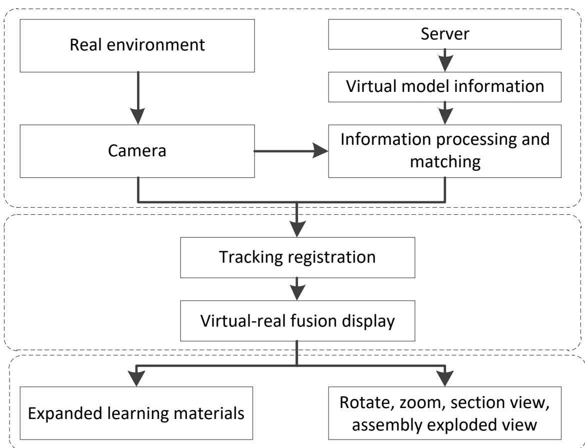


Fig. 1. Overall Design of the Proposed System.

The information input module collects the feature point information of engineering drawings in the real environment through the camera, then processes and matches them to obtain the virtual model information on the server side. The fusion display module tracks the real-world position of virtual model in real-time according to the matching and identified information from information input module, and performs superposition fusion display of real and virtual information. In the human-computer interaction module, the user interacts with the teaching assistant system through interface touch control and can perform operations such as rotation, zooming, section view of parts and explosion diagram display of assembly parts. At the same time, the user can view relevant parts introduction, network reference materials and other expanded information.

B. Development Environment

The system uses Unity3D as the development platform, builds 3D models through SolidWorks software, and uses AR technology to realize the interaction between touch instructions and the 3D model nested in the screen through Vuforia AR Toolkit. The software running system is Android. The specific development platform and tools are shown in Table II.

TABLE II. Development Platform and Tools

Tools	Functions
Unity3D	Rendering Interaction and Software Development Platform
Android NDK	Data Communication between Android and Unity3D
Vuforia SDK	Software Development Toolkit
XAMMP	Database
SolidWorks	3D modeling
Deep Exploration 6.3	3D model format conversion

C. Development Process

The implementation of the system mainly includes three steps: Creating a database; Objective management; Resource integration and publishing. The system implementation flow is shown in Fig. 2.

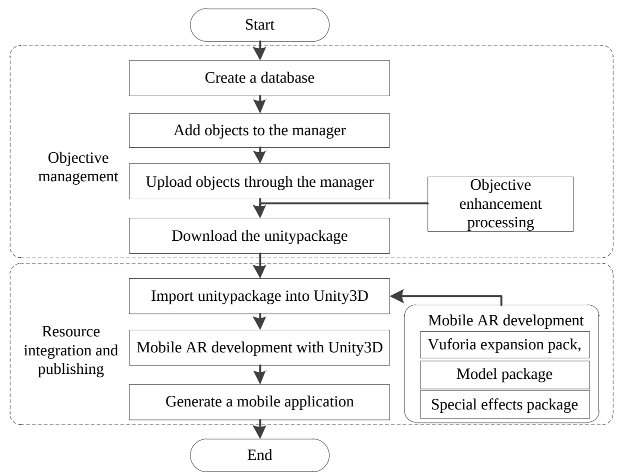


Fig. 2. System Implementation Flow.

1) Create a database.

2) Objective management. Create a 3D model, add and upload identification point data, model data, references and other data to the objective manager.

3) Target enhancement processing. Improve the recognition effect through data enhancement.

4) Setup of the development platform. Importing Vuforia AR Toolkit, model package and special effects package into Unity3D, and setting a series of initial values such as camera direction and lighting system.

5) Unity3D mobile AR development. Write Unity3D camera identification code and C# script for model controlled interaction.

6) Generate mobile application. The Unity3D project is exported as an Android project. The secondary development of mobile APP is carried out in Android Studio include complete the UI design, cache management module and communication module.

D. Implementation of Key System Functions

1) Fusion display function: The 3D modeling software SolidWorks is used to model the display parts. Because of the different recognition formats of each software, the SolidWorks 3D model is converted into .fbx file through DeepExploration software and imported into Unity3D software. Register on Vuforia website, create a cloud database and manage markers. In this system, pictures are used as markers, and the recognized objects of image target are set as 3D models of parts. The image implementation mechanism of Vuforia is completed by detecting the matching of natural feature points on the image. The feature point data of the picture or the image in the target manager is collected and then stored in the database. When the application runs, the feature points of the image are detected in real-time and the feature points of the picture stored in the data center are matched. After identifying the corresponding results, adjust the model size, position and light illumination with the predefined configuration, and then display them on the screen. The recognition effect of the display function is shown in Fig. 3.

2) Rotate, zoom and section function: The zoom is realized by the localScale() function. Firstly, the initial positions oPos1 and oPos2 of the fingers are recorded. When at least one finger moves on the screen, the real-time positions tPos1 and tPos2 of the fingers are checked. Finally, the model is enlarged or reduced by comparing the distance between the starting and ending positions where the fingers stay. The effect of model scaling is shown in Fig. 4, where the aperture represents the finger movement path.

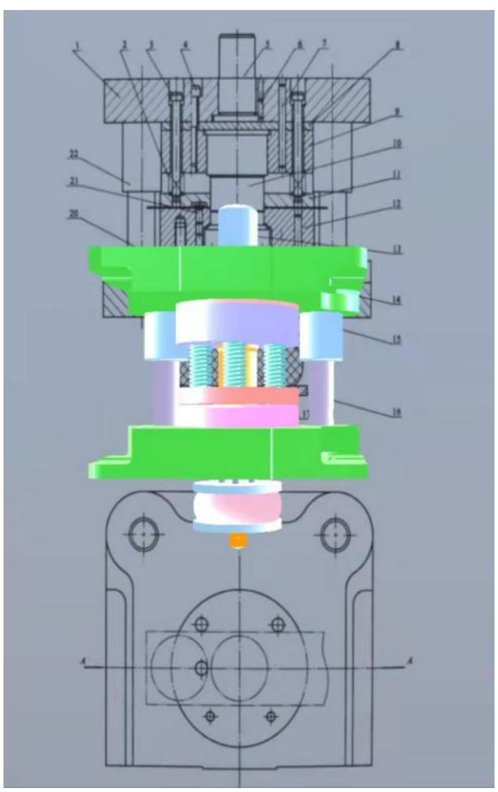


Fig. 3. The Recognition Effect of the Display Function.

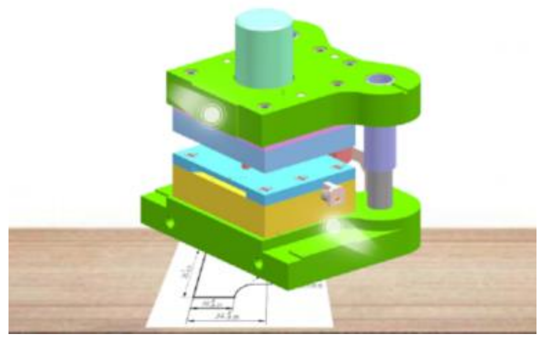


Fig. 4. Model Scaling Implementation Effect.

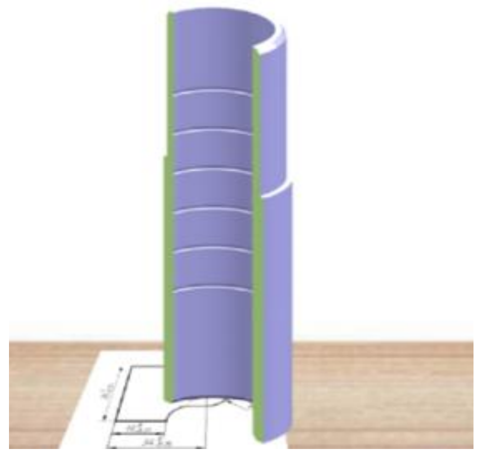


Fig. 5. Section View Effect.

Since section display cannot be directly carried out in Unity3D, it is necessary to set "true/false button assembly" to convert the overall model to the section model by establishing another parallel and independent section 3D model. The effect achieved in section view is shown in Fig. 5.

IV. Experimental Design

In order to verify the influence of the AR teaching assistant system on the teaching effect of mechanical drawing course, 154 undergraduates from Henan Institute of technology were selected as research participants. According to the large class teaching, 154 students were divided into experimental class and control class. The control class adopts the traditional teaching mode. First, the students are required to preview the knowledge before class. During the teaching process, the teachers teach the knowledge through books and PPT, and answer the questions raised by the students. Smart phones are not allowed to be used in the course of class. The experimental class adopts the AR teaching assistant system to carry out teaching activities. The specific steps are knowledge Preview - Teacher Explanation - Deep Learning - Question Answering - Opinion Feedback. First of all, students preview around the knowledge points arranged by teachers in advance. When encountering unimaginable models, they can deepen their understanding with the help of AR teaching assistant system. Teachers use teaching strategies that match technology and content to teach. Then, students enter the deep learning link.

Through the four links of cognition, construction, transfer and application, students can independently understand knowledge, and then solve problems, explain the problems that students haven't digested. Finally, students put forward the deficiencies in the teaching link, so as to promote better teaching improvement next time.

In order to better understand the effect of AR teaching assistant system in the process of mechanical drawing teaching, the teaching effect is analyzed by score analysis and questionnaire survey. At the end of the semester, the teaching effect is judged by analyzing the results of the two classes. Each student in the experimental class is required to fill in a questionnaire. From the two dimensions of learning cognition and product experience, the questionnaire sets up five aspects: exploratory learning, learning interest, User satisfaction, resources abundant degree and students' willingness to use. The questionnaire settings are shown in Table III.

TABLE III. CURRICULUM SATISFACTION QUESTIONNAIRE

Category	Questionnaire
Learning cognition	1. Using AR teaching assistant system in the classroom is more conducive to exploratory learning
	2. Using AR teaching assistant system can improve the interest in learning engineering drawing
User experience	3. The function of AR teaching assistant system meets the teaching requirements
	4. AR teaching assistant system is abundant in teaching resources
	5. I am willing to use AR teaching assistant system to assist learning

V. EXPERIMENT AND RESULT ANALYSIS

A. *Practical Application of AR Teaching Assistant System*

Through AR technology, image recognition and stereo presentation methods, this system displays stereo effects through 3D modeling, directly converts the plan into 3D objects, and enhances the interactivity in the learning process.

Taking the mechanical drawing course as an example, the application scenario of this system is introduced. Students scan the engineering drawings by opening the App on the mobile phone, and the 3D model of the displayed part is fused on the screen, as shown in Fig. 6. After the mobile phone screen is loaded with the 3D model, even if the camera is moved to other positions, the 3D model can still stay and be displayed, as shown in Fig. 7. On the mobile phone screen, students can rotate, zoom, and section model through gestures. Through 3D display from various angles, students can be helped to analyze and understand the correlation between 2D engineering drawings and 3D entities model, thus improving the learning effect.

In the process of classroom theoretical teaching, by setting specific identification pictures, the virtual 3D model is superimposed on the real teaching scene, and various parts and mechanisms are presented in front of students. Let the students observe the movement process of the 3D structure or principle of the mechanism from a close and all-around angle to deepen their understanding of the textbook knowledge.

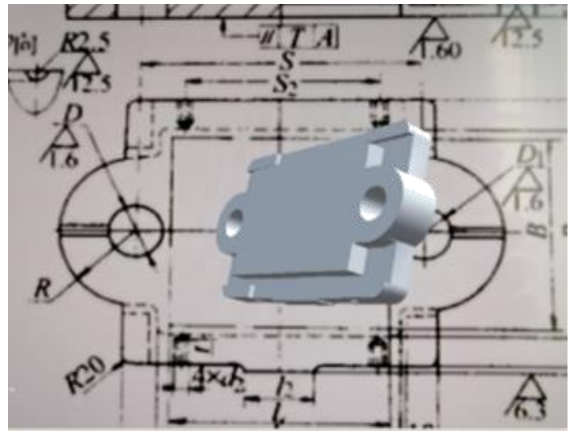


Fig. 6. Scan Drawings to Obtain 3D Models.

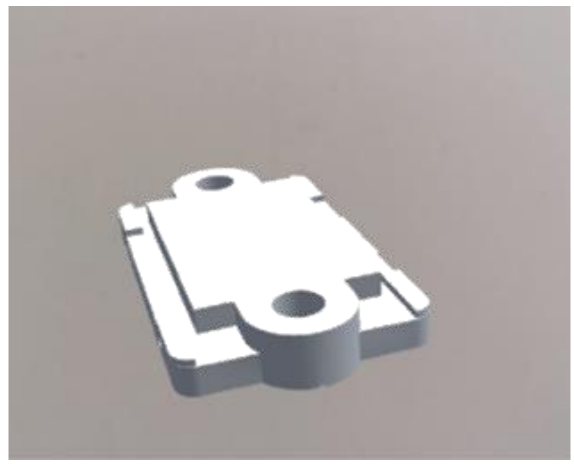


Fig. 7. Effect of Removing Background Drawing.

B. *Teaching Effectiveness Analysis*

The teaching effect of mechanical drawing course in the spring semester of 2020-2021 school year is shown in Table IV. There are 78 people in the control class and 76 people in the experimental class. The control class adopts the traditional paper teaching mode, while the experimental class adopts the AR assisted teaching system. Comparing the final examination results of the two classes, the average score of the experimental class is 84.24, and the pass rate is 100%, which is better than that of the control class. The Standard deviation of the experimental class is 9.17, which is less than 15.31 of the control class, indicating that the distribution of students' score is more concentrated. Through the statistical analysis of teaching effect, it can be concluded that after using AR assisted teaching, the results of students in the experimental class are better than those in the control class. Through the questionnaire survey, we know the students' recognition of AR assisted teaching, and most students are satisfied. Most of the gains of students focus on the enhancement of autonomous learning ability, the deepening of classroom participation, the solid mastery of knowledge, the improvement of problem-solving ability and so on. From students' feedback, it can be found that students have high recognition of AR assisted teaching mode.

TABLE IV. TEACHING EFFECT OF AR ASSISTED TEACHING

Index / Group	Fractional segment (Grade)					Average score	Pass rate	Standard deviation
	[100-90]	(90-80]	(80-70]	(70-60]	(60-0]			
Experimental Class (76 students)	27 35.5%	27 35.5%	16 35.5%	6 35.5%	0 0%	84.24	100%	9.17
Control Class (78 students)	5 6.4%	22 28.2%	25 32.1%	13 16.7%	13 16.7%	72.35	83.33%	15.31

TABLE V. DATA ANALYSIS OF CURRICULUM SATISFACTION QUESTIONNAIRE

Questionnaire Evaluation	Very satisfied	Generally satisfied	Dissatisfied	Very dissatisfied
1. Using AR teaching assistant system in the classroom is more conducive to exploratory learning	42	21	8	5
2. Using AR teaching assistant system can improve the interest in learning engineering drawing	40	24	10	2
3. The function of AR teaching assistant system meets the teaching requirements	44	24	7	1
4. AR teaching assistant system is abundant in teaching resources	34	25	13	4
5. I am willing to use AR teaching assistant system to assist learning	46	22	8	0

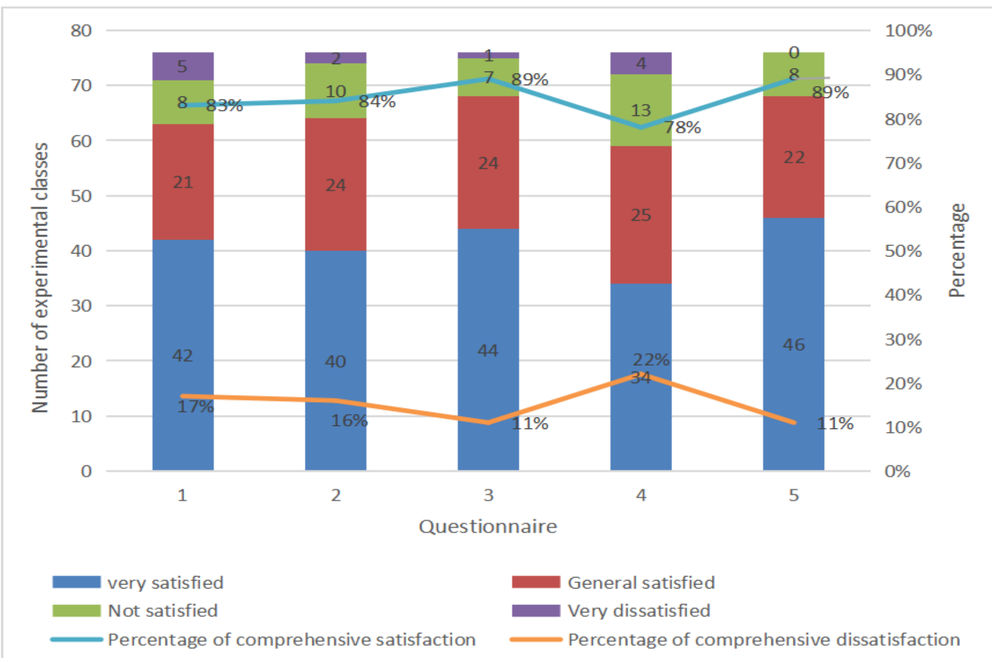


Fig. 8. Questionnaire Statistical Analysis Results.

After the examination, a questionnaire was distributed to the students in the experimental class. See Table III for questions 1-5 in the questionnaire. For the five questions in the questionnaire, students can choose four grades: very satisfied, generally satisfied, dissatisfied and very dissatisfied. A total of 76 questionnaires were distributed, and 76 were recovered, with a recovery rate of 100%. After inspection and verification, 76 valid questionnaires were found, with an effective rate of 100%. The statistical results of the questionnaire are shown in Table V. In terms of improving exploratory learning ability, 42 students expressed very satisfied and 21 students were generally satisfied, while only 8 and 5 students chose dissatisfied and very dissatisfied respectively. It shows that students agree that AR teaching assistant system can improve their exploratory learning ability. Similarly, it has been highly evaluated in improving learning interest, satisfaction with the system and usability. However, the richness of software models is slightly lower. Some students believe that the models and resources in courseware and after-school exercises need to be further improved. For more intuitive data analysis, convert the table into a chart, add the very satisfied and generally satisfied people to calculate the comprehensive satisfaction percentage, add the dissatisfied and very dissatisfied people to obtain the comprehensive dissatisfaction percentage, and draw the questionnaire statistical analysis results chart as shown in Fig. 8. It is obvious from Fig. 8 that the AR teaching assistant system has been recognized by most of the students in the experimental class.

In the process of questionnaire survey, the author also interviewed the teachers. The instructor said that most students had not been exposed to AR before the experiment, but thanks to the popularity of smart electronic devices such as smart phones, students' information literacy was relatively high, so the installation and use of teaching assistance system were relatively smooth. By using the assistant learning system, students can understand knowledge more easily, their interest and confidence in learning have been improved, the frequency and efficiency of classroom teaching interaction have become higher, and the teaching quality has been significantly improved compared with the traditional teaching mode.

C. Influence of AR on Engineering Drawing Teaching

Contemporary college students have many new characteristics. As a generation growing up in the digital environment, they can skillfully use electronic products and have a strong curiosity about new things. As a popular field in recent years, AR technology is introduced into teaching to create mobile learning environment and more opportunities for autonomous learning for students. It will also arouse students' interest, make students understand knowledge more easily and learn happily.

Integrating AR technology into some courses with higher learning difficulties will make students' learning easier and make the classroom more efficient. From another point of view, it will also reduce teachers' teaching pressure and make teachers easy to teach. The setting of every model and scene in augmented reality is carefully considered by teachers, which has certain demonstration, typicality and guidance. Students can study anytime and anywhere, reducing the disadvantages of being unable to concentrate on their study caused by external conditions. Compared with traditional learning methods, under the same time investment cost, students using AR technology assisted learning can immerse themselves in deep learning faster, so as to achieve better learning effect.

In the future, AR assisted instruction technology can be further extended to the process of experimental teaching. Due to the lack of experimental resources, colleges and universities have limited laboratory opening time, places and number of experiments which leads to the unsatisfactory effect of students' hands-on operation. Some mechanical equipment is expensive and dangerous, so it cannot be carried out in reality. Using AR technology can carry out some experiments that cannot be realized or dangerous in the real environment to solve this problem. AR technology is used to convert the required experimental equipment and instruments into digital models, import and create a virtual database, and realize virtual simulation experiment operation, which can provide students with a more flexible experimental environment and course arrangement while reducing consumables and maintenance costs. Moreover, through virtual operation, students can more accurately understand the experimental principle and content, and carry out design experiments to realize the differentiated cultivation of students, so as to implement the educational concept of "student-centered".

VI. Conclusion

AR technology can realize the superposition and supplement of virtual enhancement information and real scene information, which provides a new way for the current mechanical course teaching and development. In view of the strong practicality of mechanical courses, the traditional teaching method is too simple, and it is difficult for students to establish stereoscopic impression. A mechanical course teaching assistant system based on mobile AR technology is designed and developed. Through the integration of AR technology and traditional paper teaching materials, the 3D visualization and interactive operation of parts and mechanisms in teaching materials are realized to achieve the purpose of diversity and efficient course teaching.

According to the experimental results and questionnaire analysis, the proposed system has a beneficial impact on students' learning effect and learning motivation. The academic performance of the experimental class is significantly higher than that of the control class. Because the AR based teaching method is more likely to improve students' attention and interest, the questionnaire results show that students are more satisfied with the AR teaching method and expect to have a more functional and perfect software system. The above results show that introducing AR technology into mechanical course can make up for students' lack of practical exercise. Through the combination of AR technology and mechanical course teaching, students can better understand and master the principle of mechanical structure, manufacturing and assembly process, use and operation, and promote the cognition of spatial relationships based on visualization. This teaching method can stimulate students' learning enthusiasm and improve innovative ability and engineering consciousness.

Information technology has a revolutionary influence on the development of education, has become a new wave of educational reform, and provides new ideas for the further development of teaching and learning. This paper is a practical exploration of the application of AR in the field of education. The mobile learning mode based on AR technology has the advantages of interactivity, mobility and sharing that traditional learning does not have. The auxiliary mobile learning platform based on AR technology still has a lot of development space. In the next step, it can be studied and developed in multi-disciplinary and interactive functions.

Acknowledgment

This research was supported by Key Scientific and Technological Project of Henan Province (222102210189, 212102210324, 202102210286), Industry-university Cooperative Education Program of Ministry of Education, (202002029035), Key Scientific Research Projects of the Higher Education Institutions of Henan Province (22B460004), Educational Reform and Practice project of Henan Institute of Technology (2021-YB024) and Doctoral Fund of Henan Institute of Technology (KY1750).

REFERENCES

[1] S. Morales, L. Andrade-Arenas, and A. Delgado, “Augmented Reality: Prototype for the Teaching-Learning Process in Peru,” International Journal of Advanced Computer Science and Applications (IJACSA), vol. 13, pp. 806-815, 2022. [Online]. Available: https://dx.doi.org/10.14569 /IJACSA .2022.0130194.

[2] N. Pellas, P. Fotaris, I. Kazanidis, and D. Wells, “Augmenting the learning experience in primary and secondary school education: A systematic review of recent trends in augmented reality game-based learning,” Virtual Real., vol. 23, pp. 329–346, 2019. [Online]. Available: doi: 10.1007/s10055-018-0347-2.

[3] M. da Silva, J. Teixeira, P. Cavalcante, and V. Teichrieb, “Perspectives on how to evaluate augmented reality technology tools for education: A systematic review,” J. Brazilian Comput. Soc., vol. 25, Art. no. 3, 2019. [Online]. Available: doi: 10.1186/s13173-019-0084-8.

[4] F. Marcel, “Mobile augmented reality learning objects in higher education,” Res. Learn. Technol., vol. 27, pp. 1–10, 2019. [Online]. Available: doi:10.25304/rlt.v27.2133.

[5] E. Ibili, M. Çat, D. Resnyansky, S. Sahin, and M. Billinghurst, “An assessment of geometry teaching supported with augmented reality teaching materials to enhance students’ 3D geometry thinking skills,” Int. J. Math. Educ. Sci. Technol., vol. 51, no. 2, pp. 224–246, 2020. [Online]. Available: doi: 10.1080/0020739X.2019.1583382.

[6] S. P. Liu, W. B. An, k, Han, Z, Y, Ma, L, Zhang, and C. H. Ruan, “Research on multi-level teaching strategy mode based on mobile augmented reality,” Higher Education Exploration, vol. 8, pp. 59-64, 2021.

[7] Z. Chen, Y. Li, and Y. Zhang, “Recent advances in omnidirectional video coding for VR: projection and evaluation,” Signal Processing, vol. 146, pp. 66–78, 2018.

[8] A. A. Eldokhny and A. M. Drwish, “Effectiveness of Augmented Reality in Online Distance Learning at the Time of the COVID-19 Pandemic,” International Journal of Emerging Technologies in Learning (IJET), vol. 16, pp. 198-218, 2021. [Online]. Available: http://dx.doi. org/10.3991/ijet.v16i09.17895.

[9] F. Pettersson, “Understanding Digitalization and Educational Change in School by Means of Activity Theory and the Levels of Learning Concept,” Educ. Inf. Technol. Vol. 26, pp.187–204, 2021. [Online]. Available: https://doi.org/10.1007/s10639-020-10239-8.

[10] G. Miriam, G. Ricardo, I. Orlando C. Michael and José Luis Herrera Salazar, “Mobile Application with Augmented Reality to Improve Learning in Science and Technology,” International Journal of Advanced Computer Science and Applications (IJACSA), vol. 12, 2021. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2021.0121055.

[11] H. Köse, N. Güner-Yildiz, “Augmented reality (AR) as a learning material in special needs education,” Educ Inf Technol vol. 26, pp. 1921-1936, 2021. [Online]. Available: https://doi.org/10.1007/s10639-020-10326-w.

[12] Z. Zhang, Z. Li, and M. Han, “An augmented reality-based multimedia environment for experimental education,” Multimed Tools Appl, vol. 80, pp. 575-590, 2021. [Online]. Available: https://doi.org/10.1007/s11042-020-09684-x.

[13] M. Tezer, E. P. Yıldız, A. R. Masalimova, A. M. Fatkhutdinova, M. R. Zheltukhina, and E. R. Khairullina, “Trends of Augmented Reality Applications and Research throughout the World: Meta-Analysis of Theses, Articles and Papers between 2001-2019 Years,” International Journal of Emerging Technologies in Learning (IJET), vol. 14, pp. 154-174, 2019. [Online]. Available: http://dx.doi.org/10.3991/ijet.v14i22 .11768.

[14] Y. Zhou, "Mobile Information System for English Teaching Based on Edge Computing and VR", Mobile Information Systems, vol. 2021. Article ID 9741244, 2021. [Online]. Available: https://doi.org/10 .1155/2021/9741244.

[15] G. Molnar, S. Szuts, and K. Biro, “Use of Augmented Reality in Learning,” Acta Polytechnica Hungarica,vol. 15, pp. 209–222, 2018. [Online]. Available: http://doi.org/10.12700/aph.15.5.2018.5.12.

[16] D. Parras, D. Fernandez, T. Barbosa, M. Soler, and J. Molina, “An augmented reality tool for teaching application in the agronomy domain,” Applied Sciences-Basel, vol. 10,10(10), Article ID 3632, 2020. [Online]. Available: http://doi.org/10.3390/app10103632.

[17] X. Wei, D. Weng, Y. Liu, and Y. Wang, “Teaching based on augmented reality for a technical creative design course,” Computers & Education, vol. 81, pp. 221–234, 2015. [Online]. Available:http://doi.org/10.1016 /j.compedu.2014.10.017.

[18] S. Criollo, D. Abad, M. Martic, F. Velasquez, J. Perez, and S. Lujan, “Towards a New Learning Experience through a Mobile Application with Augmented Reality in Engineering Education,” Applied Sciences-Basel, vol. 11, Article ID 4921. [Online]. Available: https://doi.org/ 10.3390/app11114921.

[19] J. Cheng, Y. Wang, D. Tjondronegoro, and W. Song, “Construction of Interactive Teaching System for Course of Mechanical Drawing Based on Mobile Augmented Reality Technology,” International Journal of Emerging Technologies in Learning (IJET), vol. 13, pp.126-139, 2018. [Online]. Available: http://dx.doi.org/10.3991/ijet.v13i02.7847.

[20] Y. Ji, “Use of Virtual Reality Technology in Animation Course Teaching,” International Journal of Emerging Technologies in Learning (IJET), vol. 16, pp. 191-208, 2021. [Online]. Available: http://dx.doi. org/10.3991/ijet.v16i17.25337.

[21] K. Lee, “Augmented Reality in Education and Training,” TechTrends, vol. 56, pp. 13–21,2012. [Online]. Available: http://doi.org/10.1007/s1 1528-012-0559-3.

[22] C. Myung, L. Youngho, K. Sun, and J. Moon, “Design of a Markup Language for Augmented Reality Systems,” Journal of The Korea Internet of Things Society. Vol. 7, pp. 21-25, 2021.

[23] H. Zhang, G. Yang, C. Liu, J. Wang, and C. Lang, “Aided learning system of cross platform mobile terminal based on ar technology,” Journal of Graphics. Vol. 39, pp. 85-90, 2018. [Online]. Available: http://dx.doi.org/10.11996/JG.j.2095-302X.2018010085.

[24] J. Martin, P. Fabiani, W. Benesova, M. Meneses, C. Mora, “Augmented reality to promote collaborative and autonomous learning in higher education,” Computers in Human Behavior, vol. 51, pp. 752–761, 2015. [Online]. Available: http://doi.org/10.1016/j.chb.2014.11.093.

[25] X. Deng, J. Lei, M. Chen, “Application of VR in the Experimental Teaching of Animation Art,” Mobile Information Systems, vol. 2021, Article ID 4642850, 2021. [Online]. Available: https://doi.org/10. 1155/2021/4642850.

[26] D. Parras-Burgos, D. Fernández-Pacheco, T. Polhmann, M. Soler-Méndez, J. Molina-Martínez, “An Augmented Reality Tool for Teaching Application in the Agronomy Domain,” Applied Sciences, vol. 10, Article ID 3632, 2020. [Online]. Available: https://doi.org/10.3390 /app10103632.

Extended Max-Occurrence with Normalized Non-Occurrence as MONO Term Weighting Modification to Improve Text Classification

Cristopher C. Abalorio[1], Ruji P. Medina[3]
Graduate Programs
Technological Institute of the Philippines, Quezon City, Philippines

Ariel M. Sison[2]
School of Computer Studies
Emilio Aguinaldo College
Manila, Philippines

Gleen A. Dalaorao[4]
College of Computing and Information Sciences
Caraga State University
Butuan City, Philippines

***Abstract*—The increased volume of data due to advancements in the internet and relevant technology makes text classification of text documents a popular demand. Providing better representations of the feature vector by setting appropriate term weight values using supervised term weighting schemes improves classification performance in classifying text documents. A state-of-the-art term weighting scheme MONO with variants TF-MONO and SRTF-MONO improves text classification considering the values of non-occurrences. However, the MONO strategy suffers setbacks in weighting terms with non-uniformity values in its term's interclass distinguishing power. In this study, extended max-occurrence with normalized non-occurrence (EMONO) with variants TF-EMONO and SRTF-EMONO are proposed where EMO value is determined as MO interclass extensions as improvements to address its problematic weighting behavior of MONO as it neglected the utilization of the occurrence of the classes with short-distance document frequency in non-uniformity values. The proposed schemes' classification performance is compared with the MONO variants on the Reuters-21578 dataset with the KNN classifier. Chi-square-max was used to conduct experiments in different feature sizes using micro-F1 and macro-F1. The results of the experiments explicitly showed that the proposed EMONO outperforms the variants of MONO strategy in all feature sizes with an EMO parameter value of 2 sets number of classes in MO extension. However, the SRTF-EMONO showed better performance with Micro-F1 scores of 94.85% and 95.19% for smallest to largest feature size, respectively. Moreover, this study also emphasized the significance of interclass document frequency values in improving text classification aside from non-occurrence values in term weighting schemes.**

Keywords—Extended MO; normalized NO; text classification; term weighting scheme

I. Introduction

As the volume of data has dramatically increased in the digital environment due to rapid advances in the internet and developing technology, many researchers focus on the data organization, data retrieval, and data mining. One widely used technology mentioned above is text classification [1]. Text classification (TC) labels text documents with categories from predefined classes according to their content. Spam SMS and email filtering [2][3][4], author recognition [5][6], classification in clinical text and medical documents [7][8][9], sentiment analysis [10][11], and short text classification[12][13][14] are topics in research under the different subdomains category of TC in literature.

In TC, modeling in the feature vector space used Bag-Of-Words (BOW) [14], which extracts unique words representing the text document in the collection. The machine learning algorithm assigns documents to predefined classes. The VSM numerical feature vector represents each document with significant weights thru Term Weighting Scheme (TWS). TWS plays a vital role as it directly affects the classifying performance and simplifies the classifier's jobs. Three factors comprise the TWS for TC: term frequency factor (TFF), collection frequency factor (CFF), and length normalization factor (LNF). The TFF refers to the ratio of the number a term occurs [15][16][17], while the CFF [1] refers to the ratio of the information of the specific term to each document in a text collection. The LNF is used to normalize text collection containing different document lengths. TWS is categorized into two: supervised (STWS) and unsupervised (UTWS) Term Weighting Scheme [18][19]. STWS includes the class information in the weighting process, while UTWS ignores class information in setting weights to terms. Currently, STWS is preferred to use as it outperforms previous weighting strategies [20]. Its latest state-of-the-art is the SRTF-MONO, a variant of the MONO strategy [21] that added value to the non-occurrence of a term in stressing weights.

However, the MONO strategy suffers setbacks in weighting terms with non-uniformity of values in its class document frequency such as (1) a case of giving equal values to $MONO_{Local}$ leads to equal $MONO_{Global}$ weights containing different max-occurrence and non-occurrence; (2) the occurrences of interclass distinguishing power of a term proven in literature [22] [23] to give good term weighting ability leading successful classifying performance is neglected as it focused more class' document frequency for non-occurrence members.

The motivation of this study is to address the issue of MONO in its problematic weighting behavior to terms. In generating max-occurrence, the standard MONO formula neglects the utilization of the occurrence of the classes with short-distance document frequency in non-uniformity values.

These occurrences are the concentration of values representing the interclass distinguishing power essential in improving classification performance in STWS. The top classes with higher class occurrences could affect the belongingness of a document and should not be treated as non-occurrences. Thus, this study aims to enhance MONO by extending the max-occurrence group to cater to succeeding classes with higher document frequency values to set its real class distinguishing power of a term; and employing normalization to the non-occurrences to address the imbalanced distribution of the document frequency of classes.

The remaining parts of the paper are ordered as follows. The literature review that introduces several TWS and examines the state-of-the-art MONO strategy is presented in Section 2. The enhancement and required data, techniques, and evaluation are presented in Section 3. In Section 4, we present the results and discussion of the study. Finally, in Section 6, we conclude and recommend developing this study in text classification subdomains.

II. Literature Review

A. Schemes on Assigning Weights to Terms

First, Term Frequency – Inverse Document Frequency (TF-IDF), a TWS, topped the list in chronological order on research related to assigning weights to terms [24]. The IDF is one of the pioneering strategies in setting weights to terms adapted from the studies in the area of information retrieval used in text classification task proposed by Karen Spärck Jones, implied that the assigning of weights to terms should take place under the collection frequency factor (IDF) to utilize the terms effectively. This collection frequency factor is adapted to join with term frequency (TF). As an improvement in TWS for TC, Deisy et al. modified the IDF called MIDF and successfully outperformed Weighted IDF and TF-IDF [25]. The results revealed computation of MIDF is characteristically easier, and there was an improved TC performance compared with the other term weighting schemes. Sabbah et al. proposed several TWS as TF-IDF’s modification (mTFmIDF, mTF, TFmIDF, and mTFIDF) which are developed with improved results than standard TF, TF-IDF, and Entropy. They also substantiated the importance of employing ELM, NB, SVM, and KNN classifiers with popular text corpora. In another study, Debole and Sebastiani proposed several term weighting strategies, namely TF-GR, TF-IG, and TF-CHI, based on feature selection methods named Gain Ratio, Information Gain, and Chi-Square [26]. The study commenced the idea of adding the class information in weighting to terms in TC, which was later called supervised term weighting schemes (STWS). The strategy improved the classifying performance compared to the previously mentioned and traditional TWS.

A novel collection frequency factor under the STWS category, introduced by Lan et al., namely TF-RF based on relevance frequency [27]. TF-RF focused terms in its terms' class distribution considering the positive and negative ratio. This new TWS showed an improved performance than Binary, TF, and TF-IDF (unsupervised TWS) and TF-IG, TF-CHI2, and TF-LogOR (supervised TWS). Another term weighting method derived from TF-RF, namely LogTFRFmax strategy proposed by Xuan and Le Quang for TC [28]. They showed that TF-RF's classification accuracy could be increased by combining reduced TF values with RF. Liu et al. introduced a TWS intended for unbalanced text datasets, namely TF-PB [29]. TF-PB is derived using the term's interclass and intraclass distribution. They showed that classification accuracy could be increased by utilizing information on a term's inner-class distribution from unbalanced text datasets. Log-TFTRR introduced by Ko is derived from class distribution by using the negative and positive class probabilities of terms [30]. Log TF-TRR showed better performance than (TF-CHI, TF-RF) over the use collection of text data such as Korean UseNet, Reuters-21578, and 20-Newsgroups.

A study called Positive Impact Factor (PIF) introduced by Emmanuel et al. showed better TC performance concerning computing time and classification accuracy upon experiments on Classic3 text collection [31]. Altınçay and Erenel investigated previously proposed and most-used TWS for TC [32]. They discovered that the ratios and term occurrence probabilities are the reasons for relative performance differences in giving weights to terms. A new collection frequency factor introduced by Altınçay and Erenel in another study derived from the logarithm of term frequencies [33]. They showed that the lesser term frequency values bore better classification performance combined with their proposed term weighting scheme. They also indicated that the distribution of TF in the collection of the text suggests the usable form of the TF factor.

Another related study, a proposed method by Cai et al. in tagging systems wherein it modeled the resource along with the user's profile, indicating the utilization of normalized form of term frequency [34]. Badawi and Altınçay introduced another study implemented for binary text classification using a termset weighting strategy [35]. They wanted to prove using the bag-of-words approach as an effective method. They showed that their process produces document vectors successful for TC. Later, they introduced new cardinality statistic-based TWS [36]. Two collection frequency factors integrated with Bag-of-Words (BOW) were used in this new term weighting strategy. They emphasized the success of text classification performances in setting weights to terms generated by the standard BOW approach can be increased with n-term setting its values to n = 2, 3, 4.

In another study, a term weighting scheme adapted from information of the term’s document was introduced by Ren and Sohrab. [37]. They showed that TF-IDF-ICSDF outperforms five previous TWS (TF-CC, TF-IDF, TF-PB, TF-OR, and TF-RF) on 20 Newsgroups, RCV1-v2, and Reuters-21578 with Centroid, SVM, and NB classifiers. Escalante et al. used genetic programming to improve TWS [38]. They stated that genetic programming aims to acquire which combination of units makes greater discriminative TWS. They showed that genetic programming generated superior classification results than latest and popular strategy on assigning weight to terms. A new collection frequency factor for TWS introduced by Chen et al. named Inverse Gravity Moment (IGM) adapted from a statistical model [22]. IGM has two variants, namely TF-IGM and SRTF-IGM. They stated that the adequate distinguishing power is taken from the information of the interclass distribution of the documents. They showed that IGM performed better classification than the 5 TWS (RF, CHI2, Prob, ICSDF, IDF) using KNN and SVM classification algorithm on the popular text corpora (20-Newsgroups, TanCorp, and Reuters-21578). Dogan et al. improved the IGM by reorganizing the IGM formula to address similar values given to different term weights [23]. The improvement in the IGM provides better classification results than the standard IGM. Sabbah et al. implemented a hybridized term weighting strategy for terrorism activity detection in texts [39] by creating a set of features from the combination of small feature subsets taken from TF, IDF, TF-IDF, Entropy, and Glasgow TWS. They also indicated that the successful text classification could be improved by utilizing small sets of features representing the most significant terms inside the text.

With all the previous and successfully introduced supervised TWS, the concept is always the involvement of the available class information. In contrast, another collection frequency factor STWS presented by Dogan et al., called max-occurrence and non-occurrence (MONO) [21], added the value of non-occurrence or the absence of a term in the distribution of terms in the documents on each class upon stressing weights. MONO comes with two variants, TF and SRTF (the squared root of TF). SRTF-MONO variant outperforms the classification performance of the previous STW using the news dataset.

B. Assigning Weight to Terms using MONO Strategy

The standard MONO term weighting scheme [21] according to Dogan et'al:

1) Assume that the document frequency df of a term t_i or df_{t_i} from the j classes of a text collection is shown in (1). The df_{t_i} collection is sorted in descending order. The head of the left has the highest value, and the tail in the right has the lowest as shown in (2).

2) The sorted df_{t_i} is divided into two groups; one has the highest class df_{t_i} values and the other for the rest of the classes. Groups are categorized into two the max-occurrence (MO) ratio group and the non-occurrence (NO) ratio group as shown in (3).

3) MO_{t_i} value, corresponds the ratio between total document frequencies and the total number of documents available on the class with maximum occurring t_i. After calculating the MO_{t_i} value, the NO_{t_i} value is calculated as shown in (4). NO_{t_i} value is calculated on classes excluded from MO_{t_i} calculation. The value is computed from the ratio between the quantity of document frequencies and the total number of the documents of the class within the NO ratio group.

4) In order to obtain the $MONO_{Local}$ weights of a term the MO and NO is multiplied as shown in (5).

5) Finally, $MONO_{Global}$ weight of term t_i is calculated as shown in (6). In the aforementioned equation in (6), α parameter is presented. The purpose of α is to set balance to weights to global values in the weighting stage where values ranges from 5.0 to 9.0 and 7.0 as its default.

6) The two TWS based upon $MONO_{Global}$ collection frequency factor are shown in (7). The $SRTF(t_i, d_k)$ is squared root of TF values of the term t_i in text document d_k.

$$df_{t1} = \left\{d_{i1}, d_{i2}, d_{i3}, d_{i2}, \dots, d_{ij-1}, d_{ij}\right\} \tag{1}$$

$$sorted\_df_{t1} = \left\{d_{i3}, d_{i1}, d_{i4}, \dots, d_{ij}, d_{ij-1}\right\} \tag{2}$$

$$sorted\_df_{t_i} = \left\{ \underbrace{\overbrace{d_{i3}}^{C_{t_i\_max}}}_{eq} \; |, \underbrace{\overbrace{\overline{d_{i1}}, \overline{d_{i4}}, \dots, \overline{d_{ij}}, \overline{d_{ij-}}}^{C_{others}}}_{eq} \right. \qquad sorted\_df_{t_i} = \left\{ \underbrace{\overbrace{d_{i3}}^{MO}}_{eq} \; |, \underbrace{\overbrace{\overline{d_{i1}}, \overline{d_{i4}}, \dots, \overline{d_{ij}}, \overline{d_{ij-}}}^{NO}}_{eq} \right. \tag{3}$$

$$MO_{t_i} = \frac{D_{t_{i1}\_max}}{D_{total(t_i\_max)}} \qquad NO_{t_i} = \frac{D_{\overline{t_i}}}{D_{total(\overline{t_i})}} \tag{4}$$

$$MONO_{Local}(t_i) = \underbrace{\overbrace{\left[\frac{D_{t_{i1}\_max}}{D_{total(t_i\_max)}}\right]}^{MO_{t_i}}}_{eq} * \underbrace{\overbrace{\left[\frac{D_{\overline{t_i}}}{D_{total(\overline{t_i})}}\right]}^{NO_{t_i}}}_{eq} \tag{5}$$

$$MONO_{Global}(t_i) = [1 + \alpha * MONO_{Local}(t_i)] \tag{6}$$

$$TF - MONO = TF(t_i, d_k) * [MONO_{Global}(t_i)]$$
$$SRTF - MONO = SRTF(t_i, d_k) * [MONO_{Global}(t_i)]$$
(7)

C. Empirical Observations of Issues in MONO Local and Global Weights

This subsection illustrates empirical observations on weighting terms using standard MONO strategy.

Assume that there exist three terms (t_4, t_5, t_6) where the frequencies of its documents in four classes are {100, 60, 0, 0}, {200, 40, 10, 10}, and {100, 55, 5, 0} respectively. Then, assume that the entire documents of each class are uneven, which are 200, 300, 400, and 500, respectively. The standard $MONO_{Local}$ and $MONO_{Global}$ produced similar values in this scenario.

Supervised TWS considers intraclass and interclass distinguishing essential factors in specifying weights to terms [40]. Intraclass within a specific class, while interclass in multiple classes. MO represents intra-class and NO as the interclass. However, the existing interclass is not utilized because NO neglects actual interclass existence values and obtains non-existence. Class information improves classification [27], such as TF-CHI [26] [41] considers the term's intraclass distribution. Moreover, the used of interclass [22][23] outperforms the previous TWS in classification performance. Terms (t_4-t_6) using MONO ignore interclass occurrence as it selects one member of a MO group and the rest to NO group even to other classes containing higher class-document frequency. It failed to fully represent the distinguishing power of a term as it assigns equal scores, as shown in Table I in its weights.

TABLE I. RESULTS OF THE CASE SCENARIO

Terms	Document Frequencies	No. of documents in every class	MO	NO	MONO Local	MONO Global
t_4	(100, 60, 0, 0)	(200, 300, 400, 500)	0.5	0.95	0.475	4.325
t_5	(100, 40, 10, 10)	(200, 300, 400, 500)	0.5	0.95	0.475	4.325
t_6	(100, 55, 5, 0)	(200, 300, 400, 500)	0.5	0.95	0.475	4.325

III. METHODOLOGY

The enhancement of MONO strategy is categorized into two major processes. The extended max-occurrence and non-occurrence normalization is shown in Fig. 1.

MO computation selects only a single class with the highest document frequency values, previously shown in (3). In the modification, extended max-occurrence (EMO) is proposed to cater to interclass occurrence values essential for supervised term weighting schemes. EMO refers to the number of class members of the MO group comprising the ratio of the number of documents in a class where the term mostly occurs and its total number of documents in that class. In the j classes in (1), assume that $EMO = 2$ where $1 < EMO < d_{ij-1}$ then MO group covers d_{i3} and d_{i1} representing ratios as shown in (8), EMO_{t_i} calculates the new value of the weight of a term as shown in (9). In non-occurrence normalization, the original NO formula in (4) is modified. As the EMO is extended, NO members are reduced. With $EMO = 2$ then each NO member is individually calculated in percentage and normalized as shown in (10). The rest of the formulas for assigning weight to a term using the EMONO is shown in (11) for local weights and (12) for global weights on different variants.

$$sorted\_df_{t_i} = \left\{ \overbrace{d_{i3}, d_{i1}}^{MO} |, \overbrace{\overline{d_{\iota4}}, \overline{d_{\iota5}}, \dots, \overline{d_{\iota J}}, \overline{d_{\iota J-1}}}^{NO} \right\} \quad (8)$$

$$EMO_{t_i} = \frac{MO_{t_{i1}} + MO_{t_{i2}}}{D_{total}\left(MO_{t_{i1}} + MO_{t_{i2}}\right)} \quad (9)$$

$$NO_{t_{i1}} = \frac{NO_{\overline{t_{\iota1}}}}{D_{total}(\overline{t_{\iota1}})} * 100 \qquad NO_{t_{i2}} = \frac{NO_{\overline{t_{\iota2}}}}{D_{total}(\overline{t_{\iota2}})} * 100 \qquad Normalized\_NO_{t_i} = \frac{NO_{t_{i1}} + NO_{t_{i2}}}{200} \quad (10)$$

$$EMONO_{Local}(t_i) = EMO_{t_i} * \ Normalized\_NO_{t_i} \qquad EMONO_{Global}(t_i) = [1 + \alpha * EMONO_{Local}(t_i)] \quad (11)$$

$$TF - EMONO = TF(t_i, d_k) * \ [EMONO_{Global}(t_i)] \qquad SRTF - EMONO = SRTF(t_i, d_k) * \ [EMONO_{Global}(t_i)] \quad (12)$$

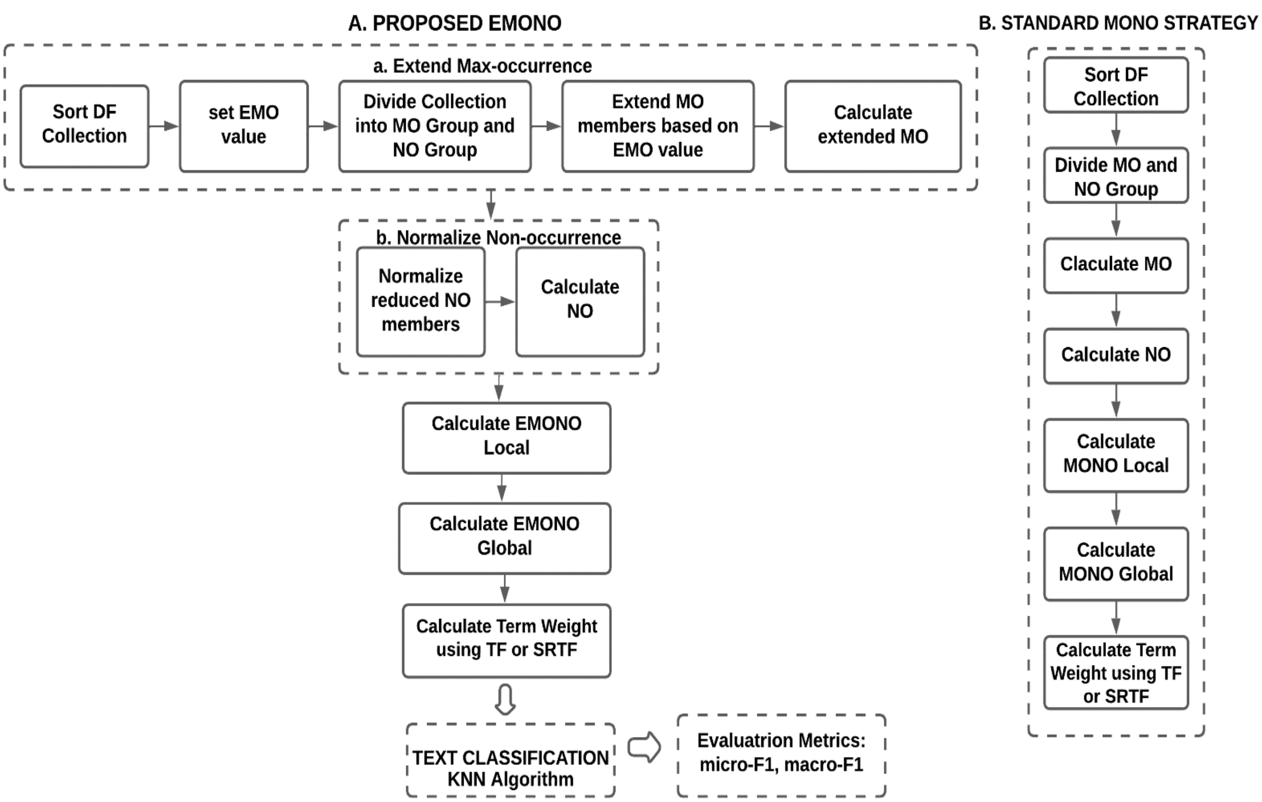


Fig. 1. Framework of the Study.

A. Dataset

Reuters-21578 is the benchmark dataset utilized in this study. It is one of the most popular unbalanced datasets preferred for text classification. Training and testing of data are taken from the Reuters-ModApte split, which is its segmentation for the experiments' training and testing. Due to removing the multi-labeled from the reading text in this dataset, the texts 'wheat' and 'corn' classes are emptied. There are a total of 7,215 documents with eight categories: earn (3735), acq (2124), money-fx (354), grain (45), crude (259), trade (332), interest (211), and ship (155).

B. Pre-Processing

In preparing the dataset, the following are the preprocessing methods [21] applied to the documents: conversion to lowercases, removing stop-words, alphabetic tokenization, Porter Stemming [42], and eliminating of seldom occurring terms (retain words occur more than one time).

C. Feature Selection

A feature selection method is preferred to manifest the proposed improvements' performance in a text classification dealing with high dimensionality. A standard statistical metric named chi-square max is used to obtain a selected number of features for this experiment. The total number for observations is 7,215 and the highest feature extracted is 9,237. The term weighting schemes are tested with top terms sorted in descending order and scored using the employed feature selection method {1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, and 9237} terms scored and sorted.

D. Learning Algorithm

The well-known and most used classification algorithm K-Nearest Neighbor (KNN) is employed in this study to classify documents from the benchmark Reuters-21578 dataset. Aside from its simplicity, it is generally utilized for performing text classification in literature [43][44].

E. Performance Evaluation

The text classification performance of this study was evaluated using micro-F1 and macro-F1 scores. Precision and Recall in (13) used to derived F1 scores preferred to use for uneven distribution of documents in the classes of multiclass text classification as shown in (14).

$$Precision_{c_k} = \frac{TP_{c_k}}{TP_{c_k}+FP_{c_k}} \quad Recall_{c_k} = \frac{TP_{c_k}}{TP_{c_k}+FN_{c_k}}$$

$$F1_{c_k} = \frac{2*Precision_{c_k}*Recall_{c_k}}{Precision_{c_k}+Recall_{c_k}} \tag{13}$$

$$Micro-F1 = \frac{2*\sum_k^c TP_{c_k}}{2*\sum_k^c TP_{c_k}+\sum_k^c FP_{c_k}+\sum_k^c FN_{c_k}} \qquad Macro-F1 = \frac{1}{c}\sum_{k=1}^{c} F1_{c_k} \tag{14}$$

IV. Result and Discussion

The results of the performance of the MONO and the proposed modification were analyzed in this section, along with the use of the classification algorithm KNN on Reuters-21578 corpus with all alpha values set to 6.0. The boldface values correspond to the highest F1 values in each micro and macro score. Table II shows the new local, and global weights generated using the proposed EMONO on case scenario with problematic weights mentioned in Table II.

Table III shows the classification performance of the proposed EMONO combined with term frequency on 1,000 and 9237 feature sizes. As EMO value implies the number of classes in MO extension, in 1000 features, EMO values 2, 3, and 4 have 0.927815207 as the highest scores for micro-F1, and only in the EMO value of 2 has 0.880240879 as the highest value for macro-F1. The proposed EMONO combined with the squared root of TF's EMO value of 2 has the highest obtained scores for micro-F1 and macro F1 with 0.948508181 and 0.893171284, respectively. In TF-EMONO 9237 features, the EMO value of 3 got the maximum micro-F1 score of 0.918671800 as the highest score, and the EMO value of 2 has 0.860379651 as the highest value for macro-F1. On the other hand, the classification performance of the proposed EMONO combined with the squared root of term frequency EMO value of 4 has the highest obtained scores of 0.957170356 and 0.909860837 micro-F1 to macro-F1, respectively.

The comparative results of the classification performance of the original MONO and proposed EMONO combined with TF and squared root of TF with 2 as the successful EMO value is shown in Table IV. The table shows that EMONO combined with both TF and SRTF have greater values than the original MONO strategy in micro-F1 and macro-F1. It is explicit that the proposed EMONO generally outperformed the original MONO strategy in all indicated feature sizes shown in the table.

Fig. 2 shows the classification performance of micro-F1 and macro-F1 in plot graph. The overall performance of the proposed EMONO is superior when combined with TF and squared root of TF explicitly shown in the plot graph.

TABLE II. EMONO WEIGHTS ON ISSUES IN LOCAL AND GLOBAL WEIGHTS

Terms	Document Frequencies	No. of documents in every class	Extended MO	Normalized NO	EMONO Local	EMONO Global
t_4	(100, 60, 0, 0)	(200, 300, 400, 500)	0.32	1	0.32	3.24
t_5	(100, 40, 10, 10)	(200, 300, 400, 500)	0.28	0.9775	0.2737	2.9159
t_6	(100, 55, 5, 0)	(200, 300, 400, 500)	0.31	0.99375	0.3080625	3.1564375

TABLE III. MODIFIED MONO ON MINIMUM AND MAXIMUM FEATURE SIZES

EMO	TF-EMONO 1,000 Features		SRTF-EMONO 1,000 Features		TF-EMONO 9,237 Features		SRTF-EMONO 9,237 Features	
	micro-F1	macro-F1	micro-F1	macro-F1	micro-F1	macro-F1	micro-F1	macro-F1
2	**0.927815207**	**0.880240879**	**0.948508181**	**0.893171284**	0.910490857	**0.860379651**	0.951876805	0.898198983
3	**0.927815207**	0.861238997	0.941289702	0.882265425	**0.918671800**	0.864875929	0.954764196	0.907354723
4	**0.927815207**	0.861238997	0.942252166	0.888250757	0.913859480	0.856517844	**0.957170356**	**0.909860837**
5	0.918190568	0.845061943	0.936477382	0.873495059	0.913859480	0.850101527	0.948989413	0.897765549
6	0.920596728	0.845835989	0.935514918	0.869062724	0.911934552	0.851904410	0.948026949	0.894754037
7	0.914821944	0.824183935	0.934552454	0.848188784	0.900384986	0.819861594	0.950914341	0.889930181

TABLE IV. COMPARATIVE RESULTS OF MONO AND MODIFIED

No. of Features	micro-F1				macro-F1			
	TF-MONO	TF-EMONO	SRTF-MONO	SRTF-EMONO	TF-MONO	TF-EMONO	SRTF-MONO	SRTF-EMONO
1000	0.9095	0.9278	0.9365	0.9485	0.7952	0.8802	0.8749	0.8932
2000	0.9100	0.9211	0.9355	0.9398	0.7964	0.8650	0.8594	0.8696
3000	0.9076	0.9196	0.9346	0.9413	0.8038	0.8794	0.8702	0.8727
4000	0.9038	0.9153	0.9418	0.9442	0.7971	0.8703	0.8808	0.8856
5000	0.9033	0.9110	0.9403	0.9485	0.8031	0.8627	0.8808	0.8932
6000	0.8985	0.9086	0.9398	0.9514	0.8028	0.8624	0.8816	0.8985
7000	0.9023	0.9090	0.9437	0.9514	0.8104	0.8615	0.8891	0.9002
8000	0.9038	0.9100	0.9480	0.9519	0.8109	0.8619	0.8942	0.8966
9000	0.9076	0.9095	0.9471	0.9524	0.8121	0.8600	0.8934	0.8984
9237	0.9081	0.9105	0.9480	0.9519	0.8123	0.8604	0.8954	0.8982

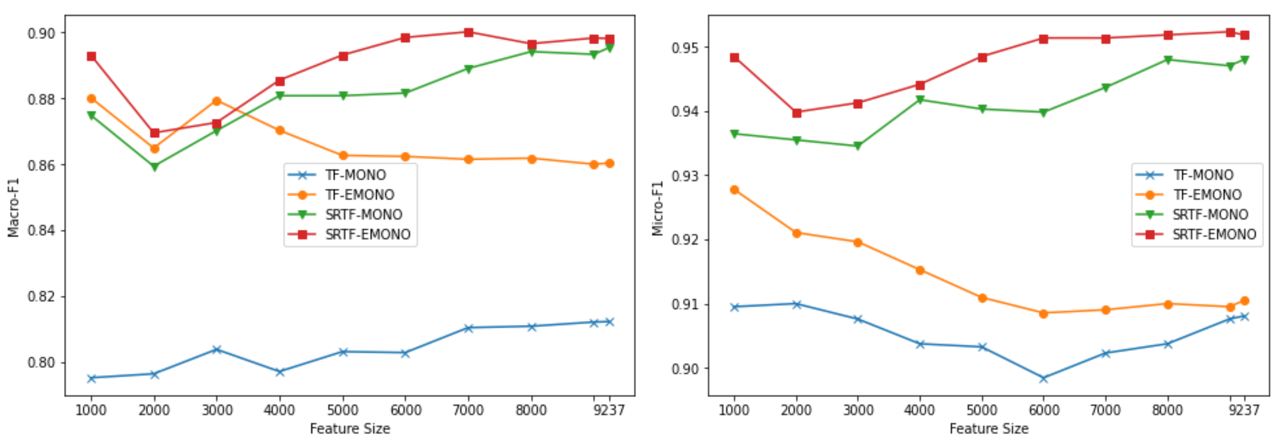


Fig. 2. Micro-F1 Acquired from 4 TWS on Reuters-21578 Dataset using KNN (k=5) with EMO=2.

V. CONCLUSION AND RECOMMENDATION

In this study, behavior on how the original MONO strategy assigned a weight to terms was comprehensively analyzed. Two new variants of the EMONO, namely TF-EMONO and SRTF-EMONO, are proposed. The MONO modifications employed in the study are the EMO parameter's utilization and the normalization of the non-occurrences. EMO parameter sets classes to cover in generating max-occurrence extension, and the normalization is utilized for class imbalance. The results of the experiments explicitly showed that the proposed improvement outperforms the variants of the original MONO strategy in all feature sizes with an EMO parameter value of 2 sets number of classes in MO extension. Even though there are feature sizes in which EMONO has a slight increase of micro-F1 and macro-F1, the increase is consistent and gradual in all features with a selected EMO value. However, the SRTF-EMONO showed better performance with Micro-F1 scores of 94.85% and 95.19% for smallest to largest feature size. It can be stated that the proposed EMONO term weighting scheme is superior in classification performance to the original MONO strategy. It is recommended that the proposed EMONO be implemented in text classification. It is also suggested to utilize it in other text classification subdomains.

REFERENCES

[1] M. Marcin Michał and J. Protasiewicz, "A Recent Overview of the State-of-the-Art Elements of Text Classification," 2018, doi: 10.1016/j.eswa.2018.03.058.

[2] O. Abayomi-alli, S. Misra, A. Abayomi-alli, and M. Odusami, "A review of soft techniques for SMS spam classification: Methods, approaches and applications," Eng. Appl. Artif. Intell., vol. 86, no. August, pp. 197–212, 2019, doi: 10.1016/j.engappai.2019.08.024.

[3] R. Pav, D. Ruano-ord, G. Silvana, and R. M, "Enhancing representation in the context of multiple-channel spam filtering," vol. 59, no. May 2021, 2022, doi: 10.1016/j.ipm.2021.102812.

[4] A. K. Uysal and S. Gunal, "A novel probabilistic feature selection method for text classification," Knowledge-Based Syst., vol. 36, pp. 226–235, 2012, doi: 10.1016/j.knosys.2012.06.005.

[5] O. Fourkioti, S. Symeonidis, and A. Arampatzis, "Language models and fusion for authorship attribution," Inf. Process. Manag., vol. 56, no. 6, p. 102061, 2019, doi: 10.1016/j.ipm.2019.102061.

[6] C. Akimushkin and D. R. Amancio, "On the role of words in the network structure of texts: application to authorship attribution," Phys. A Stat. Mech. its Appl., pp. 49–58, 2018.

[7] X. Li, M. Cui, J. Li, R. Bai, Z. Lu, and U. Aickelin, "A hybrid medical text classification framework: Integrating attentive rule construction and neural network," Neurocomputing, vol. 443, pp. 345–355, 2021, doi: 10.1016/j.neucom.2021.02.069.

[8] R. Matsuo and T. B. Ho, "Semantic Term Weighting for Clinical Texts," pp. 543–551, 2018.

[9] G. Mujtaba et al., "Clinical Text Classification Research Trends: Systematic Literature Review and Open Issues," Expert Syst. Appl., 2018, doi: 10.1016/j.eswa.2018.09.034.

[10] A. Mee, E. Homapour, F. Chiclana, and O. Engel, "Sentiment analysis using TF – IDF weighting of UK MPs ' tweets on Brexit," Knowledge-Based Syst., vol. 228, p. 107238, 2021, doi: 10.1016/j.knosys.2021.107238.

[11] G. N. H, R. Siautama, A. C. I. A, and D. Suhartono, "Extractive Hotel Review Summarization based on TF / IDF and Adjective-Noun Pairing by Considering Annual Sentiment Trends," Procedia Comput. Sci., vol. 179, no. 2020, pp. 558–565, 2021, doi: 10.1016/j.procs.2021.01.040.

[12] I. Alsmadi, "Term weighting scheme for short-text classification: Twitter corpuses," Neural Comput. Appl., vol. 8, 2018, doi: 10.1007/s00521-017-3298-8.

[13] S. S. Samant, N. L. B. Murthy, and A. Malapati, "Improving Term Weighting Schemes for Short Text Classification in Vector Space Model," IEEE Access, vol. 7, pp. 166578–166592, 2019, doi: 10.1109/ACCESS.2019.2953918.

[14] V. N. Gudivada, D. L. Rao, and A. R. Gudivada, Information Retrieval : Concepts , Models , and Systems, 1st ed. Elsevier B.V., 2018.

[15] T. Dogan and A. K. Uysal, "On Term Frequency Factor in Supervised Term Weighting Schemes for Text Classification," Arab. J. Sci. Eng., 2019, doi: 10.1007/s13369-019-03920-9.

[16] L. Chen, L. Jiang, and C. Li, "Using modified term frequency to improve term weighting for text classification," Eng. Appl. Artif. Intell., vol. 101, no. November 2020, p. 104215, 2021, doi: 10.1016/j.engappai.2021.104215.

[17] Y. He, T. Li, Y. Huang, and S. Li, "Term Weight Algorithm Oriented Terms: Low Frequency Rather Than Little Occurrences," Procedia Comput. Sci., vol. 176, pp. 838–847, 2020, doi: 10.1016/j.procs.2020.09.079.

[18] F. Sebastiani, "Machine Learning in Automated Text Categorization," vol. 34, no. 1, pp. 1–47, 2002.

[19] H. Wu, X. Gu, and Y. Gu, "Balancing between over-weighting and under-weighting in supervised term weighting," Inf. Process. Manag., vol. 0, pp. 1–11, 2016, doi: 10.1016/j.ipm.2016.10.003.

[20] A. Alsaeedi, "A survey of term weighting schemes for text classification Abdullah Alsaeedi," vol. 12, no. 2, pp. 237–254, 2020.

[21] T. Dogan and A. K. Uysal, "A novel term weighting scheme for text classification: TF-MONO," J. Informetr., vol. 14, no. 4, p. 101076, 2020, doi: 10.1016/j.knosys.2012.06.005.

[22] K. Chen, Z. Zhang, J. Long, and H. Zhang, "Turning from TF-IDF to TF-IGM for term weighting in text classification," Expert Syst. Appl., vol. 66, pp. 245–260, 2016, doi: 10.1016/j.eswa.2016.09.009.

[23] T. Dogan and A. K. Uysal, "Improved inverse gravity moment term weighting for text classification," Expert Syst. Appl., vol. 130, pp. 45–59, 2019, doi: 10.1016/j.eswa.2019.04.015.

[24] K. S. Jones, "A statistical interpretation of term specificity and its application in retrieval," vol. 60, no. 5, 2004.

[25] C. Deisy, M. Gowri, S. Baskar, and N. Ramraj, "A NOVEL TERM WEIGHTING SCHEME MIDF FOR TEXT CATEGORIZATION," vol. 5, no. 1, pp. 94–107, 2010.

[26] F. Debole and F. Sebastiani, "Supervised Term Weighting for Automated Text Categorization *," pp. 784–788, 2004, doi: https://doi.org/10.1007/978-3-540-45219-5_7.

[27] M. Lan, C. L. Tan, S. Member, J. Su, and Y. Lu, "Supervised and Traditional Term Weighting Methods for Automatic Text Categorization," vol. 31, no. 4, pp. 721–735, 2009.

[28] N. P. Xuan and H. Le Quang, "A New Improved Term Weighting Scheme for Text Categorization," vol. 1, pp. 261–270, 2014, doi: 10.1007/978-3-319-02741-8.

[29] Y. Liu, H. Tong, and A. Sun, "Imbalanced text classification: A term weighting approach," Expert Syst. Appl., vol. 36, no. 1, pp. 690–701, 2009, doi: 10.1016/j.eswa.2007.10.042.

[30] Y. Ko, "A New Term-Weighting Scheme for Text Classification Using the Odds of Positive and Negative Class Probabilities," vol. 66, no. 12, pp. 2553–2565, 2015, doi: 10.1002/asi.

[31] M. Emmanuel, S. M. Khatri, and D. R. R. Babu, "A Novel scheme for Term weighting in Text Categorization: Positive Impact factor," IEEE Int. Conf. Syst. Man, Cybern., 2013, doi: 10.1109/SMC.2013.392.

[32] H. Altınçay and Z. Erenel, "Analytical evaluation of term weighting schemes for text categorization," Pattern Recognit. Lett., vol. 31, no. 11, pp. 1310–1323, 2010, doi: 10.1016/j.patrec.2010.03.012.

[33] Z. Erenel and H. Altıncay, "Nonlinear transformation of term frequencies for term weighting in text categorization," vol. 25, pp. 1505–1514, 2012, doi: 10.1016/j.engappai.2012.06.013.

[34] Y. Cai, Q. Li, H. Xie, and H. Min, "Exploring Personalized Searches using Tag-based User Profiles and Resource Profiles in Folksonomy," Neural Networks, 2014, doi: 10.1016/j.neunet.2014.05.017.

[35] D. Badawi and H. Altincay, "A novel framework for termset selection and weighting in binary text classification," Eng. Appl. Artif. Intell., vol. 35, pp. 38–53, 2014, doi: 10.1016/j.engappai.2014.06.012.

[36] D. Badawi and H. Altıncay, "Termset weighting by adapting term weighting schemes to utilize cardinality statistics for binary text categorization," 2017, doi: 10.1007/s10489-017-0911-6.

[37] F. Ren and M. G. Sohrab, "Class-indexing-based term weighting for automatic text classification," Inf. Sci. (Ny)., vol. 236, pp. 109–125, 2013, doi: 10.1016/j.ins.2013.02.029.

[38] H. J. Escalante et al., "Term-weighting learning via genetic programming for text classification," KNOWLEDGE-BASED Syst., 2015, doi: 10.1016/j.knosys.2015.03.025.

[39] T. Sabbah, A. Selamat, R. Ibrahim, and H. Fujita, "Hybridized term-weighting method for Dark Web classification," Neurocomputing, 2015, doi: 10.1016/j.neucom.2015.09.063.

[40] U. I. Akpana and A. Starkey, "Review of classification algorithms with changing inter-class distances," vol. 4, no. November 2020, 2021, doi: 10.1016/j.mlwa.2021.100031.

[41] H. Peng, Y. Ma, Y. Li, and E. Cambria, "Learning multi-grained aspect target sequence for Chinese sentiment analysis," Knowledge-Based Syst., vol. 148, pp. 167–176, 2018, doi: 10.1016/j.knosys.2018.02.034.

[42] M. .Porter, "An algorithm for suffix stripping," vol. 40, pp. 211–218, 2006, doi: 10.1108/00330330610681286.

[43] T. Sabbah et al., "Modified Frequency-Based Term Weighting Schemes for Text Classification," Appl. Soft Comput. J., 2017, doi: 10.1016/j.asoc.2017.04.069.

[44] V. B. S. Prasath, H. Arafat, A. Alfeilat, A. B. A. Hassanat, O. Lasassmeh, and S. Ahmad, "Effects of Distance Measure Choice on KNN Classifier Performance - A Review," pp. 1–39, 2019.

Assessing Digital Readiness of Small Medium Enterprises: Intelligent Dashboard Decision Support System

Okfalisa[1]
Department of Informatics Engineering
Universitas Islam Negeri Sultan Syarif Kasim Riau
28293 Pekanbaru, Riau, Indonesia

Mahyarni[2]
Department of Management
Universitas Islam Negeri Sultan Syarif Kasim Riau
28293 Pekanbaru, Riau, Indonesia

Wresni Anggraini[3]
Department of Industrial Engineering
Universitas Islam Negeri Sultan Syarif Kasim Riau
28293 Pekanbaru, Riau, Indonesia

Saktioto[4]
Department of Physics
Universitas Riau
Pekanbaru, Indonesia

B. Pranggono[5]
Department of Engineering and Mathematics
Sheffield Hallam University
United Kingdom

***Abstract*—The implication of the Covid-19 global pandemic is driving the transition of SMEs' business towards digitalization. However, despite the use of the digital platform, many SMEs are unable to survive. Therefore, this study included a focus on Decision Support System (DSS)-based dashboard model as a new feature in assessing SMEs' digitalization readiness. The twenty-four criteria appraisals are regarded in this sense as two views of business and Information Technology (IT) dimensions which include the Fuzzy-Analytical Hierarchy Process Method (F-AHP) for the weighting measurement and Objective Matrix (OMAX) for the performance mapping analysis, and both are embedded in the Business Intelligent (BI) dashboard development. In Riau Province, Indonesia, a total of 118 SMEs were interested in this study and fact thus revealed the general performance of SMEs as rated at an "Average" level of index value 4.95 with comprehensive parameters for index contribution viz., 3.79, 3.84, 7.75, 4.68, 4.32, and 5.43 for Business Activity (BA), Transaction (TC), Marketing (MC), Management (MG), Micro Environment (MI) and Macro Environment (MA) respectively. Furthermore, the dashboard prepares a tracking and analysis system with the graphical diagram extracted from each criteria hierarchy's root cause to sub-criteria. The DSS dashboard's information and knowledge have been developed into a promotional framework for stakeholders relevant to a digital business's success and sustainability performance initiatives.**

Keywords*—Decision support system; digital readiness; fuzzy analytical hierarchy process; business intelligent dashboard; objective matrix

I. Introduction

Indonesia and other developing countries provide a historical and significant scale investment of SMEs that influenced the country's economy, both in numbers of companies, the gross value-added, and the absorption of employees. The transformation of the manufacturing industry in the digital age has led to the development of SMEs that are leading the way in integrating creative technology utilization and digital platforms for product marketing, business transactions, as well as a wide range of quality investment services from top to bottom business [1]. Furthermore, the leverage of the digital business model of performance in SMEs has been seen to have the effect of enhancing overall cross-industry equipment, lowering inventory, workers' productivity, reducing logistical costs, boosting productivity, reducing production times, and rising sales, and cutting costs [2]. The Ministry of Co-operatives, Small and Medium-sized Enterprises, and the Ministry of Communication, Information Technology, and e-commerce actors launched a program entitled Eight million SMEs Go Online, aiming to foster this digital transformation by creating social and community awareness [3]. Here, several marketplaces and e-commerce sites in Indonesia, such as Lazada, Shopee, MatahariMall.com, BliBli.com, Bukalapak, Tokopedia, and Blanja.com, the founders of successful online companies, are comprised in this program. However, due to the lack of engagement and interest from SMEs and the suffering for the very principle of information exchange by active leaders, the program did not live up to expectations, and therefore, the digital revolution struggles to work.

Previous studies indicated that SMEs' common obstacles and demand in the revolution stage include human resources and capitalization, competition, marketing strategy, innovation, government actions, preparedness for the state, market demand potential, innovation and creativity, exports, companies' resistance, and survival, knowledge of local business, women empowerment, internet and digitalization, currency conversion, manufacturing, trade, and infrastructure [4]. Despite the negative impacts on SMEs' growth globally, especially in Indonesia, it still offers attractive prospects. Furthermore, the Covid-19 virus outbreak has devastated countries' economies and has forced the digital revolution to accelerate fully. The number of SMEs influenced by this has attempted to make flexible and digitalize their market growth model to survive. However, digitalization's success requires full reinforcement from external and internal parties due to the high risk of investment. Peillon and Dubruc (2019) have identified various stumbling blocks to SMEs' digitalization, including technological hurdles to the accessible entrance of technology and artificial intelligence in promoting the quality and delivery of their business practices and services [5]. Besides this, technological capital and infrastructure adoption are difficult to enforce due to a lack of fund.

Furthermore, it affects essential changes in operational areas such as core business practices, commodities, procedures, organizational structures, managerial concepts, emerging capabilities and skills in management innovation, cultural shock, and analysis of company performance [5] and [6]. More importantly, customers revisiting and stored information becomes a huge burden to reach due to their protection and privacy breaches [7]. In a nutshell, this failure detection encourages further analysis of SMEs' readiness towards the digital business's success and sustenance.

Therefore, this study aims at analyzing SMEs' readiness in the vicinity of digitalization by incorporating ideas and perspective from governments, investors, the marketplace, and SMEs industries through the opportunities for creating a model-driven DSS dashboard. The application of F-AHP and OMAX enriched the DSS dashboard analysis as an information tracking system to assess the degree and parameters of the performance index. Furthermore, this intelligent business tool struggles with the new situation in Indonesia with current challenges and digital transformation issues in the future. The increasing sense of complexity and confusion in decision-making encourages the need for a forecasting process to assist decision-making, planning, analysis, and evaluation. Since model-driven DSS comprises mathematical simulation, the parameters are manipulated and optimized by investigating product analysis's sensitivity to assist in decision making [8]. This dashboard offers a wide range of digital readiness assessment options, thus providing a service as a new performance measurement tool.

II. Literature Review

Understanding this performance measurement dashboard provides the contribution into the ultimate digital solution as well as market plan, business model, and technology penetration. Many researchers viewed and developed e-readiness tools and approaches from different viewpoints, environments, experiences, people, and objectives [9], and unfortunately, the gap between the concept and performance of e-readiness is inevitable [10]. For example, Beacon [11] and Verdict [12] are unable to guide the organization into recovering and identifying the priority concerns for success, and currently, the General Practitioner Information System (GPIS) and New IT/IS Capability Evaluation (NICE) mechanism are missing out on the involvement of stakeholders rather than recruiting technology consultants for a thorough market review [13]. The Building Information Modeling (BIM) Maturity Matrix is a restrictive tool applicable only to its organizational context and hard to customize. Furthermore, Technology Readiness Levels as company maturity index tool deteriorates in meeting the demands of SMEs sector, and simultaneously, Lou et al. (2019) released an e-readiness construction (ERiC) framework as a promising sustainable instrument for assessing the extent of organizational precaution to shield and bring about a competitive digital economy. Like the previous one, the ERiC framework is explicitly deployed for the United Kingdom Construction Industry (CI) and case studies [14].

Meanwhile, the deployment of performance measurement tool in environmental quality assessment management is undergoing a significant transformation with the model-driven DSS, including Charkha and Jaju (2020) with the establishment of an Analytical Hierarchical Process (AHP)-DSS for determining the performance of the supply chain in the textile industry [15]. Govindan et al. (2020) created a practical decision support system based on physicians' knowledge and expertise complemented by a fuzzy inference system to control the supply chain demand [16].

Preliminary studies on SMEs such as Doltsinis et al. (2020) have been attempting to introduce additional information on the TOPSIS-DSS to repurpose manufacturing systems products [17]. A DSS-based fuzzy credit risk evaluation is recently carried out by Chang et al. (2020) using a logistic regression classifier [18]. Furthermore, the finding revealed no consideration of SME e-readiness with the growth of DSS above. There is a limitation in the DSS approaches that encourage a wide variety of crises to benefit from this paradigm's flexible solutions. In addition, Pandey et al. (2020) reviewed a range of challenges in adopting the DSS tool, including the organizations unable to set appropriate metrics to measure market impact, insufficient information about SMEs context measurement, and lack of oversight accountability executive management [19].

Furthermore, it is crucial to create a comprehensive performance as well as an appropriate metrics system as a framework to support the next generation decision concerning critical organizational goals in the e-readiness of the SME industry. A DSS should have an in-depth research identity of the analytical attribute such as a business intelligence dashboard model to simplify the operational phase of DSS and human cognitive capacities by incorporating graphical user interfaces to attract management engagement. A dashboard feature leads to higher task performance, reduces situation awareness, and nourishing a potential out-of-the-loop problem. In a nutshell, its adoption in this study offers the management decision-makers interactive analytical knowledge and will indeed be able to actively recognize the vulnerabilities SMEs face towards effective and sustained digitalization. In addition, the involvement of multi stakeholders' perspectives on digital DSS evaluation provides them with monitoring and surveillance platforms to ensure SME e-readiness achievement and successful acquirement.

III. Methodology

This study is developed by describing diversity as an instrument that examines SMEs' digital readiness through documentary research and interviews. Here, scholars asked four academicians from the reference sectors of economics and technology in both domains for their views on proposed digital preparation tools. Government officials and key actors represent the views of true collaborative partners in the creation of SME e-readiness including the Ministry of cooperatives and, SMEs, the Ministry of Communication and Information Technology in the central region of Java, Riau, investors from a telecommunication company, and central banks, marketplaces (Tokopedia and blanja.com), conventional and digital SMEs companies leading the process of ideas for establishing requirements. An unbiased online interview addresses concerns and issues related to digital adoption, Covid-19 results, SMEs' e-readiness, and factors affecting digitalization as well. For this reason, the suggested digital readiness evaluation criteria are summarized in Table I.

Next, the topography and technology practice against economic views has developed a conceptual framework that encompasses BA-IT Culture, Education, Financial Resources, and BA-IT Infrastructure for BA construction. TC is enrolled in TC-IT Culture, Education, Financial Resources, and TC-IT Infrastructure. MC-IT Culture, Education, Financial Resources, and MC-IT Infrastructure for MC. MG-IT Culture, Education, Financial Resources, and MG-IT Infrastructure for MG. MI-IT Culture, Education, Financial Resources, and MI-IT Infrastructure for MI. Finally, the ME is reviewed from ME-IT Culture, Education, Financial Resources, and ME-IT Infrastructure. This proposed framework grows into a foundation for designing the instruments. In this sense, two sorts of questionnaires are delivered to capture the respondents' information and knowledge concerning SMEs' evaluation. The first review is meant for decision-makers to assign the priority to each criterion and sub-criteria. The questionnaire fits the fuzzy AHP format of a one to nine scale, one just as equally important, three for moderately important, five for strongly important, seven for very strongly important, and nine for extremely important [25]. A panel discussion invited a community of decision-makers from four academies, five government officials, two managers of the central bank and marketplace, three investors, and three representatives from both digital and traditional SMEs to eliminate the language barriers and intervention in interpreting the questionnaires and thus, the weight of the criterion is defined.

Meanwhile, the second review has been conducted as a digital readiness assessment tool, and the questionnaire investigates the level of consensus amongst SME players on the nature of their SMEs' criteria and sub-criteria. As key players in SMEs business, the respondents are constrained to owner, service, management, operation, marketing, and production team. In this regard, 24 questions with five Linkert-scale were submitted to 118 SMEs in Riau Province and limited to SMEs food companies. The companies were considered as one of Indonesia's government focuses on creative business development, and they have a low significant impact on the spread of Covid-19 in Indonesia SMEs. Thus, it tends to increase the demand due to panic buying and social distance restrictions.

TABLE I. SMEs E-Readiness Criteria

Economical Perceives		
No.	***Criteria***	***Reference***
1	Business activities (BA) - BA is highlighted by production, distribution, and consumer protection activities.	[2] [20]
2	Transaction (TC) is spelled out as market practices on automated data collection and distribution of innovative manufacturing systems and resources from the text, sounds, and visual images.	[21] [22]
3	Marketing (MC) is reflected as ideas, products, services, technology, and refining ideas to fulfill the market demand, to determine and satisfy the needs and target market aspirations.	[23] [24]
4	Management (MG) is the motivation of SMEs to plan, govern, organize, control, maintain, and sustain the business.	[18]
5	Micro Environment (MI) is defined as internal factors that directly or do not affect the digital readiness of an SME, such as the stakeholder perceptions, marketing systems, productivity, management, operational functions, technology adoption, the market demands, policy/strategy/vision, pricing, and licensing.	[18] [2]
6	Macro Environment (ME) is characterized as external influencing factors of SMEs, viz. demographic, economic, technology, environmental, political, government support, cultural, and a broader competitor landscape.	[11] [12]
Technological Perceives		
1	IT-Culture is defined as social culture, awareness, knowledge, creativity, attitude, and behavior to emotionally adapt to digitalization, satisfaction, trust, and confidence.	[11][14] [8]
2	IT-Education is specified as the readiness to cultivate hard and soft skills for digital technology adoption, knowledge sharing capability, creativity, and innovation.	[11][14] [8]
3	IT-Financial Resources is set up as SMEs' financial asset readiness in terms of loans or capital assistance, grants, and resources supported programs through training, seminars, and exhibitions.	[5]
4	IT-Infrastructure is characterized as SMEs' readiness concerning the IT infrastructure, hardware, and software to bolster digitalization, time access, and its utilization.	[6]

Furthermore, Covid-19 offers SME food companies a chance to expand through digital transformation and adoption to lower social business interaction. The DSS dashboard system is then applied as a business intelligent model-driven DSS to automate this study's mechanism, starting with acquiring expert knowledge and team research and ending with a comprehensive review of the e-readiness assessment in survey two. Adopting a hybrid F-AHP and OMAX approach enhanced this DSS dashboard's capability as an intelligent decision-making tool for SMEs' e-readiness testing. The application is developed using Hypertext Preprocessor (PHP) and PHPMyAdmin for programming languages and MySQL databases. Blackbox and User Acceptance Testing (UAT) are designed to monitor end-users performance and compliance alongside application development.

A. F-AHP

One of the tests performed on the model-driven DSS is multi-criteria decision making (MCDM), and it has used careful analysis of decision-making processes such as AHP, Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS), Weighted Product (WP), Simple Additive Weighting (SAW), Multi-objective optimization based on ratio analysis (MOORA). Among the above approaches, AHP is one of the most widely encountered performance tests of model-driven DSS as a single or hybrid AHP. It is a robust and practical scientific procedure that minimizes terrible decision-making [26].

Furthermore, it optimally advances the amount and contribution of what-if scenarios and their granularity to the consistency of proposed alternatives [27]. In other words, it provides a powerful cascading multi-criteria decision-making tool for aggregate and objective purposes. Tariq et al. (2020) studied its capability to render the classification of unbiased and confidential parameters and aggregate the specified versatility by relative calculation [28], while Okfalisa et al. (2018) and Okfalisa et al. (2021) successfully gave its application approach to encourage and track the organization's output through the set priority values for Key Performance Indicators (KPIs) [29 and 30]. AHP attempts the preferences more than one of the amazing series selected from one to nine range for scaling based on the perceptions made by the decision-makers [31].

Nonetheless, decision-making practices and rational thinking are hard to pinpoint into a constraint scale and produce victory over language problems in translating the decisions. Furthermore, the persistent matrices for criteria don't launch completely, and information disaster occurs when good and bad scores compensate [32]. In addition, the number of criteria and alternatives used in the measurement also affects the AHP's difficulties.

In a nutshell, the F-AHP emerges as a revolutionary approach in facing the disadvantages of AHP by emphasizing the decision-making uncertainty, the fuzziness of human reasoning style, and linguistics terms issues [33]. Furthermore, it's triangular fuzzy is sufficient in altering the algebraic language of decision-makers to the level of significance of criteria [34 and 20]. It is chosen as an intuitive model since this study captures the respondents' knowledge through natural language, to deploy the weighting criteria and alternatives for digital readiness assessment. It also offers a reasonably stable mechanism to reduce the problems defined by the weighting factors. Its integration with OMAX develops into a new task and progress in this study to increase the revelation of F-AHP for digital readiness analysis. Subsequent studies have tried to link MADM, such as AHP, with OMAX [29 and 21]. However, it is not currently able to discover F-AHP convergence for OMAX. This assimilation of fuzzy logic and OMAX appears as a new contingency in performance measurement analysis.

The employment of Dashboard DSS-based F-AHP in this study follows the formula below [30].

- It follows the AHP modeling method to determine the comparison matric and calculate the consistent values ($CR \leq 0.1$) of the criteria and sub-criteria parameters.
- It then converts the AHP comparison matrix's value into the F-AHP value with the Triangular Fuzzy Number (TFN) scale. TFN is a trapezoidal fuzzy number that is widespread in F-AHP studies. It provides a linear mathematical formula to measure the criteria and alternatives with fuzzy triangular numbers that are denoted in simple parameters.
- To calculate Fuzzy Synthesis's value (Si) where M is defined as TFN number, m for the numbers of criteria, i and j as row and column of the matrix, and p as the value of variables defined in l, m, and u values.

$$Si = -1 \sum_{j=1}^{m} M_{gi}^{j} x \left[\sum_{i=1}^{n} \sum_{j=1}^{m} M_{gi}^{j}\right] \quad (1)$$

- To determine the vector value (V), M2 = (l2, m2, u2) $\geqslant$ M1 = (l1, m1, u1) are defined as a vector value.

$$V\,(M_2 \geq M_1) = \sup\,[\min\,(\pi M_1\,(x)), \min\,(\pi M_2\,(y))] \quad (2)$$

$$V(M_2 \geq M_1) = \begin{cases} 1\ if\ m_2 \geq m_1 \\ 0\ if\ l_1 \geq u_2 \\ \frac{l_1 - u_2}{(m_2 - u_2) - (m_1 - l_1)} \end{cases} \quad (3)$$

- To calculate De-fuzzification Ordinary Value (d').

For k = 1, 2, n; k $\neq$ I, we obtain the vector weight value:

$$W' = (d'\,(A1), d\,'(A2), \ldots\ldots d'\,(An))\,T \quad (4)$$

- To multiple the normalization value of fuzzy vector weights (W) by the following formula.

$$Sd(A_n) = \frac{d'}{\sum_{i=1}^{n} d(A_n)} \quad (5)$$

$$W = (d\,(A1), d\,(A2), \ldots, d\,(An))\,T \quad (6)$$

- Where W is a non-fuzzy number.
- It is ending with the calculation of the final weight of recommendation.

B. OMAX (Scoring)

OMAX is a device scoring tool based on the default KPI outcomes by keeping the matrix in a single measure, and it allows the data interpretation to be relatively simple, easy to understand, flexible, and supports the units' normalization from different measurement specifications. A test system with OMAX balances each indicator's value scale and, as a result, each parameter's similar objective level is accomplished and compared to determine its position [21]. The previous studies observed on OMAX, including Paduloh and Hardi (2020) that measured the company productivity based on the automotive paint industry's KPIs [22] and, Margareta et al. (2020) integrated OMAX and AHP to measure the Green Warehouse KPIs [23]. The above studies have clarified how OMAX and AHP allow optimum achievement of performance assessment. Therefore, this study aimed to explore future OMAX values through its conjunction with the latest F-AHP levy to assess SMEs' digital readiness.

Here, OMAX as the operating system works by the proportion of each KPIs or criteria based on the total of the intermediate value divided into three levels viz highest, middle, and lowest level complemented by the traffic light system method as depicted in Fig. 1.

Part A in Fig. 1 describes the parameters influencing the efficiency and performance of the KPI, while Part B quantifies the degree of performance from the highest (10) to the lowest (0) by considering class and interval interpolation values.

$$\Delta X_{L-H} = \frac{Y_H - Y_L}{X_H - X_L} \tag{7}$$

The interval between high and low levels is the numbers at high levels, YL as numbers at low levels, XH as variables for the highest levels, and XL as variables for the lowest levels. Part C exemplifies the value analysis of KPIs that is outlined by the following formula.

$$value = level \times weight \tag{8}$$

Meanwhile, the index variable shows the initialization of criteria as the ultimate measures of efficiency. The OMAX traffic light system remarks the KPIs were levelling with three colors: green, yellow, and red. The green rates the scale level ranged from 6 to 10, where a good category is rated from 6 to 7, the very good type is enumerated by 8, and the excellent class is racked up from 9 to 10. The yellow color warned the measurement condition far from reaching the goal and scored from 2 to 5. It prescribes the level values 2 to 3 as bad category and the values 4 to 5 as the average order. The red color indicates the below-targeted achievement, thus requiring immediate improvement, and in this case, the OMAX curves are 0 to 1, with 0 indicating the rating class and 1 the terrible section. This traffic light system approach has been successfully solved in many labelling cases. Furthermore, it is found that the traffic light system is consistent, reduces confusion, and wastes time in determining the performance of KPIs, and as a result, it boosts the labelling criteria's efficiency in digital readiness assessment.

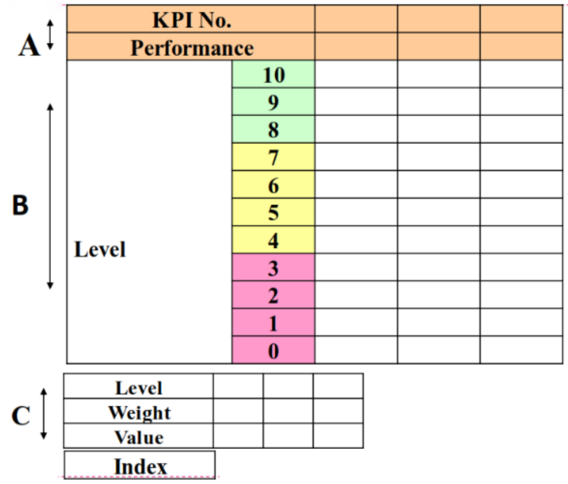


Fig. 1. OMAX Method Assessment Scheme [42].

IV. Result

A. KPIs Weighting Analysis – F-AHP

F-AHP resolves a questionnaire and appraises the significant weight values of SMEs' digital readiness parameters following Equation (1) to (6), and Table II is obtained in this sense. Initially, the AHP computing of comparison matrices and consistent values has led to the reasonableness and consistency of SMEs' entire digital readiness assessment parameters with the values 0.10, 0.08, 1.24, and 6.51 for CI, CR, Ratio Index (RI), and "λ maks" respectively. Hence, the parameter weighting values rank the criteria in Transaction (TC) of 0.29, Marketing (MC) and Micro Environment (MI) of 0.22, Management (MG) of 0.15, Macro Environment (MA) of 0.07, and Business Activity (BA) of 0.05, respectively. Next, the fuzzy synthesis values resulting from the operation performed on TFN are defined by Table II. Furthermore, from equations (4) to (6), Table II has elucidated the removal of fuzziness from the values of the ordinary and fuzzy vector weights (w). Thus, it attributes the summary of the priority weight value of the indicator and sub-indicators with the transaction (TC) as the most significant indicator (0.30) followed by Marketing (MC) and Micro Environment (MI) to a vector value of 0.24, Management (MG) provides 0.20, Macro Environment (MA), and Business Activity (BA) correspond to the vector values of 0.03 and 0.02, respectively. Furthermore, the weighted analysis reveals that TC-Education contributes to the weight values at 0,54, as the main sub-indicators of the transaction (TC). TC-Financial Resource, TC-Technical Infrastructure, and TC- Culture determines the vector values of 0.33, 0.13, and 0.08, respectively. Therefore, the priority indicators and sub-indicators for the SMEs digitization assessment model are established between SMEs and their stakeholders. This model as a performance measurement instrument is recommended as an aid for further analysis.

TABLE II. THE RECAPITULATION OF F-AHP PRIORITY FOR SMES DIGITAL READINESS VARIABLES AND SUB-VARIABLES

Variable	W	Si			Sub Variable	Si		
		L	*M*	*U*		*L*	*M*	*U*
Business Activity (BA)	0.02	0.06	0.09	0.18	BA-Culture	0.06	0.09	0.14
					BA-Education	0.43	0.25	0.41
					BA-Financial Resources	0.43	0.16	0.28
					BA-Technical Infrastructure	0.14	0.10	0.17
Transaction (TC)	0.30	0.09	0.17	0.31	TC-Culture	0.08	0.09	0.15
					TC-Education	0.54	0.28	0.44
					TC-Financial Resources	0.33	0.12	0.23
					TC-Technical Infrastructure	0.13	0.10	0.18
Marketing (MC)	0.24	0.09	0.17	0.32	MC-Culture	0.21	0.10	0.18
					MC-Education	0.27	0.19	0.35
					MC-Financial Resources	0.27	0.13	0.26
					MC-Technical Infrastructure	0.26	0.11	0.21
Management (MG)	0.20	0.08	0.16	0.29	MG-Culture	0.13	0.09	0.15
					MG-Education	0.31	0.20	0.35
					MG-Financial Resources	0.26	0.11	0.20
					MG-Technical Infrastructure	0.31	0.17	0.30
Micro Environment (MI)	0.24	0.19	0.29	0.46	MI-Culture	0.19	0.13	0.17
					MI-Education	0.36	0.14	0.28
					MI-Financial Resources	0.09	0.13	0.16
					MI-Technical Infrastructure	0.36	0.22	0.38
Macro Environment (MA)	0.03	0.06	0.11	0.21	MA-Culture	0.36	0.14	0.28
					MA-Education	0.09	0.13	0.16
					MA-Financial Resources	0.36	0.22	0.38
					MA-Technical Infrastructure	0.13	0.08	0.11

B. OMAX-Labeling Analysis

The OMAX labelling emphasized the second survey analysis that delivered 118 questionnaires for food industrial SMEs in Riau Province. It rated SMEs' degree of performance hinged on the F-AHP estimation weight values and following the Equations (7) and (8). As a result, Table III revealed the recapitulation of SMEs' readiness at Riau Province. Furthermore, the survey found that the highest criteria value for SMEs' digital readiness in Riau is Marketing (MC) with "Good" (7.75) and an "Average" for the entire MC sub-criteria. It then was followed by Macro Environment (MA) for index level of "Average" (5.34), Management (MG) for the index degree of "Average" (4.68), Micro Environment (MI) with index "Average" (4.32), Transaction (TC) with the category of "Bad" (3.84), and Business Activity (BA) with the class of "Bad" (3.78), respectively. Besides, Table III further describes the calculation performance analysis for each sub-criterion.

Fig. 2 and Table IV display the comprehensive evaluation of SMEs' digital readiness in the Riau province. In Fig. 2, out of the 118 SMEs in Riau Province, their performance was 42.37%, 36.44%, 11.86%, 5.08%, and 4.24% at the "Average" level, "Bad" degree, "Good," "Very Good," and at "Excellent" group respectively. The thorough levelling index of 118 SMEs for BA, TC, MC, MG, MI, and MA parameters is illustrated in Table IV. The root cause analysis is then addressed according to each sub-criterion indication. Thus, the stakeholders are advised in taking curative action.

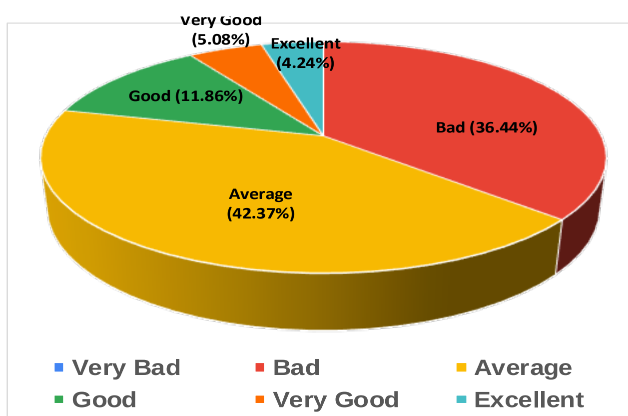


Fig. 2. The Percentage of SMEs Digital Readiness at Riau Province.

TABLE III. THE RECAPITULATION OF OMAX FOR SMES DIGITAL READINESS IN RIAU PROVINCE

Variable	Sub-Variable	Achievement	OMAX Level	Performance Value	OMAX Category	Index
BA	BA-Culture	58.47	3	0.18	Bad	3.78 (Bad)
	BA-Education	60.17	3	1.29	Bad	
	BA-Financial Resources	55.30	3	1.29	Bad	
	BA-Technical Infrastructure	62.50	3	1.02	Bad	
TC	TC-Culture	62.5	3	0.24	Bad	3.84 (Bad)
	TC-Education	58.90	3	1.62	Bad	
	TC-Financial Resources	58.26	3	0.99	Bad	
	TC-Technical Infrastructure	61.02	3	0.99	Bad	
MC	MC-Culture	69.70	5	2.15	Average	7.75 (Good)
	MC-Education	65.47	4	1.88	Average	
	MC-Financial Resources	65.04	4	1.88	Average	
	MC-Technical Infrastructure	65.25	4	1.84	Average	
MG	MG-Culture	66.74	4	1.32	Average	4.68 (Average)
	MG-Education	62.92	3	0.99	Bad	
	MG-Financial Resources	60.81	3	1.38	Bad	
	MG-Technical Infrastructure	61.86	3	0.99	Bad	
MI	MI-Culture	58.48	3	1.17	Bad	4.32 (Average)
	MI-Education	65.04	4	1.44	Average	
	MI-Financial Resources	58.48	3	0.27	Bad	
	MI-Technical Infrastructure	64.62	4	1.44	Average	
MA	MA-Culture	66.95	4	1.32	Average	5.34 (Average)
	MA-Education	65.25	4	1.8	Average	
	MA-Financial Resources	58.26	3	1.11	Bad	
	MA-Technical Infrastructure	62.92	3	1.11	Bad	

TABLE IV. THE ASSESSMENT INDEX OF SMES DIGITAL READINESS AT RIAU PROVINCE

SMEs No.	Index	Category	SMEs No.	Index	Category	SMEs No.	Index	Category	SMEs No.	Index	Category
SME1	2.36	Bad	SME10	3.14	Bad	SME19	3.81	Bad	SME28	3.64	Bad
SME2	6.94	Good	SME11	6.66	Good	SME20	2.63	Bad	SME29	4,6	Average
SME3	3.64	Bad	SME12	4.46	Average	SME21	5.28	Average	SME30	8.82	Very Good
SME4	5	Average	SME13	2.59	bad	SME22	7.34	Good	SME31	4.87	Average
SME5	6.56	Good	SME14	2.87	bad	SME23	8.1	Very Good	SME32	3.35	Bad
SME6	6.14	Good	SME15	5.92	Average	SME24	3.46	Bad	……	……	……
SME7	4.69	Average	SME16	4.92	Average	SME25	4.43	Average	SME116	4.28	Average
SME8	4.5	Average	SME17	9.82	Excellent	SME26	4.44	Average	SME117	4.28	Average
SME9	3.25	Bad	SME18	3.14	Bad	SME27	5	Average	SME118	5.26	Average

C. Dashboard System Development

Fig. 3 shows the SMEs Readiness Assessment Dashboard application architecture and design setup. It is designed according to the approach and components of the DSS framework and also eligible for two key players, which include; users derived from decision-makers comprising researchers, government engagement, the central bank, marketplaces managers, investors, and also participants designated for the end-users from 118 SMEs in the province of Riau. This application is designed to provide stakeholders with an integrated and interactive platform to rack up the weight on F-AHP and make predictions for SME digital readiness predictions. Here, two questionnaire mechanisms are built through this application. It provides an exact word, video, and text formatted to eliminate gaps when performing the job. The system interfaces are equipped with the information and knowledge activities, so focusing on the outcome of analyzes and tracking of the SMEs readiness assessment as well as scoring issues. Finally, the assessment analytical report is performed by the system as intelligent recommendations in tackling the problems identification for the stakeholders and SMEs actors as well. Fig. 4 explains the general SMEs analysis report as outcome report generate that exhibiting the performance achievement of indicators and constructs in terms of BA, TC, MC, MG, MA, and MI. Besides, the index performance percentage is also displayed as resume index of the whole SMEs in Riau Province. The graphical performance of indicators and constructs informs the SMEs achievement and provide the recommendation on how to bring about the optimal fulfillment. Fig. 5 is public OMAX calculation page in disclosing the SMEs mapping performance calculation within 10 leveling indexes into overall performance calculation. Fig. 6 resolves the performance calculation index for one SME case study namely Kebab Frozen. This page describes the graphic performance for indicators and constructs, OMAX calculation, and overall performance analysis whether this SME in good, bad, or average leveling index. Lastly, Fig. 7 expounds the mapping performance of SMEs in Riau Province based on the google map location. The rules' foundation is based on the result of DSS analysis and displayed as user interface results.

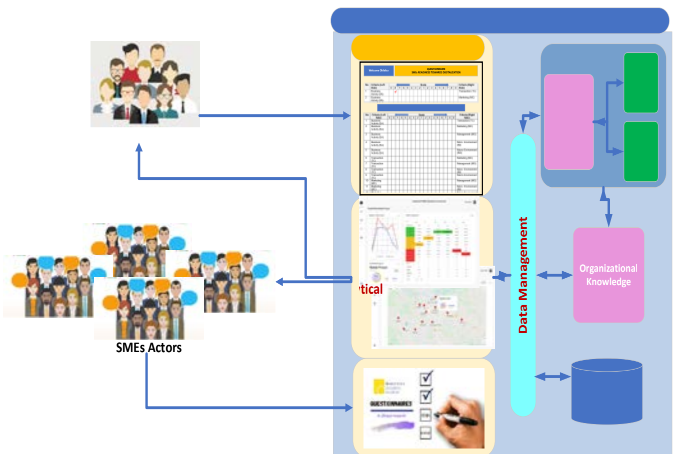


Fig. 3. System Architecture for Dashboard SMEs Readiness Assessment.

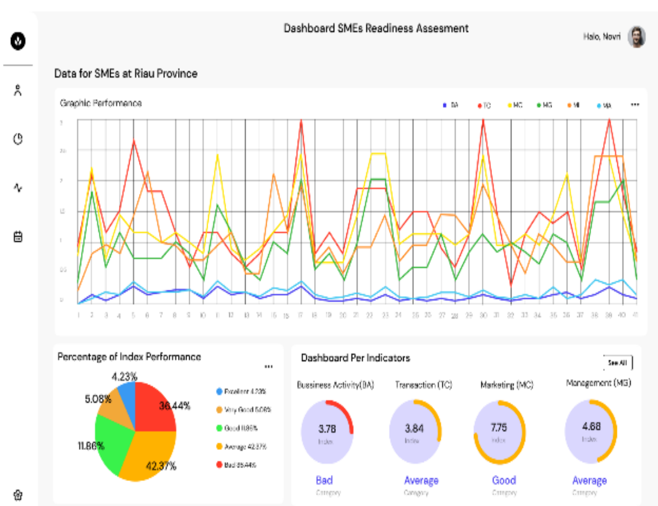


Fig. 4. General SMEs Scoring.

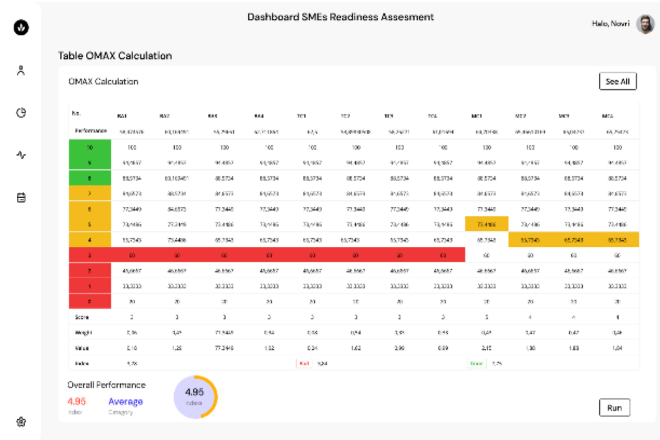


Fig. 5. General SMEs OMAX Calculation.

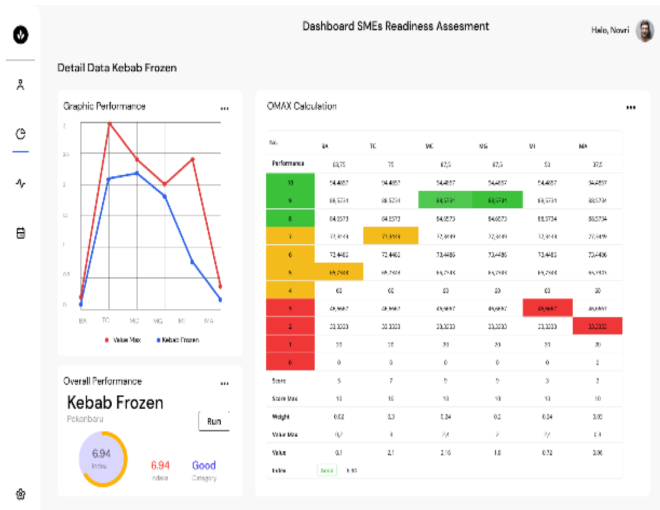


Fig. 6. SMEs OMAX Calculation for Case Study: Kebab Frozen.

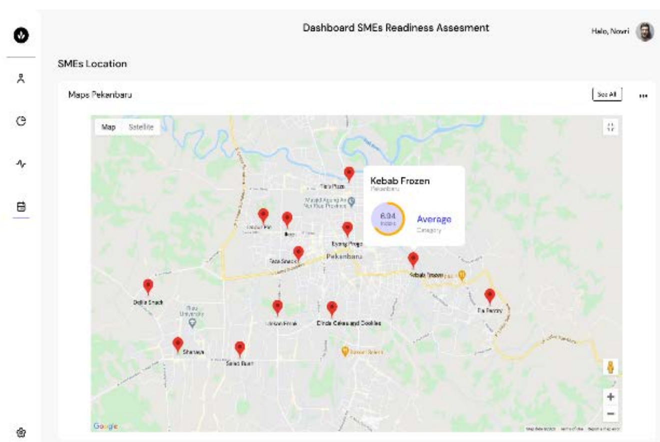


Fig. 7. SMEs Mapping Performance at Riau Province.

Furthermore, A Blackbox software evaluation was successfully conducted, and 128 functions were correctly completed to test the reliability and functionality of the dashboard, including the three parts of the AHP-questionnaires, the three aspects of Assessment-questionnaires, and the three of the SME General Analysis Task, one of SMEs surveying and mapping functions, and 118 tasks of SMEs analysis. In addition, a UAT survey is carried out by a five Linkert Scale questionnaire. In terms of user adoption of applications and functions, user-friendly applications, system interfaces, and utilization, 22 stakeholders and 118 SME participants were required to participate.

The survey showed that 85% of the respondents had no issues with the dashboard, and the application was designed to be user-friendly. The interface provided was very attractive, and 88% of respondents reported that the dashboard was constructive and will provide better clarity and understanding when reviewing formatted content, graphs, and tables. Concerning the dashboard utility, 90% of respondents have suggested that this application boost SMEs' efficiency for digital businesses' success. Furthermore, the dashboard examines SMEs' potential growth in Riau Province by defining policies, strategies, and supporting measures to improve digital business achievement.

V. Discussion

In this study, F-AHP weighting has been successfully identifying the outstanding factors in assessing the digital readiness of SMEs. F-AHP in DSS approach cascades the assessment of multi-criteria decision making in digital readiness constructs into detailed sub-indicators weighting and model development. For instance, this model provides transaction (TC) as the index of the top-weighted value down warding into sub-criteria TC-Education and TC-Financial Resource as the main criteria that would directly influence the general performance index of SMEs. In a nutshell, SMEs must take a comprehensive approach to the weighted priority of criteria and sub-criteria to enhance a sophisticated digital readiness index achievement. This assessment advises stakeholders to enforce these systems to support digitally-driven transaction settings by ultimately achieving the highest criteria and sub-criteria, such as TC-Financial Resources, Education, Culture, and TC-Technical Infrastructure. Furthermore, the issuance of regulations, policies, and understood the transaction-driven procedure, security, quality, and protections develop into the government's consideration and accountability of customers and SMEs practitioners. In addition, the adoption of technologies requires that entrepreneurs and innovative SMEs transaction flow encourage the forecasting of the future of networking, collaboration partnerships, shared services, expansion programs approaching digitalization, as well as funding requests from others both internally and externally, thus allowing digital commerce to thrive [12]. Besides, the digital transaction's success requires the involvement of leaders and empowerment by external stakeholders and employees in promoting the environment and technology in business production, management, marketing, technical infrastructure, skills, and resources [6]. The above initiative is considered growing into the greatest danger in SMEs' digitalization. Moeuf et al. (2018) substantiated that those managers and external experts play a unique role in the success and failure of a digitalized SME project [24].

Meanwhile, beholding on the OMAX calculation analysis index whereby 42.37% of SMEs in "Average" and 36.44% in "Bad" status indicates that SMEs in the Province of Riau need to make a great deal of digital sustainability commitment. Herein, the OMAX scoring and leveling index has been flourishingly accurate in mapping the SME's performance and achievement. The integration of F-AHP and OMAX equip a deep and cascading performance analysis by considering the customized perceptions of stakeholders [35]. The Riau province condition reflects the generalization of SMEs' performance on Sumatra Island due to the possible growth of SMEs on this island. Furthermore, its geographical position is adjacent to the entrance and business transactions with several Southeast Asian countries such as Malaysia, Singapore, and Thailand, providing a considerable bonus opportunity in this area for accelerating the digital businesses, finance and investors, partnerships, and networking.

Corresponding to the dashboard system development and interface are interactively designed to provide decision-makers with a helpful overview and easy to interpret using gauges, graphs, traffic light strategies, and map analysis. According to Nadj et al. (2020), the dashboard's reciprocal analytical features strengthen the DSS operational process [36]. Herein, the model-driven DSS component has been software implemented in a hybrid analysis stage using F-AHP and OMAX approaches, and the integration of these in decision-making allows for better and optimized benefit analysis. Furthermore, OMAX and fuzzy blending bring new and meaningful success measurement values [11]. Meanwhile, data management alongside organizational knowledge is processed in the knowledge repository according to the SME's digital preparation model's parameters and sub-parameters. Hence, the knowledge recommendation analysis furnishes a new and innovative problem-solving way to aid SMEs in increasing and conducting the curative actions towards their successful preparedness for the sustainability of the digital business era.

VI. Conclusion

The era of the new-normal has accelerated the implementation of digitization around the world. To sustain the nation's economy, Indonesia has made various breakthroughs to achieve digital success in all facets of life, especially for the SMEs industry, which plays an essential role in its economy. Despite SMEs having the slightest effect from the Covid-19 spread, they still need to adopt various solutions and digital acceleration to survive. The high risk of digital transformation and adoption leads the stakeholders to take action as well as digital preparation evaluation by considering the views perceived of both supply chain activities and technology. This study has successfully developed a DSS dashboard that measured SMEs' readiness towards digital business from 6 main variables and 24 sub-variables sets as criteria of SMEs' e-readiness model. The model-driven DSS application has effectively analyzed the requirements and employs the F-AHP to generate priority weight, and the views of stakeholders from government, academicians, investors, marketplaces, banking, and SME participants are considered weighted judgment in the dashboard calculation.

Furthermore, the OMAX analysis has empirically mapped 118 SMEs in Riau province with a levelling index and root cause analysis. The study reveals that SMEs' overall performance in Riau is measured by an average index, requiring more government and stakeholders' attention for corrective actions. It shows that the overall achievements of SMEs in Riau are below the average index, which requires further consideration by the governments and stakeholders' corrective action. In terms of accessibility and utility, this dashboard application was tested and confirmed to advise decision-makers in addressing the current issues based on their review findings and the feedback presented.

Acknowledgment

This study received a grant from Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia, through the DIPA funding number. 1171 / R / 2019. The authors are grateful to the support parties in this study, including the Ministry of Cooperatives and Small and Medium Enterprises, the Ministry of Communication and Information Technology, management of market places, Telecommunication Company at Pekanbaru and Central Jakarta, Department of Communication, Informatics, and Statistics of Riau Province, Integrated Business Services Center of Riau Province, Central Banks, Faculty of Science and Technology, Faculty of Economics and Social Sciences of Sultan Syarif Kasim Riau Islamic University, and several SMEs in Riau Province.

References

[1] Jebreen, I., and Al-Qerem, A., "Empirical Study of Analysts' Practices in Packaged Software Implementation at Small Software Enterprises," International Arab Journal of Information Technology (IAJIT), vol. 14, 2017.

[2] Umeh, F., Tan, C., Pan, Y. C., and Khan, H. T., "Developing the digital business readiness assessment framework (DBRAF) for fashion retail SMEs in Lagos, Nigeria," 2020.

[3] Kominfo, "Indonesian Internet Users World Number Six", https://kominfo.go.id/content/detail/4286/pengguna-internet-indonesia-nomor-enam-dunia/0/sorotan_media, 2017.

[4] Hoellthaler, G., Braunreuther, S., and Reinhart, G., "Digital Lean production an Approach to Identify Potentials for the Migration to a Digitalized Production System in SMEs from a Lean Perspective," Procedia Cirp, vol. 67, pp. 522-527, 2018.

[5] Peillon, S., and Dubruc, N., "Barriers to digital servitization in French manufacturing SMEs," Procedia CIRP, vol. 83, pp. 146-150, 2019.

[6] Singh, T., Kumar, R., and Kalia, P., "E-marketing Practices of Micro-, Small-and Medium-sized Enterprises: Evidence from India. In Strategic corporate communication in the digital age," Emerald Publishing Limited, 2021.

[7] Tavana, M., Mousavi, S. M. H., Mina, H., & Salehian, F. "A dynamic decision support system for evaluating peer-to-peer rental accommodations in the sharing economy", International Journal of Hospitality Management, 91, 102653. doi:10.1016/j.ijhm.2020.102653. 2020.

[8] Klör, B., Monhof, M., Beverungen, D., and Bräuer, S., "Design and evaluation of a model-driven decision support system for repurposing electric vehicle batteries," European Journal of Information Systems, vol. 27, pp. 171-188, 2018.

[9] Ilbiz, E., and Durst, S., "The appropriation of blockchain for small and medium-sized enterprises," Journal of Innovation Management, vol. 7, pp. 26-45, 2019.

[10] Castagna, F., Centobelli, P., Cerchione, R., Oropallo, E., and Strazzullo, S., "Assessing SMEs' Internationalisation Strategies in Action," Applied Sciences, vol. 10, pp. 4743, 2020.

[11] Khalfan, M. M., Anumba, C. J., and Carrillo, P. M., "Development of a readiness assessment model for concurrent engineering in construction," Benchmarking: An International Journal, 2001.

[12] Ruikar, K., Anumba, C. J., and Carrillo, P. M., "VERDICT—An e-readiness assessment application for construction companies," Automation in construction, vol. 15, pp. 98-110, 2006.

[13] Lou, E. C., Lee, A., and Goulding, J., "E-readiness in construction (ERiC): self-assessment framework for UK small and medium enterprise building services providers," Architectural Engineering and Design Management, vol. 16, pp. 3-22, 2020.

[14] Goulding, J., Lou, E., and Lee, A., "E-readiness in construction (ERiC): self-assessment framework for UK small and medium enterprise building services providers", 2019.

[15] Charkha, P. G., and Jaju, S. B., "Decision support system for supply chain performance measurement: case of textile industry," New Paradigm of Industry 4.0, pp. 99-131, 2020.

[16] Govindan, K., Mina, H., and Alavi, B., "A decision support system for demand management in healthcare supply chains considering the epidemic outbreaks: A case study of coronavirus disease 2019 (COVID-19)," Transportation Research Part E: Logistics and Transportation Review, vol. 138, pp. 101967, 2020.

[17] Doltsinis, S., Ferreira, P., Mabkhot, M. M., & Lohse, N., "A Decision Support System for rapid ramp-up of industry 4.0 enabled production systems," Computers in Industry, vol. 116, pp. 103190, 2020.

[18] Chang, Y. C., Chang, K. H., and Huang, Y. H., "A novel fuzzy credit risk assessment decision support system based on the python web framework," Journal of Industrial and Production Engineering, vol. 37, pp. 229-244, 2020.

[19] Pandey, N., Nayal, P., & Rathore, AS. "Digital marketing for B2B organizations: structured literature review and future research directions". Journal of Business and Industrial Marketing. 35(7), 1191–1204. 2020.

[20] Phuc, V. N., Phong, T. N., Quyen, L. H.T.T.N., and Vy, D.B.H., "Extended Fuzzy Analytical Hierarchy Process Approach in Determinants of Employees' Competencies in the Fourth Industrial Revolution," International Journal of Advanced Computer Science and Applications, vol. 10, no.4, 2019.

[21] Cruz-Rivero, L., Angeles-Herrera, D., Méndez-Hernández, M. L., Rivera-Toscano, C. D., and Mar-Orozco, C. E., "Omax and Fuzzy Logic as Productivity Tools and Ergonomic Analysis in Automotive Maintenance," International Conference on Applied Human Factors and Ergonomics, pp. 357-364, 2020.

[22] Paduloh, P., "Analysis of Productivity Based on Kpi Case Study Automotive Paint Industry," Journal of Engineering and Management in Industrial System, vol. 8, pp. 1-12, 2020.

[23] Margareta, W., Ridwan, A. Y., and Muttaqin, P. S., "Green Warehouse Performance Measurement Model for 3PL Warehousing," Proceedings of the 3rd Asia Pacific Conference on Research in Industrial and Systems Engineering 2020, pp. 180-186, 2020.

[24] Moeuf, A., Pellerin, R., Lamouri, S., Tamayo-Giraldo, S., and Barbaray, R., "The industrial management of SMEs in the era of Industry 4.0," International Journal of Production Research, vol. 56, pp. 1118-1136, 2018.

[25] Saaty, T. L., "Decision making with the analytic hierarchy process," International journal of services sciences, vol. 1, pp. 83-98, 2008.

[26] Eldrandaly, K., and Naguib, S., "A knowledge-based system for GIS software selection," International Arab Journal of Information Technology (IAJIT), vol. 10, pp. 152-159, 2013.

[27] Hodicky, J., Özkan, G., Özdemir, H., Stodola, P., Drozd, J., and Buck, W., "Analytic Hierarchy Process (AHP)-Based Aggregation Mechanism for Resilience Measurement: NATO Aggregated Resilience Decision Support Model," Entropy, vol. 22, pp. 1037, 2020.

[28] Tariq, M. I., Mian, N. A., Sohail, A., Alyas, T., and Ahmad, R., "Evaluation of the Challenges in the Internet of Medical Things with Multicriteria Decision Making (AHP and TOPSIS) to Overcome Its Obstruction under Fuzzy Environment," Mobile Information Systems, 2020.

[29] Okfalisa, Anugrah, S., Anggraini, W., Absor, M., and Fauzi, S. S. M., "Integrated Analytical Hierarchy Process and Objective Matrix in Balanced Scorecard Dashboard Model for Performance Measurement," Telkomnika, vol. 16, 2018.

[30] Okfalisa., Wresni, A., Gusman, N., Saktioto., and Kuan, Y., "Measuring the effects of different factors influencing on the readiness of SMEs towards digitalization: A multiple perspectives design of decision support system," Decision Science Letters Journal, vol. 10, 2021.

[31] Xu, J., Yu, L., and Gupta, R., "Evaluating the performance of the government venture capital guiding fund using the intuitionistic fuzzy analytic hierarchy process," Sustainability, vol. 12, pp. 6908, 2020.

[32] Wu, Z., and Abdul-Nour, G., "Comparison of multi-criteria group decision-making methods for urban sewer network plan selection," Civil Engineering, vol. 1, pp. 26-48, 2020.

[33] Liu, H., Sun, C., and Tu, J., "Application of Intuitionistic Fuzzy Evaluation Method in Aircraft Cockpit Display Ergonomics," International Arab Journal of Information Technology, vol. 18, pp. 856-863, 2021.

[34] Okfalisa, O., Rusnedy, H., Iswavigra, D. U., Pranggono, B., Haerani, E., and Saktioto, T., "Decision support system for smartphone recommendation: The comparison of fuzzy AHP and fuzzy ANP in multi-attribute decision making," Sinergi, vol. 25, pp. 101-110, 2021.

[35] Okfalisa, O., Mahyarni, M., Wresni, A., Faisal, S., Taofeeq, D., M., & Saktioto, S. "Quadruple helix engagement: reviews on shariah fintech based SMEs digitalization readiness". Indonesian Journal of Electrical Engineering and Informatics, vol. 10, pp 112-122, 2022.

[36] Nadj, M., Maedche, A., & Schieder, C. "The effect of interactive analytical dashboard features on situation awareness and task performance". Decision Support Systems. 135, 2020.

A Novel IoT Architecture for Seamless IoT Integration into University Systems

Wafa Altwoyan, Ibrahim S. Alsukayti
Department of Computer Science, College of Computer, Qassim University
Buraydah, Saudi Arabia

***Abstract*—IoT architectures play critical roles in guiding IoT system construction and enhancing IoT integration. However, there is still no standardized IoT architecture that meets the varying requirements of different IoT deployments. Although this has been a focus within the research community, no specific attention has been paid to optimizing an IoT architecture toward seamless IoT integration in educational environments. Moreover, different advanced system aspects have not been considered for designing an optimized IoT architecture. These include the need for complete security and privacy support, a highly responsive system, dynamic interactivity, and wide-range IoT connectivity. Such considerations are important considering the complexity and multidimensionality of integrating IoT in educational environments. In this paper, we introduced a novel IoT architecture with the main objective of facilitating the effective integration of IoT into university systems. It also aims at optimizing the IoT-integrated system with advanced aspects to enhance system security, responsiveness, and IoT connectivity. The proposed architecture provides a modular and scalable design of six architectural layers in addition to a vertical layer that provides security support across the architecture. Only the most relevant and critical layers are added to the architecture to maintain a practical trade-off between effective modularity and less complexity. Compared with other IoT architectures, the proposed one ensures high reliability, data management, full security support, responsiveness, and wide coverage while maintaining acceptable complexity.**

Keywords—Internet of things; architecture; smart campus; education

I. Introduction

Advancements in Information and Communication Technology (ICT) have recently enabled the effective transformation of conventional educational systems to a more efficient and sustainable level. The high dependency on technology and digital services allows substantial investment of recent technologies in enhancing traditional educational methods. The ICT revolution leads traditional learning to be moved to new paradigms such as e-learning and distance learning. This creates opportunities for further improvements regarding the educational activities and processes. Such technological advancements in the education domain can be even more optimized with the best leverage of recent technologies. One of these is the Internet of Things (IoT) technology which is projected to change several facets of the educational environment [1].

IoT provides a technological revolution that enables effective interaction between people, processes, data, and things. It facilitates the effective provisioning of smart services in different domains including healthcare [2], industry [3], and agriculture [4]. In addition, learning is one of the most perceptible human actions affected by IoT, turning the educational process into a revolutionary system in the near future. IoT enables educational environments to be significantly extended by combining physical and digital objects. IoT finds its way into education to provide advanced support in enhancing many aspects. These include streamlining educational processes, making the most use of data, and improving sustainability. IoT can serve as a catalyst for improving the current knowledge-transfer model to be more interactive and collaborative. The teaching-learning process would be actively engaging with the efficient use of IoT in education systems [5].

However, integrating IoT in educational environments is still a challenging process. Careful consideration of the complexity and multidimensionality of such environments is important before going any further through this process. At the fundamental level, architectural planning is the key to optimizing IoT integration in educational environments. The importance of developing IoT architecture lies in its critical role in guiding system construction and facilitating IoT integration. Such a challenging IoT integration can be fundamentally simplified by developing an efficient IoT architecture. Despite the wide adoption of IoT technology in different domains, there is still no standardized IoT architecture that meets the varying requirements of different IoT deployments [6]. Several design approaches have been introduced for the development of an effective IoT architecture. They vary in granularity and the number of architectural layers in addition to considering different system aspects.

However, fundamentally addressing the efficiency and complexity of integrating IoT into the specific environments of educational institutions has not been considered yet. The focus has been only on addressing the main IoT system aspects, particularly data collection, transmission, processing, and application. No particular attention has been devoted to the specific characteristics of educational environments when integrated with IoT. These include the need for complete security and privacy support, highly responsive system, dynamic interactivity, and wide-range IoT-specific connectivity. Still there is a compelling need for effective design of an IoT architecture to facilitate IoT integration and accelerate IoT adoption in these dynamic environments. This paper aims at addressing such a need by establishing an IoT architecture while taking into account advanced technological considerations. These include edge computing, IoT accessibility, cloud computing, and full security integration.

The following section, Section II, of this paper presents the related work. In Section III, an overview of the integration of IoT in education environments is provided. Section IV introduces the proposed IoT architecture. Then, a simple use case demonstrating the implementation of the proposed architecture is presented in Section V. A broad discussion is then provided in Section VI. Section VII concludes this research paper and presents the plan for future work.

II. Related Work

There has been a lack of consensus concerning IoT architectural design. Several IoT architectures have been proposed in the literature with different characteristics and properties. A well-known architecture that serves as a reference IoT architectural model is the three-layer architecture [7-9]. It provides a basic architectural design of the IoT functionality with three logical layers, namely the perception, network, and processing layers. A further improvement was made to this basic design to include the application layer at the top level [10-11]. This enables managing IoT services and applications in a more effective way for the end-users.

With the rapid development of IoT, such basic architectures become insufficient to realize the full functionality of IoT systems in an effective manner. They would not be able to efficiently fulfill the different requirements of emerging IoT applications. They can only be feasible at the initial stage of IoT system development. Therefore, other architectures were proposed with additional layers. One is the five-layers architecture which was introduced in [9] to not only realize the structure of loT from a technical point of view but also a business and operational perspective. It incorporates a business layer at the top of the architecture to effectively enable the management of IoT applications based on certain business models and strategies. The other one is the six-layer architecture proposed in [12]. It includes additional layers between the perception and network layers for IoT data monitoring, preparation, storage, and security.

In addition, there have been different proposals focusing on the establishment of specific IoT architectures considering certain common aspects. These include SDN-based implementation [13] security-specific design [14], QoS provisioning [15], and service-oriented perspective [16]. Others focused on certain IoT communication technology such as the architectures proposed in [17] for provisioning 5G-enabled IoT systems. Customized architectures considering specific IoT applications have also been proposed in the literature. These include smart agriculture control [18], tourism management [19], smart metering [20], e-health [21], building automation [22], and water resource management [23]. However, no particular attention has been yet paid to addressing the effective development of an optimized IoT architecture for educational environments.

It can also be seen that these design approaches provide no adequate support for different important aspects. These include the need for security across the different levels of the IoT systems. It is insufficient to address security at a specific layer of the architecture. Security support is critical at the different levels of data acquisition, access, communication, processing, and management. Another important consideration is the incorporation of advanced technologies that would make the deployed IoT system more sustainable. One of these technologies is edge computing which allows the system to be more responsive with very low latency. This is important considering that educational environments are more dynamic and interactive. Another dimension in this context is the need for an effective way to incorporate IoT-based connectivity into these environments. Remote educational areas can be provided with IoT-customized networking support to implement the different IoT applications and services. Therefore, IoT-based network access control is a critical consideration for the effective design of the IoT architecture.

III. IoT Integration in University Systems

For understanding the challenges of integrating IoT into educational environments, we focus in this research work on the specific and demanding case of a university system. It is critical to look into such a complex process at the fundamental level and begin with investigating IoT integration at the infrastructural scale. This would be the initial stage toward the full realization of IoT-based university educational systems. This section presents and discusses the impact of introducing IoT into conventional university systems.

Fig. 1 shows an overview of a typical global university system that incorporates a complex setup of different networking segments, topological entities, interconnections, and computing resources. It has a core infrastructure that interconnects multiple LANs using wired and wireless connectivity. It also has a separate setup for supporting VPN communication with the core system. Internet access is also maintained using a private ISP link. The system also includes a data center having a farm of IT servers for managing different services. These include private servers for web, database, mail, file, and proxy management. Moreover, a set of network entities is used to interconnect the different parts of the system. Different types of switches and routers are deployed in a hierarchical structure. These are core and distributed switches in addition to Internet and local routers implemented at the different levels of the system.

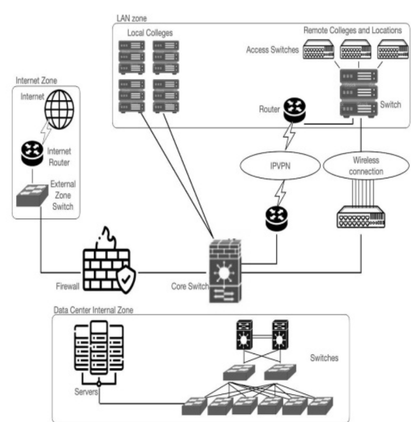


Fig. 1. A Schematic Overview of a Typical Conventional University System.

When the system is integrated with IoT, several IoT-enabling entities need to be introduced to the system. These entities are required to implement the new IoT-related functionalities of the integrated system. For example, IoT sensors and objects are required for enabling IoT data collection and acquisition. Another important functionality is connectivity management and network access control of the IoT devices. This requires special multi-function devices such as IoT gateways with multiple networking capabilities. Furthermore, IoT data in IoT-integrated systems is huge in volume and diversity thus high storage capacities and powerful processing capabilities are highly required. Expanding the current data center with additional IoT servers becomes mandatory. Otherwise, the system needs to be expanded with advanced intelligent computing technologies for effective data management.

It is clear that addressing seamless IoT integration in such systems incurs different challenges. These include the management of heterogeneous IoT devices deployed over vast and maybe remote areas. There is also a need for robust IoT access control considering a variety of IoT communication technologies of distinct properties. Moreover, IoT data management is critical at the different levels of the system. A considerable challenge in this context is addressing big data processing and analytics with a scalable, cost-effective, and flexible solution. Approaching such a challenge using advanced technologies such as cloud computing is the key to efficient and robust IoT data computation and management. Complementing this with the initial processing of the IoT data at the access level would improve resource utilization and performance. However, a trade-off should be maintained between this strategy and the limited capabilities of typical IoT devices at this level of the system. On top of all that, security support is another critical challenging aspect that needs to be fully addressed. Basic security support is insufficient as IoT data need to be secured throughout the different levels of the IoT-integrated system.

IV. Proposed IoT Architecture

The current IoT architectures provide no support for critical advanced aspects as explained in Section II. The importance of these aspects can be seen when considering the complexity and dynamics of the education environments as discussed in the previous section. Taking into account these considerations, a novel IoT architecture is proposed in this section. The following subsections introduce the proposed architecture, present its main components, and discuss potential smart applications.

A. Architecture

The proposed IoT architecture is designed as a multi-layer architecture with the main objective of facilitating the seamless integration of IoT into educational environments. It also aims at optimizing the integrated IoT system with advanced aspects such as complete security support, responsive edge processing, and customized IoT connectivity. The proposed architecture provides a modular, scalable, and simple design that can effectively meet the high volume of IoT data and application demands in the targeted environments.

The proposed IoT architecture consists of six hierarchical layers and one vertical layer. A bidirectional connection flow is established between adjacent hierarchical layers. These are the IoT-object, access, edge-computing, infrastructure, cloud, and application layers that will be discussed in the following sub-section. The focus here is on bridging the gap between IoT and legacy systems in educational environments. It encompasses most of the functionality that is implemented in the legacy systems into the infrastructure layer whereas the other layers facilitate the integration of the distinctive IoT functionalities. The architecture also has a vertical layer for providing complete security support across the different hierarchical layers of the architecture. This layer has a horizontal connection to each other layer to allow securing the system in a comprehensive manner. Fig. 2 shows an architectural representation of the proposed IoT architecture.

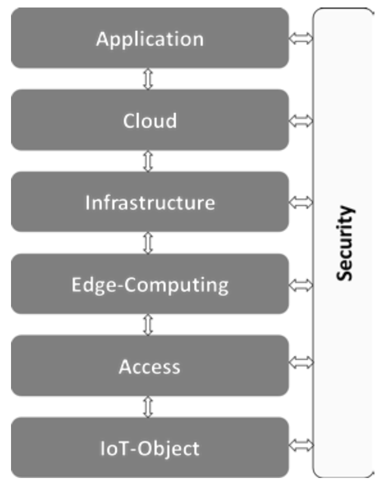


Fig. 2. Architectural Representation of the Proposed IoT Architecture.

The number of layers is selected to ensure effective granularity, modularity, and scalability. That is, each separate and distinctive functionality of the integrated system is embedded in a different layer. This would make IoT integration more simplified and standardized while considering all the essential components of the system. In addition, only the most relevant and critical layers are added to the architecture to maintain a practical trade-off between effective modularity and less complexity. The following part of this subsection describes and discusses each layer of the proposed IoT architecture.

1) IoT-object layer: The bottom layer of the architecture is a physical layer that incorporates all the different IoT objects. These include IoT sensors and actuators in addition to IoT-enabled physical objects. This layer manages IoT data collection and acquisition in addition to controlling requests destined for IoT devices. Furthermore, different aspects including deployment strategy, devices heterogeneity, resource management, and device identification are considered at this level.

2) Access layer: Management of the connectivity of IoT devices with the system is addressed in the access layer. This functionality can be realized using different IoT-based wireless communication technologies. These would include Bluetooth Low Energy (BLE), NB-IoT, 6LowPAN, ZigBee, LoRaWAN, and Sigfox technologies in addition to WiFi and cellular connectivity. This layer controls network access of the IoT devices through different entities such as gateways, aggregators, and access points. It ensures that IoT data collected at the bottom layer is received and delivered to the upper layers for further handling and processing.

3) Edge-computing layer: Data transmitted from local IoT subnets through the access layer all the way to further layers can be pre-processed meanwhile. This would enable acting on IoT data at early stages in almost a decentralized manner. Such functionality is realized using the edge-computing layer by initially processing data at different edge nodes such as gateways and local servers. IoT data can be filtered, aggregated, deformed, compressed, and validated for optimized data provisioning and management. Early mining and analysis of the IoT data can also be performed at this layer.

4) Infrastructure layer: The core computing and communication of the system is handled at the infrastructure layer. It is the heart of the proposed architecture interconnecting the different parts of the system. The design of this layer enables encompassing and abstracting the core functionality of the university legacy system. This is a key design aspect for facilitating effective integration of the different IoT functionalities into the system.

5) Cloud layer: Data management and analysis are realized in this layer with full reliance on cloud technology. IoT data storage and processing are offloaded from the computing infrastructure to the cloud system. The main functionality performed at this layer is effectively managing storage capacities and processing capabilities for IoT data. This requires proper access control to cloud servers and resources. Another consideration in this layer is enabling intelligent IoT data analytics to provide added-value IoT services.

6) Application layer: the application layer defines many application-related aspects such as how to utilize the processed data, manage user access to the system, and request IoT services from the underlying layer. Processed IoT data is used at the application layer to produce application-specific services. These can be implemented for different IoT educational applications such as the smart classroom and smart library applications. It is the top layer of the proposed architecture facilitating direct interaction with IoT application users. Therefore, it is responsible for controlling user access to IoT applications and enhancing user experience.

7) Security layer: Security management is realized by a vertical layer that is logically interfaced with each level of the architecture. The design of this layer allows incorporating the different security mechanisms into the hierarchy of IoT-integrated systems. These mechanisms include data encryption, user authentication, access authorization, threats isolation, attack detection, and trust control.

B. Main Components

To realize the proposed architecture, a set of new IoT elements need to be effectively incorporated into the integrated IoT system. This is presented in Fig. 3 which demonstrates how the legacy university system shown in Fig. 2 can be integrated with IoT. The incorporation of these elements is important to efficiently build up the backbone of the IoT-enhanced university system. Every new element is added to support the functionality of a specific layer. The main elements are presented and discussed in this subsection as follows.

1) IoT devices: IoT objects are physical entities embedded with sensors, firmware, processing units, and electronics to enable sensing, sending, and receiving data. IoT devices constitute the main source of data in IoT systems. The most common IoT devices are sensors, actuators, microcontrollers, smart wearables, and wireless cameras. These can be easily provisioned by the proposed architecture at a different scale. The management of these devices is realized at the IoT-object layer to facilitate software/firmware updates, device capability provisioning, and remote diagnostics.

2) IoT connectivity: IoT connectivity can be realized using a number of different IoT-oriented connectivity options with distinctive features. They vary in coverage range and power consumption while sharing some properties such as low data transmission rate. These can be classified into short and long-range communication protocols. For example, ZigBee and Bluetooth have short communication ranges whereas LoRaWAN and NB-IoT provide support for long-range communications. The proposed architecture is flexible enough to incorporate different communication protocols and support any potential connectivity requirements. It also supports different communication models such as the point-to-point and point-to-multipoint models. This would enable machine-to-machine communications which allow devices to establish connections among themselves without human intervention.

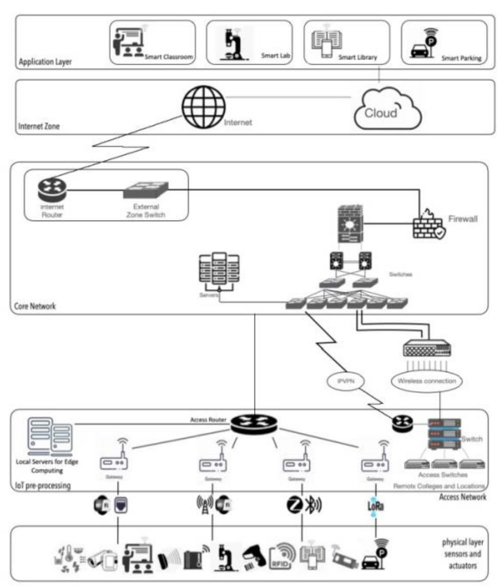


Fig. 3. A Schematic Overview of an IoT-Integrated University System.

3) IoT gateway: Gateways in IoT systems are essential devices to perform different functions. These include IoT data forwarding, protocol conversion, IoT node management, and security support. Gateways are typically deployed at the edge of the IoT network interconnecting IoT nodes to the infrastructure. The network stack of most of the typical IoT devices has no support for IP-based Internet connectivity which can be provided using a gateway. Gateways also provide an additional abstraction to address the heterogeneity and interoperability of IoT devices. In addition, new generation IoT gateways can also provide local caching and pre-processing of IoT data. In the proposed architecture, gateways are essential elements that operate at the access layer to provide all the aforementioned functions. Computing resources of the gateway devices can also be utilized for edge processing at the edge-computing layer of the architecture.

4) IoT edge servers: Edge computing can be realized in IoT systems by deploying edge servers in close proximity to the IoT end nodes. These servers then provide the required computing and storage resources to pre-process IoT data before being transmitted to the cloud. This would help in improving system reliability and availability. Edge servers operate at the edge-computing layer of the proposed architecture with the ability to support a distributed computing model. That is, edge servers deployed across the system can collaborate to provide real-time IoT data pre-processing in a scalable and distributed manner. This is important to improve system responsiveness as the speed and the volume of IoT data is increasing. However, edge servers are usually of limited computing and storage capabilities, unlike the largescale and powerful cloud servers. Therefore, it is critical to have a data management and processing framework to offload computation-intensive data processing to cloud servers.

5) IoT cloud: Cloud computing provides computing resources such as processors, storage, software, and networks as a service. It enables cost-effective data management, on-demand self-service, powerful data computing, and scalability. The proposed architecture enables the development of a cloud-based IoT system to enhance system flexibility and reliability. Relying on the cloud would make IoT integration easier and more effective. It enables unlimited, cost-effective, and on-demand data storage management without any infrastructure limitations and network restrictions. Given that IoT data is typically large and unpredictable, offloading IoT data processing and analytics to the cloud is a feasible strategy to alleviate considerable burdens on the infrastructure.

6) IoT data: IoT data is collected and acquired in different forms and from different IoT objects. These can be status, actionable, location, and automation data in structured and unstructured formats. IoT data is characterized by having a high volume, large-scale streaming, diversity, high dimension, time and space coloration, and high-noise environments. In the proposed architecture, IoT data is handled throughout the different levels of the system. It is collected at the IoT-object layer and then get forwarded at the access layer to the core infrastructure of the system. During that, the data is pre-processed for initial filtration, aggregation, deformation, compression, and validation. Full management and processing of the IoT data are performed at the cloud layer which produces processed application-specific data. Mechanisms and algorithms for data format converting, machine learning, data mining, and reasoning can be performed on IoT data. The basic support that needs to be provisioned for IoT data is storage capacity and processing capabilities.

7) IoT services and interfaces: Once having the IoT-integrated system is established, different IoT services can be provided considering different applications. These services can be developed for different smart activities such as environment monitoring, surveillance, control and automation, face recognition, and productivity support. In addition, IoT users are provided with the required services and interfaces to interact effectively with the system in the context of different IoT applications. All these functionalities are managed at the application layer of the proposed architecture.

C. Potential Smart Applications in Educational Environments

The proposed IoT architecture enables meeting the requirements of different potential smart applications in educational environments. These can include the smart classroom, smart lab, smart library, and smart parking applications. This would facilitate the development of an effective IoT ecosystem that can enhance the teaching-learning process, enrich the educational experience, and improve administrative activities to further limits. This subsection discusses the aforementioned smart applications which can be easily and effectively deployed using the proposed architecture.

All these smart applications require extending the current infrastructure of the university system with additional IoT entities. The proposed architecture supports the effective management of all these entities at the different layers of the architecture. In the smart classroom and lab setups, entities such as environmental and activity sensors, control actuators, and IoT-enabled objects are incorporated. The smart library application would mainly rely on RFIDs technology to tag and manage library resources. For the smart parking application, Infrared and Ultrasonic sensors are used to monitor parking lots. Smart cameras can also be deployed in these applications for smart recognition and AI-based services. For all the different applications, IoT device management is realized at the IoT-object layer.

These IoT devices are mostly main-powered, particularly in the cases of indoor smart applications. Therefore, the connectivity of such devices can be simply maintained using Ethernet and WiFi connections. This would ensure a high data rate and more reliable connections without adhering much to any IoT-oriented concerns regarding energy efficiency. It also makes network access control easier as no additional gateways are required. The proposed architecture also enables implementing IoT devices with other indoor connectivity options such as ZigBee and 6LowPAN if needed. These would be implemented in remote areas where poor or no WiFi coverage exists. In this case, appropriate gateways need to be installed for interconnectivity management with the IP infrastructure. In the case of outdoor smart parking, it is challenging to have accessible Ethernet and WiFi connectivity. Therefore, long-range wireless communication protocols such as LoRaWAN and NB-IoT would be of significant use in these cases. The deployment of a smart parking setup using these protocols requires the provisioning of smart gateways that provides IP interconnectivity. The proposed architecture supports interoperability among all these different communication standards at the access layer.

Massive IoT data streams would be continuously generated in such interactive and dynamic applications. For large-scale deployments in educational environments, the demanding applications of smart classroom and lab would incur a high volume of IoT data. Accordingly, it can highly benefit from edge processing to provide real-time services. Multiple local edge servers can be installed at local LANs to pre-process IoT data streams. Distributed edge processing can also be realized by managing the operations of the widely deployed edge servers using a virtual controller. Such functionality can be practically implemented at the edge-computing layer of the proposed architecture. In the case of the smart parking application, LoRaWAN gateways enabled with the edge computing capability are a feasible solution. It would provide sufficient IoT data handling given the average flow of data in such a predictive application. It is important to note that the proposed architecture supports having both local edge servers and edge-enabled gateways to improve edge computing in the system.

For all the different applications, integrating the IoT ecosystem requires no considerable modifications and updates to the core infrastructure of the legacy university system. The smart classroom application, for example, may only require additional wireless access points to be installed and connected to a local LAN network. Even in the smart parking situation, any installed gateway is interfaced with the core infrastructure through a LAN access router and managed at the access layer of the architecture. In all the different cases, the core infrastructure would only experience an increase in the data rate and may require provisioning more bandwidth for QoS assurance. It though needs no further resource provisioning for data processing and management as the cloud layer is responsible for such functionality.

V. Use Case

The example presented in Fig. 4 shows an overview of the architectural structure and implementation components of a smart packing application in an educational environment. It demonstrates how the core computing infrastructure is seamlessly integrated with the IoT resources and functionality. The proposed architecture enables a seamless deployment strategy. It starts by connecting the core domain to the access and edge computing domains on one side while being interfaced with the cloud domain on the other side.

Only gateways with average computing resources are installed for each parking block to manage network access of smart parking sensors. LoRaWAN is the feasible connectivity option in this case considering that no easy access to local LANs would be available in outdoor areas. LoRaWAN-enabled sensors and gateways are deployed as needed to have full coverage. On the other hand, smart gateways can be used to provide sufficient resources for pre-processing IoT data traffic streams. That is, streamed data in this case is typically received at average volume from the smart parking sensors. Only very basic sensor data indicating basic information such as the occupancy state of a parking lot is streamed in a textual format. Data can be filtered and aggregated while being timestamped at the edge-computing layer before being forwarded further to the cloud. However, car plate recognition can also be implemented to receive images from smart cameras at the entrance of a parking zone. For such a demanding AI-based service, intensive processing is performed at the cloud servers. At the top layer, smart parking services are provided using interactive user interfaces which are connected to back-end cloud servers.

Cross-layer implementation of full security support is achieved using different security mechanisms. These include but are not limited to authentication, authorization, encryption, and trust management. However, the system is scalable enough to implement additional security solutions at the different levels of the architecture. It is evident that the proposed architecture succeeds in guiding system construction and enhancing the adoption of the smart parking application. Seamless and flexible integration of the introduced IoT resources and functionality was achieved effectively in a plug-and-play fashion.

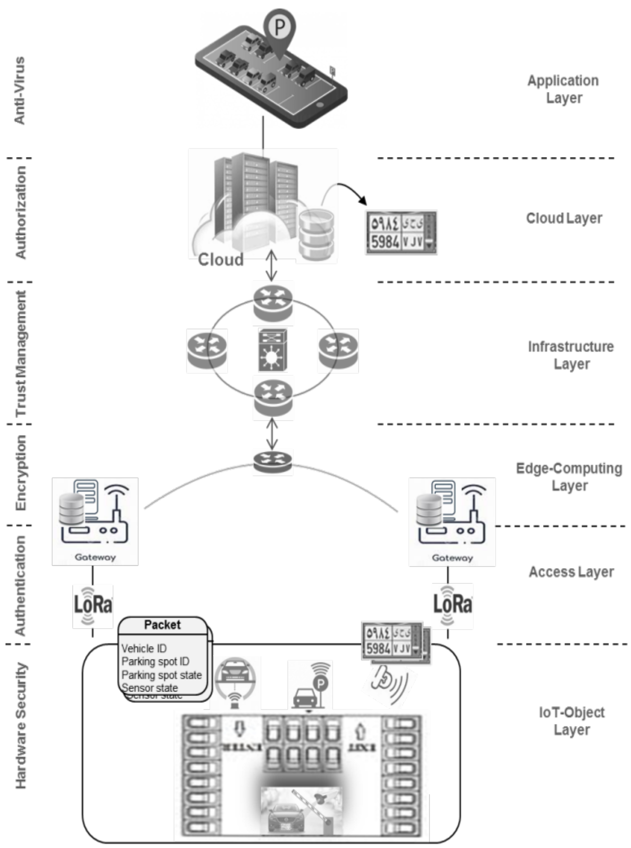


Fig. 4. Overview of a Simple use Case of a Typical Smart Parking Deployment.

VI. Discussion

It can be seen that one of the distinctive properties of the proposed architecture is the split of the networking functionality into two different layers: the access and infrastructure layers. Given the key role of networking support in any IoT-based system, this strategy is important to simplify the integration of IoT communications into the core networking infrastructure of conventional systems. IoT comes with different network access models that are managed at a separate level of the system. Isolating access control would enable effective connectivity among IoT devices and seamless transmission of IoT data over to the core network system.

In addition, the proposed architecture realizes the importance of edge computing for optimizing the responsiveness of the IoT-based system. By placing data pre-processing closer to the physical layer, it ensures that unacceptable latency is considerably alleviated irrespective of the expected high volume and dynamics of IoT data. Having IoT data handled closer to IoT end devices in the system would also improve data management and quality. Real-time processing becomes easy to implement for providing a variety of time-sensitive IoT services. Moreover, such a strategy helps in realizing more advanced support regarding different networking aspects such as user mobility and context-aware smart services.

Another important feature of the proposed architecture is encapsulating most of the functionality of the conventional systems into the infrastructure layer. This also includes all the networking entities and computing resources that constitute the core of these systems. Such a design approach would abstract the core functionality of the existing systems from the IoT integration process and avoid getting into the system complexity. This would be further enhanced with the reliance on cloud computing to realize effective IoT data management and processing. Accordingly, the proposed architecture emphasizes the importance of incorporating the cloud layer to achieve better reliability and less complexity. Furthermore, having a security layer that is interconnected to the whole system ensures the ability to provide complete security support. This design principle of the proposed architecture enables providing the different security services required at each layer.

Table I provides a comparison of the proposed IoT architecture against a set of different architectures. The inherent properties of the proposed architecture can meet different requirements that are critical to university systems. Compared with the other architectures, it provides a salable and modular design that allows elastic expansion of resources and entities to meet different IoT application requirements. It can also grow hierarchically into a different model incorporating other technologies. For example, the architecture can be adapted to accommodate the Blockchain technology in an additional sub-layer or instead of the cloud layer to realize a more decentralized data storage and processing. Encapsulating and abstracting the core computing and communication of the IoT-transformed system into the infrastructure layer makes the architecture more flexible for adaptation.

TABLE I. IoT Architectures Comparison

Architecture	**Scalability**	**Data Management**	**Security Support**	**Responsiveness**	**Wide IoT Coverage**	**Complexity**
[7]	L	L	L	L	L	L
[9]	L	M	L	L	L	L
[10]	M	M	L	L	L	M
[12]	M	M	M	L	L	M
The Proposed Architecture	H	H	H	H	H	M

Compared with other basic architectures, the complexity of the proposed architecture is maintained at an acceptable level given the provisioned set of functionalities. Every layer of the architecture provides essential support to facilitate IoT integration and enhance the IoT system. Having a vertical layer for full security provision would add to its complexity but at the benefit of providing significant support of security at each level of the architecture.

VII. Conclusion

Seamless integration of IoT in legacy university systems is still a considerable challenge. The proposed architecture in this paper efficiently addresses such a challenge to facilitate IoT integration and accelerate IoT adoption. It comes with a modular and scalable design allowing the effective abstraction of the legacy infrastructure of the university system. It is based on a seven-layer model that incorporates advanced technological considerations including combined edge-cloud computing in addition to effective IoT accessibility. In addition, the architecture is optimized toward complete security support using a vertical layer covering the whole system. It also provides responsive edge processing to support real-time communication with low system latency. The architecture also supports customized IoT connectivity and simplifies the deployment of heterogeneous IoT communication technologies. In comparison with other IoT architectures, the proposed one ensures high reliability, data management, security support, responsiveness, and wide coverage while maintaining acceptable complexity. The focus of the future work will be on studying how the proposed architecture can be extended to other specific IoT use cases. Another aspect that will be investigated is incorporating other technological advances such as Blockchain technology.

References

[1] M. Selinger, A. Sepulveda and J. Buchan, "Education and the Internet of Everything", Cisco Consulting Services EMEAR Educational Team, October 2013, [online] Available: http://www.cisco.com/c/dam/en_us/solutions/industries/docs/education/education_internet.pdf.

[2] S. Selvaraj and S. Sundaravaradhan, “Challenges and opportunities in IoT healthcare systems: a systematic review,” SN Applied Sciences, vol. 2, no. 1, pp. 1-8, 2020.

[3] B. Chandrayan and R. Kumar, “IoT integration in industry—a literature review,” Recent Advances in Mechanical Engineering, pp. 9-17, 2020.

[4] J. Ruan, J. Hua, Z. Chunsheng, H. Xiangpei, S. Yan, L. Tianjun, R. Weizhen, and C. Felix, “Agriculture IoT: Emerging Trends, Cooperation Networks, and Outlook,” IEEE Wireless Communications, vol. 26, no. 6, pp. 56-63, 2019.

[5] M. Al-Emran, S. I. Malik and M. N. Al-Kabi, "A survey of Internet of Things (IoT) in education: Opportunities and challenges" in Toward Social Internet Things (SIoT): Enabling Technologies Architectures and Applications, Cham, Switzerland:Springer, pp. 197-209, 2020.

[6] B. Dhanalaxmi and G. A. Naidu, "A survey on design and analysis of robust IoT architecture," in International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), 2017, pp. 375-378.

[7] I. Mashal, O. Alsaryrah, T.Y. Chung, C.Z. Yang, W.H. Kuo, and D.P. Agrawal, “Choices for interaction with things on Internet and underlying issues,” Ad Hoc Networks, vol. 28, pp. 68-90, May 2015.

[8] O. Said and M. Masud, “Towards internet of things: Survey and future vision,” International Journal of Computer Networks, vol. 5, no. 1, pp.1-17, 2013.

[9] M. Wu, T.J. Lu, F.Y. Ling, J. Sun, and H.Y. Du, “Research on the architecture of Internet of Things," in the 3rd International Conference on Advanced Computer Theory and Engineering(ICACTE), 2010, pp. V5-484-V5-487.

[10] M. Burhan, R. Rehman, B. Khan, and B.-S. Kim, “IoT Elements, Layered Architectures and Security Issues: A Comprehensive Survey,” Sensors, vol. 18, no. 9, p. 2796, Aug. 2018.

[11] N. V. Lopes, F. Pinto, P. Furtado and J. Silva, “IoT architecture proposal for disabled people,” in IEEE 10th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), 2014, pp. 152-158.

[12] P. Sethi and S. R. Sarangi, “Internet of Things: Architectures, Protocols, and Applications,” Journal of Electrical and Computer Engineering, vol. 2017, Article ID 9324035, 2017.

[13] K. K. Karmakar, V. Varadharajan, S. Nepal and U. Tupakula, "SDN-Enabled Secure IoT Architecture," IEEE Internet of Things Journal, vol. 8, no. 8, pp. 6549-6564, April 2021.

[14] R. T. Tiburski et al., "Lightweight Security Architecture Based on Embedded Virtualization and Trust Mechanisms for IoT Edge Devices," in IEEE Communications Magazine, vol. 57, no. 2, pp. 67-73, February 2019.

[15] N. Lo and I. Niang, “A Comparison of QoS-Based Architecture Solutions for IoT/Edge Computing Environment,” Emerging Trends in ICT for Sustainable Development, pp. 355-364, 2021.

[16] S. Maurya and K. Mukherjee, “An energy efficient architecture of IoT based on service oriented architecture (SOA),” Informatica, vol. 43, no. 1, pp. 87–93, 2019.

[17] H. Rahimi, A. Zibaeenejad and A. A. Safavi, “A Novel IoT Architecture based on 5G-IoT and Next Generation Technologies,” in IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2018, pp. 81-88.

[18] S. Verma, R. Gala, S. Madhavan, S. Burkule, S. Chauhan and C. Prakash, “An Internet of Things (IoT) Architecture for Smart Agriculture,” in 4^{th} International Conference on Computing Communication Control and Automation (ICCUBEA), 2018, pp. 1-4.

[19] M. Nitti, V. Pilloni, D. Giusto, and V. Popescu, “IoT Architecture for a Sustainable Tourism Application in a Smart City Environment,” Mobile Information Systems, vol. 2017, Article ID 9201640, 2017.

[20] J. Lloret, J. Tomas, A. Canovas and L. Parra, “An Integrated IoT Architecture for Smart Metering,” IEEE Communications Magazine, vol. 54, no. 12, pp. 50-57, December 2016.

[21] O. Debauche, S. Mahmoudi, P. Manneback, A. Assila, “Fog IoT for Health: A new Architecture for Patients and Elderly Monitoring,” Procedia Computer Science, vol 160, pp. 289-297, 2019.

[22] D. Sembroiz, S. Ricciardi, D. Careglio, “A Novel Cloud-Based IoT Architecture for Smart Building Automation,” Security and Resilience in Intelligent Data-Centric Systems and Communication Networks, Eds.; Academic Press: Cambridge, MA, USA, 2018; pp. 215–233.

[23] M. Muñoz, J. Gil, L. Roca, F. Rodríguez, and M. Berenguel, “An IoT Architecture for Water Resource Management in Agroindustrial Environments: A Case Study in Almería (Spain),” Sensors, vol. 20, no. 3, p. 596, Jan. 2020.

Software Defined Network based Load Balancing for Network Performance Evaluation

Omran M. A. Alssaheli, Z. Zainal Abidin, N. A. Zakaria, Z. Abal Abas
Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Malaysia Melaka, 76100, Melaka, Malaysia

***Abstract*—Load balancing distributes incoming network traffic across multiple controllers that improve the availability of the internet for users. The load balancing is responsible to maintain the internet availability to users in 24 hours by 7 days a week. However, the internet become unavailable since the load balancer is inflexibility, costly, and non-programmable for settings adjustment especially in managing the network traffic congestion. An increasing user using mobile devices and cloud facilities, the current load balancer has limitations and demands for the deployment of a Software-Defined Network (SDN). SDN decouples network control, applications, network services, and forwarding roles; hence makes the network more flexible, affordable, and programmable. Furthermore, it has been found that SDN load balancing performs intelligent action, efficient and maintains better QoS (Quality of Service) performance. This study proposes the application of SDN-based Load Balancing since it provides pre-defined servers in the server-farm that receive the arrived Internet Protocol (IP) data packet from various clients in the same number of loads and process orders for each server. Experiments have been conducted using Mininet™ and based on several scenarios (Scenario A, Scenario B, and Scenario C) of network topologies. Parameters used to evaluate the load balancing in SDN are throughput, delay, and jitter. Findings indicated that scenario A gives a high throughput, scenario B and C produce a low jitter values and scenario C produces the lowest delay. The impact of SDN brings a multi-path adaptive direction in finding the best route for a better network performance.**

***Keywords*—Load balancing; software-defined network (SDN); SDN load balancing; network performance; Mininet**

I. Introduction

The network started in the late 1990s with electronic messaging (e-mail), followed by file transfer protocol (FTP) and more network services are used such as access multimedia files (audio or video) or content distribution in TCP/IP architecture [1]. An increasing usage of video streaming of the network user has shown in recent trends. There will be 5.3 billion total users by 2023 [2]. Moreover, the enormous growth of users is due to the use of applications in cloud services, such as Internet of Things (IoT) and Data Science. These applications and network services demand a huge volume of IP traffic transmissions in a high-speed data communication infrastructure.

In the legacy network, load balancing segregates inbound network packets that coming into the network and outbound the packets across the network through multiple controllers. In a simple definition, load balancing is to ensure availability of network services and applications to users for every 24 hours in 7 days. As an analogy, load balancing performed as a traffic policemen standing in the middle of junctions and giving directions for vehicles to take turns to maneuver to avoid traffic jams or accidents.

Moreover, load balancing is responsible to balance an enormous volume of IP traffic across two or more Wide Area Network (WAN) that links without using complex routing protocols such as Border Group Protocol (BGP). The application of the balancing service produces an equilibrium network session over multiple connections in order to spread out the amount of bandwidth used by each Local Area Network (LAN) users, for example, browsing websites and accessing email.

Video streaming and gigantic data transmission from various users to another, demand another mechanism to be integrated with the load balancing in solving problems of delay, packet loss and high bandwidth utilization. Therefore, the objective of this study is to analyze the new mechanism in the load balancing and evaluate the network performance (jitter, delay and throughput) based on scenarios (A, B and C).

The following Section II describes the literature review of load balancing. Section III explains on materials and method used to perform the experiment environment of SDN using scenarios to represent the network topology. Section IV explains about results and findings of the study. Section V is the discussion that presents the challenges of doing the research. Section VI summaries the SDN load balancing and network performance evaluation based on scenarios.

II. Literature Review

LAN and WAN needs an implementation of a load balancing mechanism to competently distribute arriving packet traffic over a collection of server farm. The importance of the load balancing located at servers and routing clients requires a maximum speed, and capacity utilization across servers in order to guarantees that no server is strained that reduce the efficiency [3] of network performance.

Therefore, load balancing manages the increasing number of IP traffic in the internet for a better network performance. In WAN environment, when one of servers is failed, the load balancing device reroutes traffic [4] to available server in the farm. As the additional server is included to the farm, the load balancing instantly sends requisition to the server group for acknowledgment and provides the alternative route for the traffic packet to travel.

Nevertheless, a major issue in the load balancing is the difficulty in managing large network using only a single server [5]. A single server in the topology design produces a bottleneck [6] condition as soon as all data packets queueing at the interface in the same time to get to the destination.

In fact, the load balancing is made of hardware-based-networks controller and requires system maintenance in managing huge IP traffic. On one hand, the manufacturer has hard coded the programmable controls that is easier to install and settings at the load balancer device. On the other hand, the load balancer performs a complicated route [7]–[9], difficult maintenance process, and is time-consuming.

Another issue related to the load balancing is the existing TCP/IP construction framework was not planned to fulfil the need of the large scale of video content distribution [10] and a high number of Internet users. To change the TCP/IP architecture involve enormous amount of expenditure. For a quick solution, a cost-effective approach to overcome the problem of TCP/IP stack is implemented using the software-defined networking (SDN) [5] since it creates another visualization layer for content processing and content distribution.

Load balancing networks are non-programmable [11] since the network administrator unable to perform new settings and the configuration is set by default that is session based load balancing. Fig. 1 shows the load balancing that uses the hardware-based-controller at the control plane for each data plane.

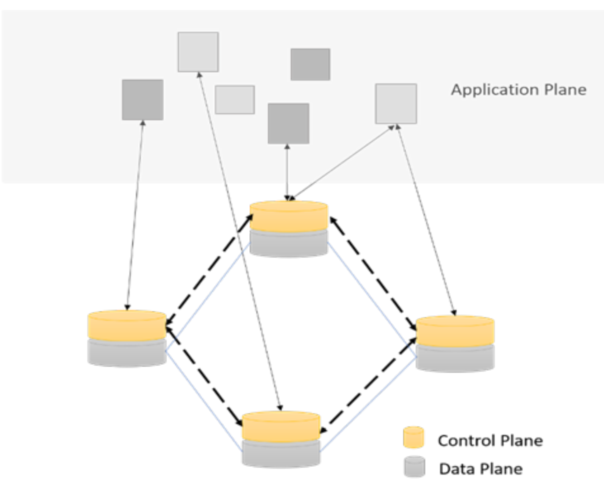


Fig. 1. Load Balancing in Network Architecture.

The configuration and setup are fixed and programmed according to functionalities and services to be provided. The control plane is situated at the data plane that increases the cost of implementation and installation of the networking infrastructure. The load balancing uses several techniques such as round robin, the least connection approach and software-defined networking load balancing.

The round robin approach is capable to forward total traffic packet by responding to DNS requests with a list of IP addresses from the nodes through poor connections and low bandwidth links. Round robin algorithm uses heterogeneous servers [12] with different link of quality, measure the load condition, and variation of security restrictions based on several scenarios of network environment. For example, a set of identical servers is allocated to provide the same services such as centralized management and control; scalable, reliable and monitoring performance. Although each server has its own IP address, it is set up to utilize the same domain name. The DNS server keeps track of all IP addresses linked with Internet domain names. When a request for an Internet domain name and its related IP address is received, the entire addresses are delivered and return in a rotating order. This architecture considers only the distribution of the incoming traffic without considering the server's side [13]. Nonetheless, round robin algorithm predicts that every server has the same capability and resource specifications such as CPU and RAM, to handle equivalent loads.

The least connection approach brings the recent server load into account. The invitation is sent to the server that has served with the lowest number of connections. Each server is assigned to a unique number. If the number of alive connections on two servers is the same, the higher weighted server gets a new request. Every server in a pool is assigned an agent, which announced to the load balancer on its latest load. This real-time data is used to determine the server that should be used to best handle requests [14], which avoid overloading a server through numbers of server connections. However, in measuring the present acquaintances, the server capacity cannot be examined.

A SDN load balancing approach produces an efficient and has a higher speed [15] of network performance. SDN load balancing occupies an important position to solve over-load traffic problem in the network. In fact, the SDN itself is an evolving field in the networking systems and is highly in demand [2]. Google, Facebook, Yahoo, and Microsoft are adopting SDN through open standards development. A SDN produces a flexible, scalable, cost-effective and adaptive features that is ideal for high-bandwidth and complex application in the content distribution. The incorporation of a few low-level features of the network application instead of hardware implementation helps the network administrators to be more effective in managing multiple servers and complex networks.

In addition, SDN allows data to be obtained based on content rather than relying on the hostname and IP address of the device. This is because video or multimedia content is allocated in the cloud platform provider and less concern on the host identification or IP address of the cloud server location. With the combination element from SDN helps load balancing to be able to access data at anywhere and anytime.

SDN is deployed in various networks [16] such as organization and campus networks, data centers and Internet Exchange Points. Moreover, SDN architecture integrates the network control and forwarding function that allow a dynamic and programmable configuration in a cloud-based network monitoring for the new generation of network management. To improve the network performance, SDN offers a simulation platform for a better performance compared to the legacy network management.

The SDN load balancing is more lightweight [17] in terms of CPU utilization and time intervals. SDN load balancing provides features of stability, reliability and scalability [18] during data transmitting and receiving from one server to another. The advantage of SDN load balancing reduces the congestion but increase the speed in data transmission.

In SDN, load balancing performs as an “aware-routing” protocol, which is an essential element that aids availability and scalability, resulting in the shortest possible application response time. Millions of individuals are linked to the internet, resulting in increased web traffic, network congestion, and packet losses, thus, the use of load balancing strategies improves network efficiency. Moreover, the SDN load balancing brings benefits in terms of:

A. *Enhance end-to-end Network Quality-of-service (QoS)*

SDN load balancing enhances the overall networking system's efficiency and QoS [19],[20]. In terms of latency, reaction time, and network performance, QoS provides a better user experience.

B. *Optimize Resource Utilization*

Resource usage is critical, and it must be optimized for maximum efficiency. Bandwidth, processor, connection, and memory utilization [21] are all network resources that must be exploited.

C. *Decrease Transmission Latency*

The term "transmission latency" [22] states time taken for a switch to send information. Latency is affected by numerous aspects of switch performance, including congestion and data packet size. The switch load state is represented by congestion in the link, and SDN accumulates the data packets transferred within a transmission rate and session. Latency is a network performance characteristic that must be less than 100ms in order to maintain a good data delivery.

D. *Minimize Response Time*

Response time means time intermission of a server demands and transmit information are achieved [23]. Thus, the load balancing algorithm uses a distributed SDN network to minimize the response time.

E. *Avoiding Bottlenecks*

Network congestion creates bottleneck at the load balancer [24]. To avoid bottleneck, the SDN load balancing is configured to avoid the switch or controller to get overloaded. Configuration options that are optimized reduce resource use while increasing efficiency, scalability, and response time. The network performance is more effective, there is failover prevention and reduce bottlenecks.

F. *Maximize the throughput*

The SDN load balancing maximize throughput [25] during data transmission. A high throughput is vital for a good network performance since how much data could be delivered within a conversation, which traffic packets are distributed evenly to many nodes and in various types of platforms. The size of traffic packets is delivered in the same kilobytes over a period of time and from one node to another.

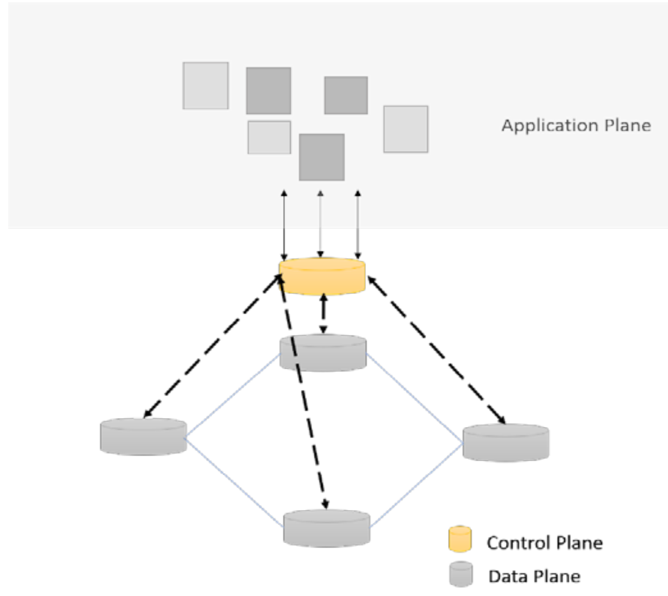


Fig. 2. SDN Load Balancing in Network Architecture.

Fig. 2 illustrates the SDN load balancing in the network architecture. The use of controller has been minimized and cost of installation is reduced. The data plane and the control plane are separated at a different location and space. The purpose is to allow users who use a huge multimedia data to access information from various locations regardless the operating system platform and the network architecture in a small expenditure.

The performance of SDN load balancing is based on parameters, which are throughput, delay and jitter.

Throughput provides information on the successful data packet travelling from one start point to the end point in a provided period. Packet arrival at the destination node in a fast speed is the key indicator for the best service performance in a network. Using throughput, the problem of packet loss is investigated to find the root cause of the problem. Load balancing maximizes throughput by decreasing response time intervals and decreasing data traffic jams.

Based on Equation 1, Mathematical calculation for the throughput can be calculated:

$$\text{Throughput, } \tilde{y}(N) = c\left(1 - \frac{1}{1+\left(\frac{1+\beta}{1-\beta}\right)N}\right) \tag{1}$$

Delay is the time taken for a message or data packets to travel from the source of destination, arrive at the destination node and get back to the source of destination. This is called as the round-trip time (RTT) of the network. Based on Equation 2, Delay is calculated:

$$Pr(E = x)$$
$$\begin{Bmatrix} \sum_{i=0}^{\infty} f_i\,(a), \sum_{i=0}^{\infty} f_i\,(b), \quad x = 0 \\ \sum_{i=0}^{\infty} f_i\,(a), f_{2x+i}\,(b) + \sum_{i=0}^{\infty} f_i\,(b), f_{2x+i}\,(a), x > 0. \end{Bmatrix} \tag{2}$$

Jitter is defined as a variation in the delay of data packets flowing through the network from one node to another node. The jitter situation occurred when data packets do not reach the destination node in the same arrangement as they were sent. The inconsistency of delay values during data transmission determines the quality of network services provided to Internet users. An outcome of the jitter is a higher a jitter value, a higher change of the delay and packet loss happened in the network. Based on Equation 3. Jitter is calculated:

$$\text{Jitter, } \sigma_c = lim_{n=\infty} \sqrt{\frac{1}{N} \sum_{n=1}^{N} (\Delta T_m)^2} \quad (3)$$

Thus, in this paper, the implementation of SDN load balancing with multiple servers is proposed for load balancing since it helps to prevent the bottleneck problem and reduces the data packet congestion in the network.

III. Materials and Method

The method of network configuration in this study is implemented according to simulation-based-experiments using Mininet software [26] in a simple LAN [27] environment. Multiple servers are configured since a single server creates congestion in the network.

The performance measurement and monitoring response are analyzed through different scenarios of load balancing using network controller software tool. There are 3 scenarios implemented, which are scenario A, scenario B, and scenario C. Every scenario consists of SDN Controller, switches, and numbers of clients' connection, which is important for the evaluation. The servers act as a server pool connected to the SDN switch controller and as a data packet reaches at the SDN switch controller, the next selected server appears in the list of all servers on the network system. As a result, every server in the database handles orders with the same number of loads.

A. Scenario A

In this scenario, the topology is based on a simple design, which consists of four clients and two servers. The controller was created using an OpenFlow controller. Fig. 3 illustrates the topology consists of a switch controller, two servers and four hosts.

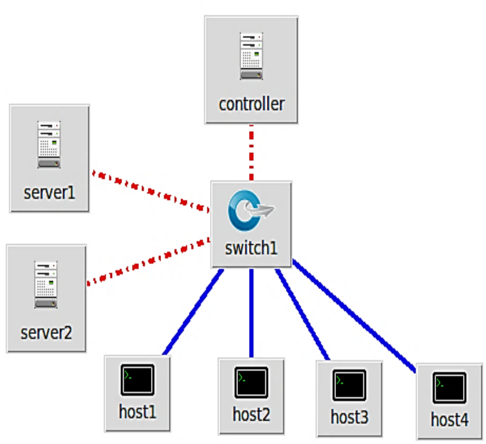


Fig. 3. Topology of Scenario A.

The load is utilizing all of the CPU assets on both web servers. Assume the load on those servers gets to be overwhelming and the Internet location execution diminishes drastically, fair basically includes a third server to the cluster giving extra assets. The arrange activity would at that point be disseminated over three servers as restricted to two.

B. Scenario B

The second scenario consists of four clients, and there are four server pools connected to the switch controller, which is the OpenFlow controller. Each host is located at a dedicated server, which helps to increase the outputs performance. The network performance is measured and monitored based on the delay, throughput and jitter values. Expected outcome of this experiment is to produce a lower delay, better throughput and lower jitter. Fig. 4 illustrates the scenario B experiment setup.

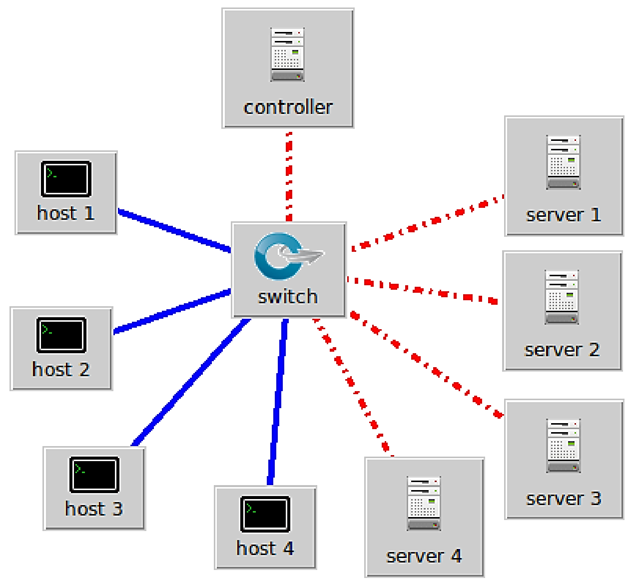


Fig. 4. Topology of Scenario B.

C. Scenario C

The third topology is scenario C consists of one switch controller connected to the four server pools, which support eight clients. The design is similar to the previous two topologies, but there is an additional number of clients. The load balancers lag in these design topologies where the response time and latency increase the load balancers. The SDN controller and the servers remain the same as in Scenario B, and the number of hosts is doubled into 8 hosts. Fig. 5 illustrates Network C.

For all scenarios, the OpenFlow controller is configured using the IP address of 10.0.1.1 and connected to the switch through servers with IP addresses of 10.0.0.1 and 10.0.0.2. The port number of 6633 is used for listening the network packet. The data is captured and tabulated as shown in Table I, Table II and Table III at Section IV.

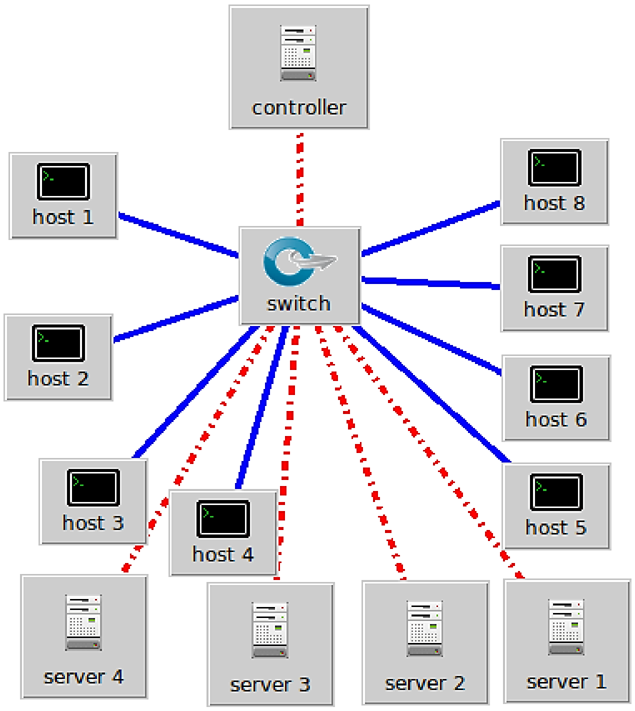


Fig. 5. Topology of Scenario C.

IV. Results and Findings

The network performance is analyzed according to metrics, namely throughput, delay and jitter for every scenario. Comparison is done using simulation and outcomes are evaluated in scenarios based on the number of clients and servers. The result shows that the SDN load balancing is measured in traffic packet of Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) traffic.

A. Throughput

SDN load balancing used a pox controller load balancing connected to number of hosts. Table I illustrates the result of throughput based on scenarios obtained from the experiment conducted.

TABLE I. Throughput for Every Second

Number of Request	**Scenario**		
	A	***B***	***C***
10	41,826	2016	1253
20	16,349	1324	4302
30	17,578	2255	5649
40	33,072	1901	4339
50	52,576	2235	6355
60	55,906	4857	4544
70	72,538	3991	4348
80	66,886	4768	5580
90	69,595	2990	6082
100	76,580	3839	5591

Based on the number of traffic packet requested and transmitted by the server, Scenario A shows larger amount of throughput as the number of traffic request increased. Meanwhile, the Scenario B and Scenario C illustrates an almost same range of values of throughput. Fig. 6 illustrates the throughput values in a form of graph for Scenario A, Scenario B and Scenario C for a better visualization and comparison.

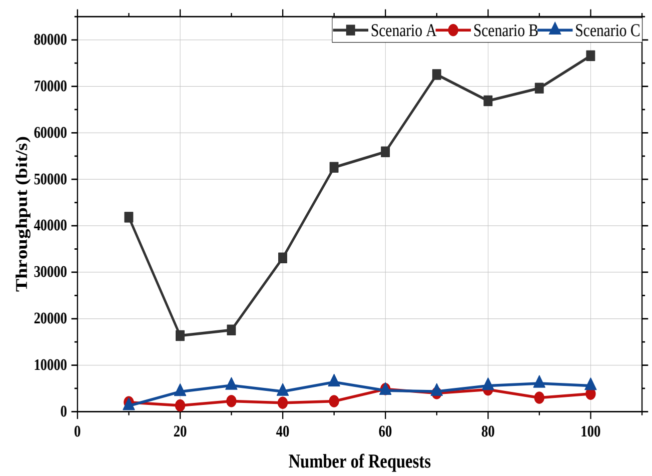


Fig. 6. Throughput Performance based on Scenarios.

Scenario A has given the highest throughput among other scenarios. Even though Scenario A only has two servers but with the SDN based switch controller helps the data transmission to be in effectively and efficiently delivered. Scenario B and Scenario C demonstrate a low throughput value. This is because data packets lost during transmission will cause poor or slow network performance, while poor performance indicates problems such as packet loss. Using performance to measure network speed is beneficial for troubleshooting because it can root out the exact cause of network slowdowns and alert administrators to packet loss-related issues.

B. Delay

The purpose of the SDN load balancing method is to reduce network lag and improve link load balancing by optimizing route calculation and multipath scheduling. Table II provides the delay values for Scenario A, Scenario B, and Scenario C.

Based on Table II, Scenario B indicates a highest delay even though a dedicated server has been provided for each node. Another finding shows that Scenario C shows a higher delay than Scenario A but changed after the number of requests increased at value of 20. After a peak at number of requests of 30, Scenario A keeps showing a constant delay. Fig. 7 presents the delay in a form of representative graph.

Based on results from Table II, Scenario C shows a continuous delay in time taken to transmit the packet. Scenario B shows imbalance condition of delay as the network is monitored to be slow in performance even though a dedicated server is provided to each node, which the concept of single server is not applicable using SDN controller. As a result, Scenario C performs a slow in speed of data transmission in the network performance, Scenario B consists of the highest delay due to type of data packets travels depending on the application that the user used. Scenario A shows unstable condition at the beginning but then the delay is increasing as the number of packets are requested by clients.

TABLE II. DELAY FOR EVERY SECOND

Number of Request	Scenario		
	A	*B*	*C*
10	10,160	21,383	31,273
20	12,738	57,435	18,223
30	71,322	51,088	20,816
40	49,251	83,701	35,595
50	41,975	88,362	30,841
60	44,300	47,294	52,105
70	41,278	69,933	58,059
80	51,581	69,052	57,458
90	56,749	125,331	58,777
100	60,214	104,335	72,069

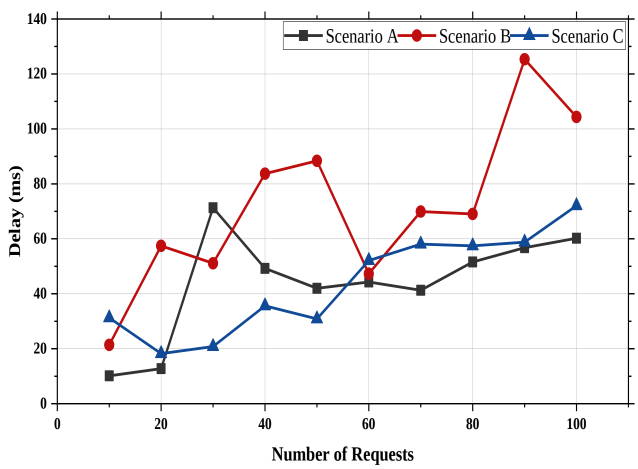


Fig. 7. Delay Performance based on Scenarios.

C. Jitter

Using SDN load balancing, jitter is measured to find the variation of delay in arriving packets to destinations. The variation of jitter represents the pattern or behavior when there is a route changed or congestion.

Table III shows the value of jitter obtained from the Scenario A, Scenario B and Scenario C. Based on the jitter values, Scenario B and Scenario C produce almost the same pattern of jitter since both scenarios using four servers with different number of clients. Finding shows that the jitter produces the similar variation of delay of the server that is designed with or without a dedicated server to a particular client.

Fig. 8 shows Scenario A has given an inconsistency condition compared to Scenario B and Scenario C. Scenario A demonstrates that the topology A built a large data packet queue that causes a huge delay and bursts of jitter. Finding found that Scenario A has a higher jitter than Scenario B and Scenario C because it processes the packet request more to transmit to the destination. Thus, it disrupts in the idle condition of transmitting data packets. The jitter is represented in a variance in time delay in milliseconds for data packets over a network.

Scenario B and Scenario C gives a reliable and scalable network performance. The jitter condition is noticeable as the graph pattern shows an almost constant variation form. For example, a user using Cisco WebEx or Microsoft Teams for online meeting has been several times disconnected. Not only that, user cannot hear the voice of another user or the voice quality has been distorted. Therefore, the longer the data packet arrives, the greater the negative impact of jitter on video and audio quality.

TABLE III. JITTER FOR EVERY SECOND

Number of Request	Scenario		
	A	*B*	*C*
10	1.127	0.998	1.000
20	2.237	0.986	1.006
30	0.001	1.008	1.008
40	0.007	1.013	1.002
50	5.481	1.001	1.016
60	6.852	1.021	1.001
70	0.032	1.004	1.034
80	4.440	1.019	1.009
90	2.307	1.011	1.017
100	1.006	1.007	1.029

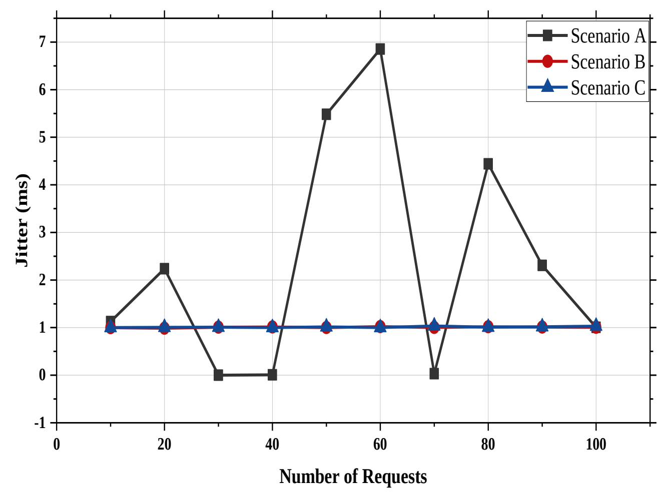


Fig. 8. Jitter Performance based on Scenarios.

V. Discussion

The outcome of this study showed that the application of SDN controller in load balancing brings a better scheduling mechanism, regardless of the number of clients requesting the packet to reach the destination. In addition, the number of servers deployed does not directly affect the overall network performance, especially in the case of network congestion and delays. Metrics are used to measure the SDN controller based load balancing in the network are throughput, delay and jitter.

According to the findings of this study, a SDN load balancer requires evidence regarding on the network's throughput. A start node delivers traffic to an end node to measure traffic on the path, and each packet is time-stamped. The packets are routed across the network's switches without bias decision, and the receiver determines the entire end-to-end delay of a packet based on the time-stamp provided. Similarly, to path measurement, each SDN switch stamps each packet with a time-stamp, which the receiver uses to calculate the transmission delay for each link in the path.

Another significant finding is that the excessive delay is caused by the SDN switch's relatively poor CPU throughput. A delay of 30 milliseconds or more, on the other hand, produce network distortion. The jitter must be less than 30 milliseconds for the video transmission to perform properly. Higher receiving jitter slows the network performance; produce a packet loss and audio quality concerns. The SDN controller behaves as expected, for instance, the response time tends to reduce as we add additional servers. When we compare a server farm with two servers to one with three servers, the response time is cut in half, but the result remains nearly constant as the number of servers increases.

VI. Conclusion

Load balancing is crucial for ensuring the availability of network to users. Nonetheless, a high demand of Internet services from users creates an adoption of load balancing. The state-of-the-art indicates that SDN load balancing is highly recommended due to features such as lightweight, stability, reliability and scalability. SDN load balancing increases user interaction by enabling the administrator to monitor the condition of the servers, check the status of the balancer through log files, and set to disable mode for the balancing features at the specific faulty servers.

The SDN controller has been analyzed in the load balancing method for network performance evaluation, which is based on three simulated scenarios. Scenarios A, B, and C each used a number of servers and multiple clients connected to the servers to produce a different outcome, which is based on the delay, jitter and throughput. The network administrator is able to modify the SDN controller settings in a less expensive and more user-friendly way for a better network performance.

Further investigation on algorithms based on the field of artificial intelligence are needed to reduce the complexity of implementing SDN load balancing in enabling autonomous scheduling and intelligent routing functions.

Acknowledgment

Thank you Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Malaysia Melaka.

References

[1] T. Guesmi, A. Kalghoum, B.M. Alshammari, H. Alsaif, A. Alzamil, “Leveraging Software-Defined Networking Approach for Future Information-Centric Networking Enhancement,” Symmetry, vol. 13, no. 441, pp. 1-19, 2021.

[2] “Cisco Annual Internet Report (2018–2023) White Paper,” 2020. [Online]. Available: www.cisco.com. [Accessed September 2, 2021].

[3] G. Singh and K. Kaur, “An improved weighted least connection scheduling algorithm for load balancing in web cluster systems,” International. Research. Journal. Engineering. Technology, vol. 5, pp. 1950–1955, 2018.

[4] C.-T. Yang, S.-T. Chen, J.-C. Liu, Y.-W. Su, D. Puthal, and R. Ranjan, “A predictive load balancing technique for software defined networked cloud services,” Computing, vol. 101, no. 3, pp. 211–235, 2019.

[5] T. Semong, T. Maupong, S. Anokye, K. Kehulakae, S. Dimakatso, G. Boipelo and S. Sarefo, “Intelligent load balancing techniques in software defined networks: A Survey,” Electronics, vol. 9 no.7, pp. 1-24, 2020.

[6] D. Shen, W. Yan, Y. Peng, Y. Fu and Q. Deng, “Congestion Control and Traffic Scheduling for Collaborative Crowdsourcing in SDN Enabled Mobile Wireless Networks,” Wireless Communications and Mobile Computing, article no. 9821946, pp. 1-12, 2018.

[7] Q. Youssef, M. Yassine and A. Haqiq, “Secure Software Defined Networks Controller Storage using Intel Software Guard Extensions,” International Journal of Advanced Computer Science and Applications, vol. 11, no. 10, pp. 475-481, 2020.

[8] H. Polat and O. Polat, “An Intelligent Software Defined Networking Controller Component to Detect and Mitigate Denial of Service Attacks,” Journal of Information and Communication Technology, vol. 20, no. 1, pp. 57-81, 2021.

[9] H. Babbar, S. Rani, D. Gupta, H.M. Aljahdali, A. Singh and F. Al-Turjman, "Load Balancing Algorithm on the Immense Scale of Internet of Things in SDN for Smart Cities,” Sustainability, vol. 13, no. 9587, pp. 1-16, 2021.

[10] N. Fotiou, “Information-Centric Networking (ICN),” Future Internet, vol. 12, no. 35, pp. 1-2, 2020.

[11] I. Alam, K. Sharif, F. Li, Z. Latif, M.M. Karim, S. Biswas, B. Nour and Y. Wang, “A Survey of Network Virtualization Techniques for Internet of Things using SDN and NFV,” ACM Computing Surveys, vol. 53, no. 2, Article 35, pp. 1-40, 2020.

[12] M. Alotaibi and A. Nayak, “Linking handover delay to load balancing in SDN-based heterogeneous networks,” Computer Communications, vol. 173, pp. 170-182, 2021.

[13] O. Hohlfeld, J. Kempf, M. Reisslein, S. Schmid, and N. Shah, “Scalability Issues and Solutions for Software Defined Networks,” IEEE Journal on Selected Areas in Communications, vol. 36, no. 12, pp. 2595-2602, December 2018.

[14] B. Alankar, G. Sharma, H. Kaur, R. Valverde and V. Chang, “Experimental Setup for Investigating the Efficient Load Balancing Algorithms on Virtual Cloud,” Sensors, vol. 20, no. 7342, pp. 1-26, 2020.

[15] M. Hasan Al Bowarab, N. A. Zakaria, Z. Zainal Abidin, “Load Balancing Algorithms in Software Defined Network,” International Journal of Technology and Engineering, vol. 7, no. 6S5, pp. 686-693, 2019.

[16] D. O'Briain, D. Denieffe, D. Okello and Y. Kavanagh, “Enabling models of Internet eXchange Points for developing contexts,” Development Engineering, vol. 5, pp. 1-8, 2020.

[17] H. Mokhtar, X. Di, Y. Zhou, A. Hassan, Z. May and S. Musa, “Multiple-level threshold load balancing in distributed SDN controllers,” Computer Networks, vol. 198, pp. 108369, 2021. https://doi.org/10.1016/j.comnet.2021.108369.

[18] S. Liang, W. Jiang, F. Zhao and F. Zhao. "Load Balancing Algorithm of Controller Based on SDN Architecture under Machine Learning," Journal of Systems Science and Information, vol. 8, no. 6, pp. 578-588, 2020. https://doi.org/10.21078/JSSI-2020-578-11.

[19] K. Tolga Bagci and A. Murat Tekalp "SDN-enabled distributed open exchange: Dynamic QoS-path optimization in multi-operator services," Computer Networks, vol. 162, pp. 1–10, 2019. https://doi.org/10.1016/j.comnet.2019.07.001.

[20] S. Khan, F.K. Hussain and O.K. Hussain, "Guaranteeing end-to-end QoS provisioning in SOA based SDN architecture: A survey and Open Issues," Future Generation Computer Systems, vol. 119, pp. 176–187, 2021.

[21] H. Xue, K.T. Kim and H.Y. Youn, "Dynamic Load Balancing of Software-Defined Networking Based on Genetic-Ant Colony Optimization," sensors, vol. 19, no. 2, pp. 1-17, 2019.

[22] M.R. Belgaum, S. Musa, M.M. Alam, M.M. Suud, "A Systematic Review of Load Balancing Techniques in Software-Defined Networking," IEEE Access, vol. 8, pp. 98612-98636, 2020.

[23] S. Jamali, A. Badirzadeh, M.S. Siapoush, "On the use of the genetic programming for balanced load distribution in software-defined networks," Digital Communications and Networks, vol. 5, no. 4, pp. 288–296, 2019.

[24] X. Shi, Y. Li, H. Xie, T. Yang, L. Zhang, P. Liu, H. Zhang, Z. Liang, "An OpenFlow-Based load balancing strategy in SDN," Computers Materials and Continua, vol. 61, no. 3, pp. 385–398, 2019.

[25] K. Ramya, M. Sayeekumar, G. Karthik, "Software defined networking based solution in load balancing for media transfer in overlay network," Journal of Computational and Theoretical Nanoscience, vol. 17, no. 1, pp. 43–47, 2020.

[26] S. Prabakaran and R. Ramar, "Software Defined Network: Load Balancing Algorithm Design and Analysis," The International Arab Journal of Information Technology, vol. 18, no. 3, pp. 312-318, May 2021.

[27] O. M. A. Alssaheli, Z. Zainal Abidin, N.A. Zakaria, "Mininet Network Emulator: A Review," International Journal of Computer Science and Network Security, vol. 19, no 9, pp. 147-155, 2019.

An Efficient Productive Feature Selection and Document Clustering (PFS-DocC) Model for Document Clustering

Document Clustering using PFS-DocC Model

Perumal Pitchandi
Department of Computer Science and Engineering
Sri Ramakrishna Engineering College, Coimbatore, India

***Abstract*—In mining, document clustering pretends to diminish the document size by constructing the clustering model which is extremely essential in various web-based applications. Over the past few decades, various mining approaches are analysed and evaluated to enhance the process of document clustering to attain better results; however, in most cases, the documents are messed up and degrade the performance by reducing the level of accuracy. The data instances need to be organized and a productive summary have to be generated for all clusters. The summary or the description of the document should demonstrate the information to the users' devoid of any further analysis and helps in easier scanning of associated clusters. It is performed by identifying the relevant and most influencing features to generate the cluster. This work provides a novel approach known as Productive Feature Selection and Document Clustering (PFS-DocC) model. Initially, the productive features are selected from the input dataset DUC2004 which is a benchmark dataset. Next, the document clustering model is attempted for single and multiple clusters where the generated output has to be more extractive, generic, and clustering model. This model provides more appropriate and suitable summaries which is well-suited for web-based applications. The experimentation is carried out in online available benchmark dataset and the evaluation shows that the proposed PFS-DocC model gives superior outcomes with higher ROUGE score.**

Keywords—Benchmark standards; document clustering; productive feature selection; multiple clustering; web applications

I. Introduction

With the vast expansion towards the web and Internet applications along with the growth of mobile phones leads to the growth of enormous textual information [1]. This drastic explosion of data generation not only produces mess over document clustering and summarization. This complexity is not only encountered by the humans' but also by the machines which lag in processing the massive data generated from various sources like (applications, technologies, and organizations) [2]. The evaluations towards the huge amount of data are generally non-structural and quite a challenging task. The drastic eruption of documents over the web necessitates the path for document clustering and summarization process [3]. It attempts to give a shorter version of documents by maintaining the necessary information. The extensive insight of data makes the researchers to take appropriate decision by document clustering [4]. Thus, document clustering turns to be an essential approach in the growing world.

The document clustering or summarization helps in attaining a wider insight towards the data and offer decision making process [5]. For example, various social media applications like Facebook, Twitter and so on are used for personal causes for political and marketing purposes [6]. Recently, most of the political campaigns are made over these social media sources all over the world to reach the supporters in various regions. Therefore, the process of extracting the textual data is essential for successful political and marketing strategies [7]. Various real-time applications of document clustering are not constraint with these political and marketing strategies. For example, it is also employed for compressing the content for searching the outcomes over search engines along with the keyword for direct subscription towards the application [8]. Moreover, a proficient document clustering process over social media resources can preserve the user's trust relies on navigation among various contents [9].

The document summarization process includes huge challenge and the preliminary attempt is performed in 1950's when it uses features like phrase and word frequency to extracting essential sentences [10]. It is also considered as a huge demand in the field of research owing to its applicability. The finest way of summarization has to preserve the preliminary factors while assisting the users to have better insight towards the enormous volume of data in a faster manner [11]. The preliminary idea behind document clustering is to gather the more essential information in a clustered or with a compressed manner for certain tasks/users [12]. The clustering is also depicted as the gathering of data instances or the shortest document version which is gathered from the machine to attain most essential information is specific manner without human interventions. Moreover, the foremost definition is provided by [13], as 'text is gathered from one or more documents that provides essential information based on the source content and provides the shorter version of it". Based on this definition, there are three different factors that have to be concentrated: 1) clustering can be done with one or more documents; 2) clustering should preserve the essential parts of original content and 3) clustering have to gather the original source content without any reduction or alteration with original content [14].

There is various classification of document clustering process. Moreover, the process of document summarization is partitioned as: abstractive or extractive manner. The former model is to understand the textual content of the document profoundly and expresses the text in shorter manner. Subsequently, the target is to extract the document content to choose the most essential information [14]. It is extremely harder for the machine to generate the clustering of multiple documents which is smoother and understandable by the humans. In common practise, extractive approaches are generally used. From the various categories of document clustering process, recently, learning approaches are used for various documents clustering process [15]. The extractive process can be either supervised or unsupervised. In the former model, the problem is based on binary classification where the classes are defined with the summary; similarly, in the latter model, the ultimate target is to attain representative sentences. This research proposes a novel Productive Feature Selection and Document Clustering (PFS-DocC) model which is beneficial to handle the supervised and unsupervised challenges in an interpretable way. The anticipated model possesses the following characteristics:

1) Here, the challenges identified in clustering are considered as a single-objective problem. The clustering process attempts to identify the underlying data structure and provides the information for further classification purpose. Therefore, it enhances the performance of clustering algorithm.

2) The features are extracted with dynamical process via selective manner for all clusters. The clustering process should include the weight of the document by label discrimination to cluster the document.

3) The sentences are chosen in a way that it produces the clustering process in a non-redundant and coherent manner. The complex documents are placed at the top while remaining sentences are selected to gather the essential information with the redundancies.

The proposed Productive Feature Selection and Document Clustering (PFS-DocC) model obviates the requirement of feature engineering in a document clustering Even though, the most crucial phase over learning process in feature selection and extraction, various work concentrates in sentence clustering process. In recent time, various attempts to make to predict the optimal feature set for clustering process. This process considers the feature relevance as binary issues, that is, whether the features are attained from feature patterns. The overview of the Productive Feature Selection and Document Clustering (PFS-DocC) model is shown in Section 3. The samples of the document are chosen based on the feature vectors. The final outcomes need to similar group of samples with the features of similar group. The weighted features show similar features with clustering. In document clustering process, these clusters specify whether the document is efficient. The preliminary contributions of this clustering process are given below:

1) This work introduces the theoretical model based on productive manner. Here, a novel concept is the process of document clustering. This model facilitates the process of clustering the documents which helps in choosing the document. More specifically, the process of designing the clustering model is to measure the document sentences by labelling ‘0’ and ‘1’, respectively.

2) The proposed Productive Feature Selection and Document Clustering (PFS-DocC) model have the ability to measure the significance of the features by class discrimination which is clustered with various dataset over the reported dataset.

3) Here, evaluation is done with online available dataset to compute the clustering process in an efficient manner. It validates that the clustering process is less redundant and possess more information in a competitive manner.

4) Also, based on the comparison with prevailing approaches, Productive Feature Selection and Document Clustering (PFS-DocC) model gives added advantages which are less interpretable. It clearly states that the process tracks the cluster of document which is essential to explain the decision performed by the end-users.

II. Related Work

Document clustering is considered as an unsupervised approach for semantic clustering with the similar documents. The embedded documents are determined as a vector space and predict the neighbours over the space along with the clustering model based on word extraction which is extensively utilized. There are various investigations that enhance the performance with cluster initialization and automatic parameterization. Moreover, these approaches consider that all the provided documents are autonomous and do not determine the relationship strength among them. The document clustering model helps to get rid of various limitations that determine the relationship along with the document significance which is extensively investigated.

Network-based document clustering [16] is determined based on the interconnection among the documents and carry out document clustering based on network characteristics. It is depicted as the graphs that comprises of vertices related to the edges. Based on this analysis, generally it is considered as the vertices pair related with the edges to project semantic relationships. Then the assumption is based on the link strength and authority over the provided documents measured and the documents are clustered based on provided parameters [17]. This document clustering model performs hyper-linked web document classification based on academic papers and society, and citations. These approaches are utilized to demonstrate the semantic relevance among the news [18]. Therefore, it employs various kinds of meta-data and applied to various document ranges. This model is utilized to link a document that relies on content. It inter-connects various shared words with preliminary text documents. Therefore, documents are clustered relies on dependent association among the prevailing network; even in case of meta-data with absolute completeness.

The process is formally defined to carry out network-based document clustering. Kusner et al., [19] depicts network based document clustering formally with probabilistic generative model and utilized to cluster the documents. Therefore, the modelling of probabilistic generative model is anticipated and not applicable for multi-label clustering where the document is provided for multiple clusters. Moreover, the model is not suitable for various domains which encounter highly complex documents like certain documents and mobile applications which are allocated to multiple clusters [20]. Based on various analyses, network based document clustering offers multi-labelling process. Here, neighbourhood graph-based weighted matrix is used to evaluate the relationship strength among the documents with clustering process, concept factorization, and matrix factorization [21].

Moreover, diverse ranking models like search rank over search engines, paper classification and hubness value process which are utilized to compute the link strengths and document significance between the documents [22]. Hubness values are extensively utilized for evaluating the document significance. For instance, HITS and PageRank approaches are used for analysing the flow of web pages to search the documents and allocate higher value authorization to possess enormous number of inter-links [23]. Thus, these approaches are adopted over various documents and assigns higher authority values with huge amount of inter-links. Moreover, these approaches are considered to be the favourable older documents and assigns low authority values for all the newer documents. Thus, meaning-based search engine is adopted to handle these issues and projects the meaning-based information with document significance and un-important factors [24]. Thus, it enhances the processing speed.

Thus, search engine based significant ranking documents are based on semantic relevance and concentrates on internal meaning information [25]. The limitations over these methods are extremely prone for abusing which specifies internal inclusion of essential irrelevant words in context to actual documents. Based on various approaches, the proposed model makes use of document significance with indices that are autonomous independently with the document content like number of downloads over the mobile apps [26]. The given model preliminarily reduces the abuse by handling these issues over the network-based document significance examination.

The embedding documents are considered as the conversion of documents which includes word set with latent vectors [27]. It is utilized to evaluate the distance among the provided documents and consequently clusters the similar documents during document clustering model [28]. This model is extensively utilized for embedding document techniques which is composed of inverse document frequency and term frequency, topic modelling approaches termed as Latent Dirichlet allocation [29]. Various investigators consider document clustering by adopting topic modelling document embedment with k-means algorithm. The functionality of topic modelling is enhanced using the measure of documents with network modelling. This enhanced model uses document clustering.

Additionally, various researches are underway with word/document embedment with neural network approaches. The representative NN model is composed of word2vec which identifies the similar words form the input words. Similarly, Doc2vec predicts the word that offers the input document. In recent times, Doc2Vec, LDA, and TF-IDF are adopted to include the documentation [29]. The performance of document-clustering process is improved with the adoption of semi-supervised approaches that include the construction of initial-clusters which relies on words and enhances the similarity among the documents over the provided clusters via learning process. Word2Vec-based documents are used to predict, classify, and visualize social network neighbourhood [30]. Subsequently, embedding algorithm is alike of word2vec with certain exception and identifies the neighbourhood indeed of context words. In this research, a novel Productive Feature Selection and Document Clustering (PFS-DocC) model is proposed to reflect the document significance based on feature selection and document clustering. This model provides better performance based on consistent document information, document meta-data, and information clustered with input document. It is explained in the section given below.

III. Methodology

This research model includes three different processes: pre-processing, feature selection, and summarization. The evaluation is done with MATLAB environment using DUC 2004 dataset. The comparison is done with various metrics like accuracy, precision, F1-score, recall, ROUGE 1 and ROUGE 2 score. Also, the evaluation is done with DUC 2003 and DUC 2004 benchmark dataset. An extensive analysis is done with a proposed Productive Feature Selection and Document Clustering (PFS-DocC) model. Fig. 1 depicts the block diagram of proposed PFS-DocC model.

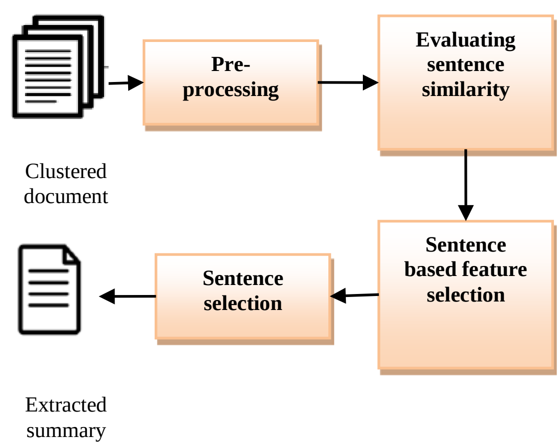


Fig. 1. Block Diagram of Productive Feature Selection and Document Clustering (PFS-DocC) Model.

A. Dataset

The DUC 2004 uses paper documents, newswire from TDT and TREC collections. The data is used for training, summarization exploration from the produced by machine translation. The task involves summarization by question and represents various tasks. The official ROUGE measures of DUC 2004 were 1-gram, 2-gram, 3-gram, and 4-gram and longest sub-string scores. The manual summarization is used for running ROUGE was provided to available participants. Thus, the truncated summaries longer than the targeted length before evaluation and generates summaries lesser than target length. The maximal target length was depicted based on bytes (punctuation, whitespace, and alphanumeric) included. The maximal target length for short summaries was 75 bytes. The shorter summaries are 665 bytes.

B. Pre-processing

The data (document) pre-processing is composed of linguistics, tokenization which provides a mathematical mode. It transforms the document content into sequence of terms which avoids punctuation and carries out removal of stop word (‘a’, ‘an’, ‘in’, ‘etc.’) are removed. There are enormous numbers of stop words.

C. Productive Clustering

The target of adopting productive clustering is used as a process of information retrieval. The user needs to scan the provided descriptors for relevancy measure and demonstrate that the clusters are relevant by manual processing of various document instances. The iterative process uses multiple stages of productive clustering to assist user for predicting the appropriate documents. The initial clustering is provided with description or clusters to the users who selects clusters of own interest. The text instances over the chosen clusters are merged and clustered. This process is continued with appropriate set of documents. The automatic description of quality is crucial for facilitating users to predict which clusters the relevant text.

The productive clustering is performed by initially clustering and predicts the set of features related with cluster. It facilitates appropriate clustering algorithm (see Algorithm 2) to be adopted. The chosen features provide best information to the users based on the users’ content (cluster). The preliminary process is to demonstrate the clusters with likely words over the cluster. But, the features are not optimal for establishing discrimination among various clusters. The scoring criterion includes information gain (mutual information).

D. Feature Selection

For the provided clusters, the prediction of instances from the input clusters handle the conventional classification problem and selection of appropriate feature subset is more essential. The selection of smaller subset with maximal feature prediction is a complex task. In smaller feature subset, step-wise similarity measure is carried out to enhance the classification performance. The model should fit with the features which cannot scale the features that are encountered with the textual data. The proposed model should trace number amount of features. The feature selection process has to ensure the process by positive correlation with target class, that is, feature occurrence rate of provided class which is higher than average rate.

Algorithm 1: Evaluating sentence score for similarity measure

Input: Array of sentences
Output: Similarity scores
1. Average weighted matrix [n][n];
2. Scores [n];
3. for $i \rightarrow 1\ to\ n$ do;
4. for $j \rightarrow 1\ to\ n$ do;
5. predict = $identity - similarity\ (s[i], s[j])$;
6. Average similarity matrix [i][j] = average value (id);
7. end
8. end
9. score = id (average similarity matrix);
10. return scores;

The productive clustering model is composed of two preliminary tasks: identifying the original occurrence of features based on cluster allocation and identifying the instances of certain cluster with smaller feature dimensionality set that functions as the cluster descriptions. This task offers an objective to automatically choose from clustering with various numbers of clusters. The cluster is related with various feature distributions with lesser frequency instances over the clusters. When the instances are allocated with similar cluster have same feature distribution where the cluster allocation is productive of feature occurrence. The successive task is the prediction of clustering membership with dependency over the selected clusters. The total information attained by clustering increases with clusters; however the complexity of finding the cluster membership increases with the fine-grained clusters. Also, the added numbers of clusters are more inherent and provide better trade-off among the number of features and prediction performance. The traceability process needs to be performed with number of available clusters and features for multi-document clustering process. The feature selection model is evaluated and chosen for candidate set.

In cluster creation process, the set of probable clusters are generated that varies from the total number of clusters which arises from various clustering process or various data specification. For all clusters, the model is trained to find the feature occurrence from allocated clusters. The association among the clusters are more productive. The feature subsets are chosen based on the selection mechanism. Specifically, feature subsets are predicted with positive constraints by changing regularization process. Every stage is allocated with standard modelling like clustering and model selection.

The productive framework is provided by allocating clusters that gives flat clustering. The clustering process is effectually executed with sparse data when the similarity among the data is utilized and changes the number of clusters to generate set of clustering process $\phi = \{\phi_{\mathbf{1}}, \dots, \phi_{\boldsymbol{K}}\}$. Here, various clustering process are chosen from the feature vectors that are attained from cluster assignments. The probability of the feature set occurrence is provided with cluster assignment $\boldsymbol{y} = [\boldsymbol{y}^{(\mathbf{1})}, \dots, \boldsymbol{y}^{(\boldsymbol{C})}]$ which is expressed as in Eq. (1):

$$prob\ (X = 1|y) = \frac{1}{1+e^{(-X_0)}} \tag{1}$$

The above equation is expressed with binary features where X_0 is bias compactness with coefficient of parameter vector $u' = [x_0, x]$ and constant features are added with cluster allocation. The cluster is a supervised learning problem which includes both feature selection and training process. The cluster $c \in \{1, \ldots, C\}$ with feature subset prediction is expressed as in Eq. (2):

$$\arg\min E[Loss\ (x^{(c)}, \hat{y}^{(c)} + \ \alpha\, \Omega\ (w^{(c)}) \tag{2}$$

Here, $w^{(c)}$ is weighted co-efficient for providing the feature ranking with original cluster vectors. The feature constraints are related with the clusters positively. The feature subset reduces the computational cost of the provided model. For the provided cluster, the probability instances are allocated with the cluster as conditional random variables. It is expressed as in Eq. (3):

$$prob\ (Y = 1|x) = \ f_{w'}(x) \tag{3}$$

It is provided as the bias compactness with the integration of co-efficient and constant which is included at the feature vectors. The minimization problem with appropriate solution in expressed as in Eq. (4):

$$\arg\begin{matrix}min\\ w'\end{matrix} - \ln L(w') + \ \frac{\tau}{2}\ \left||w|\right|_2^2 \tag{4}$$

Generally, the features are related with various non-zero coefficients. The equivalent constraints provide solution when the coefficients are zero. It is shown in Eq. (5):

$$\arg\begin{matrix}min\\ w'\end{matrix} - \ln L(w') + \ \gamma\ \left||w|\right|' \tag{5}$$

Here, γ influences the number of features with non-zero co-efficient where the larger value of γ which yields better solution with non-zero co-efficient. These non-zero co-efficient are provided with chosen features. The suitable feature subsets are determined by sweeping the γ values. The feature subset and the weighted co-efficient are used to choose appropriate feature subset. Consider a feature subset $S_1, \ldots, S_j$ for certain cluster and optimal feature subset is chosen with Eq. (6):

$$\hat{j} = \arg min - \ln L\big(w_j\big) + \ \left|S_j\right| \ln\sqrt{n} \tag{6}$$

Here, S_j is set of features that do not include feature subset. The coefficients are generated from the provided subset. The bias value sometimes influences the chosen subsets. The numbers of interpretable features are stable over the sample size variations. In practical condition, the feature subsets are restricted based on the size. The user needs to deal with enormous features to demonstrate the clustering process. Sometimes, the limit may reduce the productive performance; also it reduces the computational complexity with number of feature subsets during evaluation process. The analysis is done with publicly available dataset. The anticipated model is based on set of predictive features. The numbers of features are restricted with positive correlation among the clusters and classes. The positive constraints are provided based on classification performance.

For the computation of cluster predictions, here f1-score is used for individual clusters or classes where the summarization is resulted with the average of computed F1-score. The data instances are allocated with multiple clusters and not allocated with various available clusters. The instance possesses equivalent weight among distributed among the assigned values. The un-allocated instances are determined based on the valid group of added clusters. The mutual information is extracted from the discrete variables partitioning. The computation is done with automatic selection of total clusters whether the numbers of clusters correlate the maximal information content. The major drawback associated with existing approaches is the evaluation of multi-modal distributions of all features with higher computational complexity $O\ (N^2)$. The redundancy elimination is done with candidate features by setting the divergence among the multi-modal distributions. The scalability is done with the features of higher score over the targeted clusters.

Algorithm 2: Document clustering

Input: Array of sentences
Output: sentence score
1. Similarity matrix [n][n];
2. Array scores [n];
3. for $i \rightarrow 1\ to\ n$ do;
4. for $j \rightarrow 1\ to\ n$ do;
5. DocC [i][j] = measure similarity (S[i], S[j]);
6. end
7. end
8. DocC = Similarity matrix;
9. Hyper-linked similarity matrix;
10. for $i \rightarrow 1\ to\ n$ do;
11. score [i] = average summarization;
12. end
13. return scores;

This work concentrates in computing the appropriate selection of number of clusters with the Productive Feature Selection and Document Clustering (PFS-DocC) model. This model enhances and maximizes the information attained by the clustering algorithm. The experimentation is done to compute the information among the original clusters and the chosen clusters are varied based on proportional cluster number. The productive document clustering facilitates both the number of features and clusters which is utilized to determine the cluster. The user needs to select appropriate range data clusters with computational feasibility. The user needs to enhance the range of more optimal clustering process. The anticipated model is utilized to any data with weighted features. The productive r with productive features and cluster assignment is prediction with feature subset. The outcomes are attained based on the every cluster with minimal amount of feature subset which is essential to identify the instances that belongs to certain clusters. The productive clustering model is used to predict the cluster membership of given document. The relevance of the proposed PFS-DocC model is efficient to give higher amount of information with reduced data redundancy. The section below discusses the numerical outcomes attained with the analysis of proposed PFS-DocC model.

IV. Numerical Results and Discussion

The performance of the proposed Productive Feature Selection and Document Clustering (PFS-DocC) model based on clustering, information extraction, and non-redundancy and overall processing is evaluated. Some metrics like accuracy (%), recall (%), F1-score (%), and precision (%) are measured. For this evaluation, online available DUC 2004 dataset is a generic model for document clustering. It includes 50 clusters of new documents. These clusters include the summaries of various human references which are considered by the researchers for extracting the outcomes. It is essential to set the length of document clusters. The clusters over DUC 2004 organize 665 bytes where the pre-processing step is extremely needed for accuracy evaluation. Here, some essential pre-processing steps are performed with text documents. Generally, the documents are processed to predict the document source information from textual components. The initial process needs to eliminate the information tags such as <TEXT>, <DOC>, and so on for processing the documents.

The experimentation performance is measured with evaluation toolkit known as ROUGE which is a recall based evaluation metrics. It computes the efficiency of document clustering for evaluating the summaries generated by the humans. The ROUGE score evaluates the number of successive terms. After the completion of pre-processing steps, the similarity measures among the sentences are evaluated using the proposed Productive Feature Selection and Document Clustering (PFS-DocC) model. The probability occurrences of the words from the input clusters are used to identify the productive words. The clusters are summarized with the clusters over the dataset. The outcome of the discriminant analysis is measured with metrics like True Negative (TN), True Positive (TP), False Positive (FP), and False Negative (FN) are known as correct predictions with negative samples, correct prediction with positive instances, incorrect prediction with positive samples, and incorrect predictions with negative instances, respectively. It is expressed as in Eq. (7) - Eq. (10):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{7}$$

$$F1-score = 2 * \frac{Precision*Recall}{(Precision+recall)} \tag{8}$$

$$Recall = \frac{TP}{TP+FP} \tag{9}$$

$$Precision = \frac{TP}{TP+FP} \tag{10}$$

The simulation is carried out in MATLAB environment. Here, six different methods along with the Productive Feature Selection and Document Clustering (PFS-DocC) model are compared. The six methods are FLSA (ProbIDF), FLSA (Normal), FLSA (IDF), FLSA (Entropy), LDA, and LSA respectively. Similarly, metrics like Accuracy (%), F1-score (%), Recall (%), and precision (%) is evaluated. The accuracy of proposed Productive Feature Selection and Document Clustering (PFS-DocC) model is 98.9% which is 1.9%, 7.9%, 3.9%, 1.9%, and 38.9% higher than the prevailing methods. The F1-score of PFS-DocC is 99% which is 29.7%, 27.6%, 1.3%, 3.5%, 7.8%, and 1.3%, respectively. The recall of PFS-DocC is 99% which is 27%, 26%, 4%, 6%, 10%, and 4% higher than the other models. Similarly, precision of PFS-DocC is 99% which is 33%, 30%, 4%, 6%, 10%, and 4% higher than other models. All these process includes 50 topics. It is shown in Table I. Fig. 2 depicts the performance metrics evaluation. Fig. 3 depicts the F1-score computation.

TABLE I. Comparison of Performance Metrics

Methods	**Accuracy (%)**	**F1-score (%)**	**Recall (%)**	**Precision (%)**	**Topics**
FLSA (ProbIDF)	97	69.3	72	66	50
FLSA (Normal)	91	71.4	73	69	50
FLSA (IDF)	95	97.7	95	95	50
FLSA (Entropy)	97	95.5	93	93	50
LDA	60	91.2	89	89	50
LSA	57	97.7	95	95	50
PFS-DocC	98.9	99	99	99	50

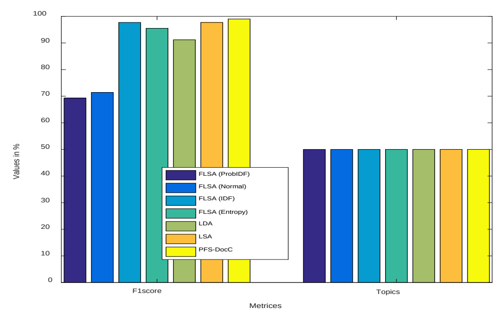


Fig. 2. Performance Metrics Evaluation.

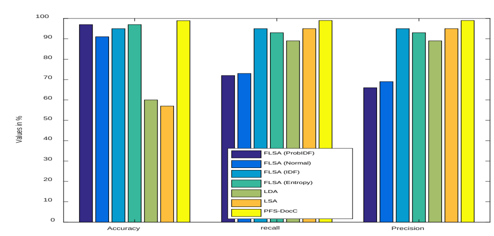


Fig. 3. F1-Score Computation for 50 Topics.

Table II depicts comparison of ROUGE 1 score and ROUGE 2 score with the evaluation toolkit. The comparison is done for ExDoS, Banditsum, HSSAS, summaRunner, NN-SE, LEAD-3, and PFS-DocC respectively. Rouge 1 score is 45 which is 3%, 4%, 3%, 6%, 10%, and 6% higher than other models. Rouge 2 score of PFS-DocC is 2%, 1.9%, 3%, 4%, 7%, and 4.7%, respectively. Finally, Rouge L score of PFS-DocC is 39 which are 4%, 7%, 4%, 2%, 1.5%, and 1% higher than other models (see Fig. 4). Table III shows the amount of information extracted, non-redundant, overall percentage achieved. PFS-DocC based information extraction is 30%; however for other approaches it is 27%, 23.5%, 20.5%, 17.6%, 13.5%, and 13% respectively (see Fig. 5). The avoidance of non-redundant data from PFS-DocC is 25% where the other data is 22.5%, 22.6%, 16.5%, 19.5%, 21%, and 23% respectively (see Fig. 6). The overall performance of PFS-DocC w.r.t information extraction and non-redundancy avoidance is 27%; whereas for other models it is 25%, 18.5%, 21.6%, 16.8%, 20.8%, and 22%, respectively.

TABLE II. ROUGE SCORE EVALUATION

Methods	Rouge 1 score	Rouge 2 score	Rouge L score
ExDoS	42	18.5	35
Banditsum	41	18.6	32
HSSAS	42	17.5	35
SummaRunner	39	16.5	37
NN-SE	35	13.5	37.5
LEAD-3	39	15.8	38
PFS-DocC	45	20.5	39

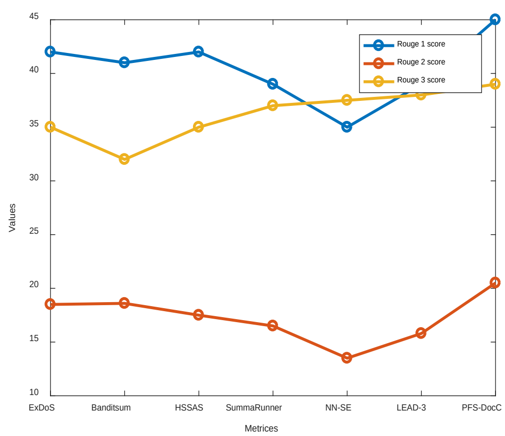


Fig. 4. ROUGE Score Computation.

TABLE III. INFORMATION EXTRACTION AND NON-REDUNDANCY PERCENTAGE

Methods	Information extraction	Non-redundancy	overall
ExDoS	27%	22.5%	25%
Banditsum	23.5%	22.6%	18.5%
HSSAS	20.5%	16.5%	21.6%
SummaRunner	17.6%	19.5%	16.8%
NN-SE	13.5%	21%	20.8%
LEAD-3	13%	23%	22%
PFS-DocC	30%	25%	27%

Table IV depicts the comparison of PFS-DocC without feature extraction is done with benchmark datasets, like DUC2002-ROUGE 1, DUC2002-ROUGE 2, Main-ROUGE 1, Main-ROUGE 2, DUC2004-ROUGE1, and DUC2004-ROUGE 2. The values of PFS-DocC (without feature extraction) are 53, 26.7, 42.5, 19, 55, and 57 respectively. Similarly, the values of PFS-DocC are 46, 22.5, 39.7, 15, 50, and 53 respectively (see Fig. 7 and Fig. 8).

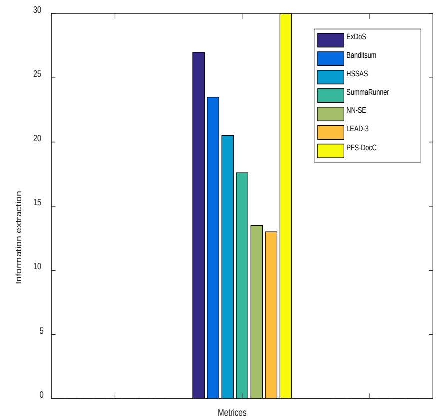


Fig. 5. Information Extraction.

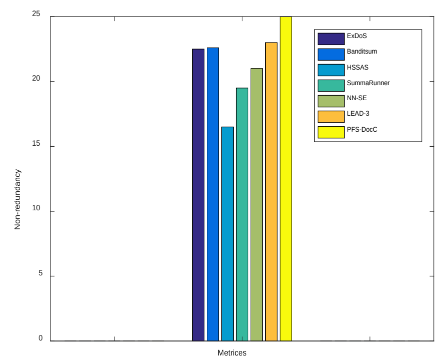


Fig. 6. Non-Redundant Data Extraction.

TABLE IV. PFS-DOCC COMPARISON (WITH / WITHOUT FEATURE SELECTION)

Methods	DUC2002-ROUGE 1	DUC2002-ROUGE 2	Mail-ROUGE1	Mail-ROUGE2	DUC2004-ROUGE 1	DUC 2004 – ROUGE 2
PFS-DocC	53	26.7	42.5	19	55	57
PFS-DocC + feature extraction	46	22.5	39.7	15	50	53

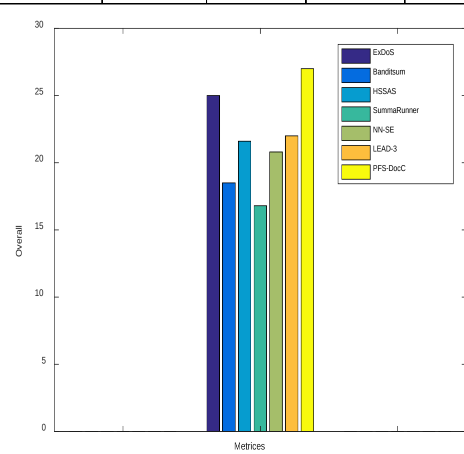


Fig. 7. Overall Performance Measure of PFS-DocC.

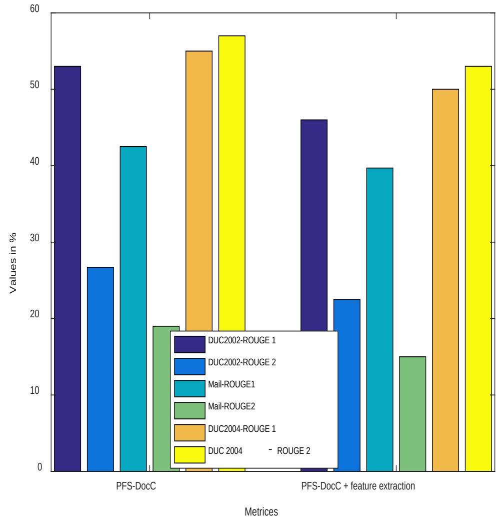


Fig. 8. PFS-DocC Performance (with / without Feature Extraction).

The similarity score is examined based on the sentence that is extracted from the central score. Initially, the document clusters are converted to connected sentence using various similarity scores. Based on the experimentation, the proposed PFS-DocC model enhances the summarization process attained from document clustering. After the extraction process, the sentences are provided with high score and include the summary length. It is essential to improve the extracted sentences which do not possess any redundant information. Therefore, to diminish the redundancy over any sentences with the similarity measure of extracted summary the proposed PFS-DocC is used.

V. Conclusion

This research concentrates on proposing a novel Productive Feature Selection and Document Clustering (PFS-DocC) model with three essential steps that includes background knowledge, pre-processing, feature selection, and summarization (clustered document). It is to enhance the performance of the proposed PFS-DocC model. Here, DUC 2004 online available dataset is used for evaluation. The input from the dataset is given for pre-processing and further process is carried out. The similarity and the correlation among the clustered document are examined and summarized to extract the essential features for provided document. Therefore, the proposed PFS-DocC model enhances the performance of the clustering algorithm. The simulation is done with MATLAB environment.

The performance of the PFS_DocC model is evaluated with the adoption of DUC 2004 benchmark dataset. The performance is measured using the ROUGE score toolkit. Various metrics like accuracy, F1-score, recall, and precision are measured for PFS-DocC model with 98.5% accuracy and 99% F1-score, recall, and precision. The outcome of the proposed PFS-DocC model is higher when compared to other approaches like FLSA (ProbIDF), Prob (Normal), FLSA (IDF), FLSA (Entropy), LDA, and LSA respectively. Similarly, the comparison is done with two benchmark dataset known as DUC 2003 and DUC 2004 for evaluating the performance of PFS + DocC with and without feature selection process. Also, the information extracted and the non-redundant data evaluation is also done for the PFS + DocC model. The performance show better trade-off in contrast to prevailing approaches. However, there is a constraint, as the proposed PFS + DocC model does not provided for classification. It will be concentrated in future along with the optimization process.

Acknowledgment

I thank the Management, Principal and Head of the department to support and provide the resources to carry out this research work.

References

[1] Vinaitheerthan Renganathan, Text mining in biomedical domain with emphasis on document clustering,' Healthcare Inform. Res., 23 (2017), 141_146, http:// 10.4258/hir.2017.23.3.141.

[2] Cheng and M. Lapata, Neural summarization by extracting sentences and words, 2016, arXiv:1603.07252. [Online]. Available: https://arxiv.org/abs/1603.07252.

[3] Yang, X. Cai,Y. Zhang, and P. Shi, Enhancing sentence-level clustering with ranking-based clustering framework for theme-based summarization, Inf. Sci. 260 (2014), 37-50.

[4] Hong, M. Marcus, and A. Nenkova, System combination for multidocument summarization, in Proc. Conf. Empirical Methods Natural Lang. Process. 2015, 107-117.

[5] Wang, W. Lam, Z. Ren, and L. Bing, ``Salience estimation via variational auto-encoders for multi-document summarization,'' in Proc. 31st AAAI Conf. Artif. Intell. 2017, 1-9.

[6] Cao, F. Wei, L. Dong, S. Li, and M. Zhou, ``Ranking with recursive neural networks and its application to multi-document summarization,'' in Proc. 29th AAAI Conf. Artif. Intell. 2015, 1-9.

[7] Ren, Z. Chen, Z. Ren, F.Wei, L. Nie, J. Ma, and M. De Rijke, Sentence relations for extractive summarization with deep neural networks, ACM Trans. Inf. Syst. 36 (2018), 1-32.

[8] Ren, F. Wei, Z. Chen, J. Ma, and M. Zhou, A redundancy-aware sentence regression framework for extractive summarization, in Proc. 26th Int. Conf. Comput. Linguistics Tech. Papers. 2016, 33-43.

[9] Cao, F.Wei, S. Li,W. Li, M. Zhou, and W. A. N. G. Houfeng, Learning summary prior representation for extractive summarization,' in Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics. 2 (2015), 829-833.

[10] Hong and A. Nenkova, Improving the estimation of word importance for news multi-document summarization, in Proc. 14th Conf. Eur. Cha ter Assoc. Comput. Linguistics. (2014), 712-721.

[11] Cao, W. Li, S. Li, and F. Wei, Retrieve, Rerank and rewrite: Soft template based neural summarization, in Proc. 56th Annu. Meeting Assoc. Comput. Linguistics. (2018), 152-161.

[12] Tang, L. Yan, Z. Yang, and Q. H. Wu, Improved document ranking in ontology-based document search engine using evidential reasoning, IET Software. 8 (2014), 33–41.

[13] Huang and X. X. Zhou, Knowledge model for electric power big data based on ontology and semantic web, CSEE Journal of Power and Energy Systems. 1 (2015), 19–27.

[14] MacQueen, Some methods for classification and analysis of multivariate observations, in Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. 1 (1967), 281–297.

[15] Vega-Pons and J. Ruiz-Shulcloper, A survey of clustering ensemble algorithms, International Journal of Pattern Recognition and Artificial Intelligence. 25 (2011), 337–372.

[16] Vega-Pons, J. Correa-Morris, and J. Ruiz-Shulcloper, Weighted partition consensus via kernels, Pattern Recognition. 43 (2010), 2712–2724.

[17] Zhang, Y.-F. Pu, S.-Q. Yang, J.-L. Zhou, and J.-K. Gao, An ontological Chinese legal consultation system,' IEEE Access. 5 (2017), 18250-18261.

[18] Kim, H. Jang, H. J. Kim, and D. Kim, A document query search using an extended centrality with the Word2vec,' in Proc. ICEC, Suwon, South Korea, 2016, Art. no. 14.

[19] Kusner, Y. Sun, N. I. Kolkin, and K. Q. Weinberger, from Word Embeddings to Document Distances,' in Proc. ICML, Lille, France, 2015, 1-10.

[20] Koniaris, G. Papastefanatos, and Y. Vassiliou, Towards automatic structuring and semantic indexing of legal documents,' in Proc. PCI, Patras, Greece, 2016.

[21] Zhang, Y.-F. Pu, and P. Wang, An ontology-based approach for Chineselegal information retrieval,' in Proc. CENet, Shanghai, China, 2015, 1-7.

[22] Zhang, Y. Xu, and W. Zhang, Clustering scientific document based on an extended citation model, IEEE Access. 7 (2019), 57037–57046.

[23] Yoon, J. Lee, S.-Y. Park, and C. Lee, Fine-grained mobile application clustering model using retrofitted document embedding, ETRI J. 39 (2017), 443−454.

[24] Nayak, Fine-grained document clustering via ranking and its application to social media analytics, Soc. Netw. Anal. Min. 29 (2018).

[25] Kim, D. Seo, S. Cho, and P. Kang, Multi-co-training for document classification using various document representations: TF–IDF, LDA, and Doc2Vec, Inf. Sci. 477 (2019), 15–29.

[26] Chen, F. S. C. Tseng, and T. Liang, An integration of WordNet and fuzzy association rule mining for multi-label document clustering, Data Knowl. Eng. 69 (2010). 1208−1226.

[27] Duan, Y. Li, R. Li, R. Zhang, X. Gu, and K. Wen, LIMTopic: A framework of incorporating link based importance into topic modeling, IEEE Trans. Knowl. Data Eng, 26 (2014), 2493–2506.

[28] Chali and S. A. Hasan. Query-focused multi-document summarization: Automatic data annotations and supervised learning approaches, Natural Language Engineering. 18 (2012), 109-145.

[29] Liu, J. Flanigan, et al. Toward Abstractive Summarization Using Semantic Representations. In HLT-NAACL (2015).

[30] Li, D. Cheng, L. He, et al. Joint Event Extraction Based on Hierarchical Event Schemas from FrameNet[J]. IEEE Access, 7(2019), 25001-25015.

Smart Elevator Obstruction Detection System using Image Classification

Preethi Chandirasekeran[1]
B.Tech. Student, Computer Science and Engineering
Vellore Institute of Technology, Chennai
Kelambakkam - Vandalur Rd, 600127 Tamil Nadu, India

Shridevi S[2]
Centre for Advanced Data Science
Vellore Institute of Technology, Chennai
Kelambakkam - Vandalur Rd, 600127 Tamil Nadu, India

***Abstract*—This paper proposes an approach that leverages real-time Image Classification to improve elevator safety. Elevators are a necessity for most multi story buildings. As a result, they play a crucial role in the lives of millions of people around the world. Despite this, there has been limited advancement in the technology used for elevator door operators. In the current system, elevators use multiple infrared transmitter/receiver pair of sensors to detect obstructions between the doors. This does not effectively detect smaller objects such as pets, small children, pet leashes etc. between the elevator doors which has led to thousands of tragic fatalities. This paper proposes an approach to tackle this challenge by leveraging Binary Image Classification to determine whether there is an obstruction between the elevator doors. This study includes the construction of a novel dataset of over 10,000 images and a comprehensive evaluation and comparison of several Machine Learning models for the proposed system. The results have produced novel findings that can be used to significantly improve safety and reliability of elevator door operators by preventing tragic fatalities every year.**

***Keywords*—Binary image classification; machine learning; deep learning; elevator safety**

I. Introduction

In today's world, there are millions of multi-story buildings across the globe. Not only do elevators improve the convenience of vertical transportation, but they also play a vital role in providing accessibility for people with disabilities. It is clear that elevators have become a necessity in our daily life. It is estimated that the door operator works approximately 1.75 million times in 10 years [1].

In the early 2000s elevators were fitted with a retractable door edge such that if the elevator door came in contact with a person or obstruction while closing, the doors would stop closing. Modern elevator door operators use non-contact detection device which includes a photoelectric detection device, ultrasound monitoring device, red light curtain detector, electro-magnetic induction detector, and other inputs to detect obstructions between the doors [2]. However, many cases have shown that such technology has limitations when detecting small obstructions such as leashes, pets or children [3]. Because of this inefficiency, there are thousands of fatalities every year. Despite the widespread use of this elevator technology and the serious consequences of its inefficiencies, there has been limited research on the technology used for elevator door controllers. Although there has been growing interest in developing Smart Buildings and Smart Cities, elevator door control remains an underexplored research area. There is a large scope to use modern techniques from computer vision and artificial intelligence to improve the existing elevator door control technology in a cost-effective manner. Widespread implementation of such improved systems has the potential to save lives and prevent injuries. This paper seeks to tackle the aforementioned challenge of improving the reliability and efficiency of elevator door operators.

Machine learning and image classification techniques – to the best of our knowledge – have not yet been used to address the challenge of improving the safety and reliability of elevator door operators.

The contribution of this paper is to:

- Propose a system that leverages machine learning and image classification techniques to overcome the shortcomings of existing elevator door control operators.
- Construct a novel dataset of over 10,000 images which serve as input to the proposed system. The images are captured by simulating the 8 positions along the top and sides of the elevator car door. Brightness augmentation is performed to improve the robustness of the system in different daylight conditions.
- Evaluate and compare the speed and efficacy of several machine learning models, namely Linear SVC, k-NN, Decision Tree, Random Forest, and CNN.
- Identify the most suitable classifiers for the proposed system based on the results of this study.
- Assess the feasibility of the proposed system based on the results of this study.

This paper proposes an approach to address the problem of elevator door operator obstruction detection to improve safety of the passengers entering and exiting the elevator in real-time.

The following section describes related research works followed by a description of the materials and methods. This includes details regarding the dataset construction, proposed system, models considered and evaluation metrics. Then the results of the study are presented. This is followed by the conclusion and future scope of the paper.

II. Related Work

In “Discussion on Improving Safety in Elevator Management”, [4] Feng ShuanChang et al. discuss the need to improve elevator technology given its direct relation to people’s lives. The authors also elaborate on the complexity of the elevator user unit and the lack of ownership for elevator maintenance among stakeholders which often leads to neglect of elevator safety.

Magota et al. [5] improve the opening-and-closing speed control of the elevator doors using ILQ design method with frequency shaping. The authors focused on developing a smoother and quicker elevator door control system rather than addressing the challenge of efficient obstruction detection.

Zeng et al. [6] propose the use of the atmega32 control chip and ov7620 image acquisition chip to tackle this challenge. The authors adopt an image processing-based approach by implementing Semi-Neighborhood Averaging Algorithm and Background Difference Method on the image pixels to identify the moving obstruction.

Various systems have been proposed in existing literature to predict the failure of elevators based on knowledge graphs [7] and neural networks [8]. Yanbin Guo et al. [9] propose a system to monitor the elevator for the most common faults based on multi-sensor fusion. This system includes a web client to view the live reports of the system and a camera to monitor the passengers inside the elevator car.

Amruthnath and Gupta [10] research the use of unsupervised machine learning algorithms, such as hierarchical clustering, k-means and fuzzy c-means, for predictive maintenance of elevators.

In “Machine Learning Modelling for Failure Detection of Elevator Doors by Three-Dimensional Video Monitoring” [11] Chih-Yu Hsu et al. propose a method for detection of elevator failure by using three-dimensional video monitoring. This method involves extracting the signal from the dynamic distances between elevator doors and modelling by trapezoidal curves. The failure detection is done by identifying these dynamical signal curves of the elevator doors using trained classifiers.

Wan et al. [12] study the use of least square support vector machine for diagnosing elevator various malfunctions using vibration signals from the elevator functions. The authors optimize the parameters of LS-SVM using K-fold cross validation.

Mishra and Huhtala [13] propose an algorithm to extract highly information features from the elevator data which is used to improve fault detection. This algorithm uses profile extraction and deep autoencoder feature extraction on the dataset.

The use of big data for fault warning of the elevator system has also been explored [14] [15]. Ming Zihan et al. [16] propose an alternate method for monitoring the elevator system by leveraging IoT.

Olalere and Dewa [17] employ a remote condition monitoring (RCM) approach for proactive maintenance of elevators. This approach uses vibration and machine-room temperature data acquired through IoT devices. Systems to protect passengers from the accidental movement of the elevator car have also been proposed in previous research [18] [19].

With the prevalence of big data, data mining technology and artificial intelligence have shown great promise in detecting the failure and evaluating the risk of elevators [20]. Wang et al. [21] study the use of Hadoop to design the data mining and analysis platform. The authors also use Hadoop to improve and parallelize the K-means and Apriori algorithms to mine the massive amounts of data from the elevator monitoring database.

A majority of existing research investigates the use of modern technology to improve maintenance of the elevator by remotely monitoring the overall elevator system for signs of faulty behavior or damage. However, related works do not address the shortcomings of existing elevator technology in effectively detecting obstructions between the doors.

In this paper, we tackle the aforementioned shortcomings by proposing a novel approach that leverages binary image classification to efficiently detect obstructions between the elevator doors. An image dataset of over 10,000 images was constructed to simulate the working of the proposed system. Several image classifiers were built, evaluated and compared on both the augmented and original image datasets.

III. Materials and Methods

A. Dataset

The dataset was collected by simulating the 8 camera positions, as shown in Fig. 1, using a functional elevator in a multi-story building. This consisted of two cameras positioned at the top view of the elevator and 3 cameras at equidistant positions along each elevator car door. Each camera recorded videos of the various obstructions passing through the elevator car doors. Every fifth frame was extracted from the videos to compile the image dataset. The code to perform this frame extraction is displayed in Fig. 2. Each image was manually labelled as either ‘Obstruction’ or ‘No Obstruction’.

Brightness augmentation was performed on the images to simulate various lighting conditions. Thereby, improving the robustness of the obstruction detection at different hours during the day. Each image was augmented using 12 different brightness factors ranging from 0.5 to 1.6 as shown in Fig. 3. The code used to perform augmentation is displayed in Fig. 4. Grayscale augmentation was also explored however it had minimal effect on the performance of the binary image classification models. The images were divided into two datasets, namely side view images and top view images. A sample of the side view and top view images are shown in Fig. 5 and 6 respectively. The images captured have a size of 1,280 x 720 pixels as well as a horizontal and vertical resolution of 96 dpi. Each dataset is fairly balanced, as shown in Table I. The total size of both datasets together is 10,836 images.

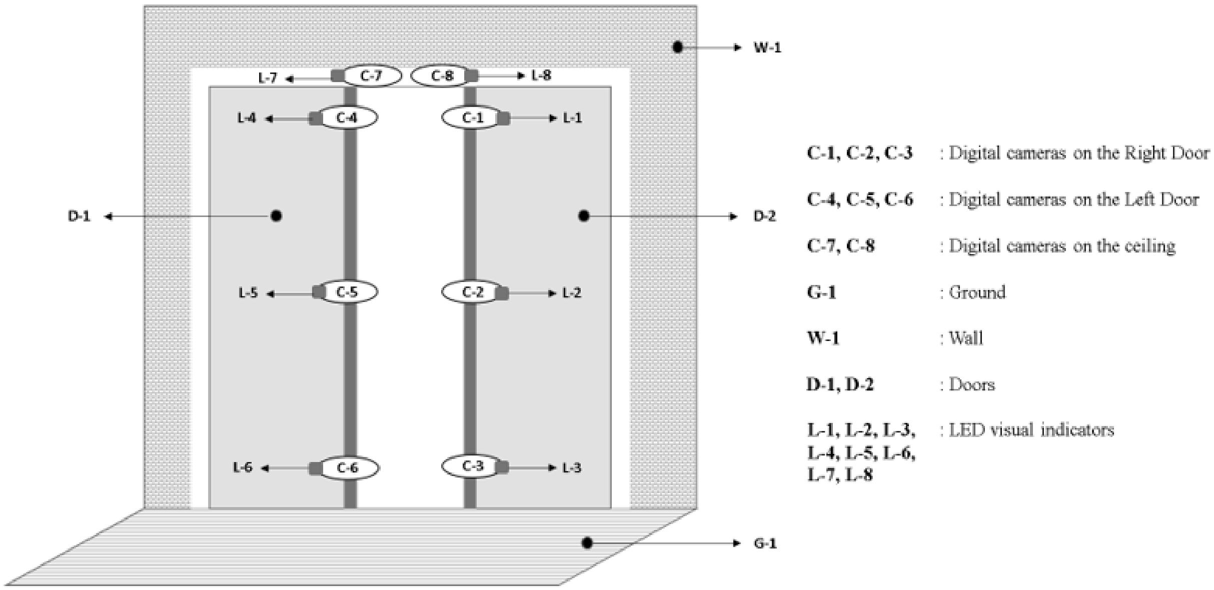


Fig. 1. Positioning of the Cameras and LEDs in the Proposed System.

```
vidcap = cv2.VideoCapture('Video.mp4')
success,image = vidcap.read()
count = 0
while success:
 if count%5==0:
 cv2.imwrite("Data Collection\\Video\\frame_%d.jpg" % count, image)
 success,image = vidcap.read()
 print('Read a new frame: ', success)
 count += 1
```

Fig. 2. Code Extracting Every Fifth Frame from a Video.

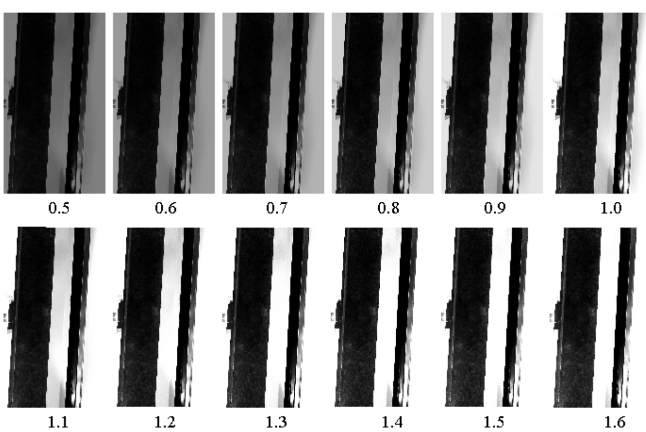


Fig. 3. Twelve levels of Brightness Augmentation used on the Dataset.

```
All_files=os.listdir(InDir_path)
for curr_fileName in All_files:
 full_in_file_name = InDir_path+"\\"+curr_fileName
 curr_length = len(curr_fileName)
 part_1=curr_fileName[0:(curr_length-4)]
 part_2=curr_fileName[(curr_length-4):99]

 try:
 im=Image.open(full_in_file_name)
 enhancer = ImageEnhance.Brightness(im)
 factorList=[0.5,0.6,0.7,0.8,0.9,1,1.1,1.2,1.3,1.4,1.5,1.6]
 counter=0
 for factor in factorList:
  im_output = enhancer.enhance(factor)
  full_out_file_name =
OutDir_path+"\\"+part_1+"_"+str(counter)+part_2
  im_output.save(full_out_file_name,"JPEG")
  counter=counter+1
 except IOError:
 print("cannot process file",full_out_file_name)
```

Fig. 4. Code to Perform Brightness Augmentation.

B. Proposed System

The elevator car carries riders between the floors of a multi-story building by moving up and down in the elevator shaft. Most elevators are automatic and do not have any personnel present to operate the elevator. Elevators have two doors, one fixed to the frame, mounted on the wall and one attached to the elevator car. When the elevator reaches the destination floor and parks, both of these doors open synchronously.

Fig. 5. Sample Side View Image.

Fig. 6. Sample Top View Image.

TABLE I. IMAGE DISTRIBUTION IN THE DATASET

View	Label	Number of Images
Side View	No Obstruction	4,128
Side View	Obstruction	3,204
Top View	No Obstruction	1,848
Top View	Obstruction	1,656

The proposed obstruction detection system will use 8 digital cameras as input devices. This includes three cameras mounted along each side of the elevator car door and two cameras mounted at the top of the elevator frame. Using multiple cameras located at various positions provides different views/perspectives of the obstruction passing through the elevator doors. This improves the effectiveness of the proposed system for detecting small or narrow objects which is a significant advantage over existing elevator door control technologies.

These cameras will capture images periodically every few milliseconds. These images are processed by the hardware and software system mounted in the elevator cars. Each image is passed to a pre-trained image classifier. The pre-trained model will perform binary image classification to determine whether there is an obstruction between the elevator doors. The proposed system will have two separate models for top view classification and side view classification respectively. By leveraging binary image classification using machine learning and deep learning, the proposed system will be able to detect obstructions with high accuracies in real-time.

If an obstruction is detected, then the signaling mechanism is triggered and a signal will be sent to the door controller to stop and reverse. Each camera has an LED light fastened next to it. If an obstruction is detected one of the images captured, the light next to the corresponding camera will turn on to notify the passengers about where the obstruction is detected. This also helps simplify the maintenance of the proposed system. If there is no obstruction detected, then the closing process will continue. The system continues to monitor for any obstructions until the gap between the elevator doors is too small to capture a digital picture. The methodology of the proposed system is illustrated in Fig. 7.

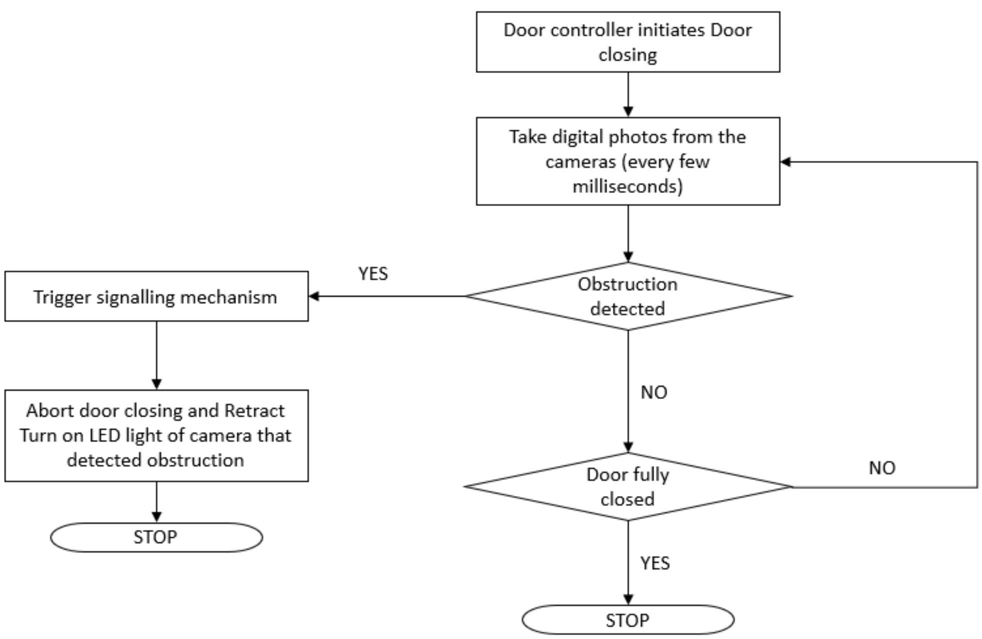


Fig. 7. Methodology of Proposed Elevator Obstruction Detection System.

C. Models

Several machine learning and deep learning models were built and evaluated to determine the most suitable model for the proposed system. These classifiers were chosen based on their ease of use, solid performance, and popularity in related works which leverage binary image classification for real-world applications.

1) Linear SVC: Linear Support Vector Classifier (SVC) is a supervised machine learning algorithm. It can be used for both regression and classification challenges. In this paper, Linear SVC is used for image classification. The data points are plotted in the n-dimensional space and the algorithm identifies the best fit hyperplane which divides the data into categories.

2) k-NN: The k-Nearest Neighbours (k-NN) classifier has proven to be an effective classification model on numerous occasions. This model uses the labelled training data to predict the output. The predicted value is determined as the majority in the 'k' nearest neighbours. For the k-NN classifiers in this paper, k=3.

3) Decision Tree: Decision Tree (DT) is an algorithm which has a flowchart-like tree structure. The leaves of the tree represent the target or outcome variables. In this case, the leaves can take the value 'Obstruction' or 'No Obstruction'. The decision tree classifiers implemented in this paper have a depth of 3.

4) Random forest: Random forest is a combination of multiple tree classifiers where each tree is based on independently sampled random vector values. Since it consists of multiple trees, it is called a forest. While this algorithm requires more computational power and time than decision trees, it helps improve the accuracy and reduce overfitting.

5) CNN: Convolutional neural network (CNN) is a type of deep learning algorithm which establishes learnable biases and weights to various aspects in an input image in order to differentiate one from another. The CNNs implemented in this paper have the classic architecture as a Sequential model with Convolution layers, Pooling layers and ReLU layers.

D. Evaluation

The speed and correctness of the predictions made by the classifier are crucial to determine which models would be suitable for the proposed system. The computation time of each model for test data prediction was calculated in seconds. Various metrics, namely accuracy, precision, recall and F1 score, were used to evaluate and compare the performance of the models.

In this paper, the binary classification task is defined such that each image is classified as either 'No obstruction' or 'Obstruction'. The True Positive (TP) attribute of the confusion matrix is defined as the number of images which contain an obstruction and were classified as 'Obstruction' by the model. Similarly True Negative (TN) is the number of images which contain no obstructions and were classified as 'No obstruction', False Positive (FP) is the number of images which contain no obstructions but were incorrectly classified as 'Obstruction and False Negative (FN) is the number of images which contain obstructions but were incorrectly classified as 'No obstruction'

These confusion matrix metrics were used to calculate accuracy, precision, recall and F1 score. Accuracy is the number of correctly classified images over the total number of images. Accuracy is calculated using all four of the aforementioned components of a confusion matrix (1).

$$\text{Accuracy} = \frac{TN+TP}{TN+FP+TP+FN} \quad (1)$$

Precision is the ratio of true positives to all positives. It gives greater importance to false positives. A precision value closer to 1 indicates that the model is performing well. Precision is calculated using the True Positive and False Positive attributes. (2).

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

Recall is a very important evaluation metric for the proposed system because it assigns more weightage to false negatives. A false negative in the proposed system means that the obstruction is present but it is not detected by the model. As a result, the doors will close normally and the person may be injured. Thus, it is crucial to minimize the number of false negatives. Recall is calculated using the True Positive and False Negative attributes. (3).

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

F1 score is defined as the weighted average of precision and recall (4). If the F1 score is closer to 1, it indicates that the classifier has performed well.

$$\text{F1 score} = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (4)$$

The models were built and evaluated separately for top view image classification and side view image classification.

IV. Result and Discussion

The models and algorithms elaborated in the previous section were implemented and compared. Table II provides a detailed summary of the performances of each model for top view image classification. It can be seen that the k-NN model performed the best out of all the classifiers with an F1 score of 0.994 and accuracy of 99.8%. This was followed closely by the CNN model which had an accuracy of 98.9%. Among the models implemented on the original image dataset, both the Linear SVC and Random Forest models performed remarkably well with the same accuracy of 98.6%. All of models performed well with respect to the considered evaluation metrics.

Table III provides a detailed summary of the performances of each model for side view image classification. It can be seen that the k-NN and CNN models performed the best with accuracies of 99.8% and 99.3% respectively. Among the models implemented on the original dataset, k-NN performed the best with an accuracy of 95.4%.

TABLE II. MODEL EVALUATION AND COMPARISON FOR TOP VIEW CLASSIFICATION

Model Name	**Dataset**	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Linear SVC	Top View images (Original)	0.986	0.99	0.98	0.99
k-NN	Top View images (Original)	0.973	0.98	0.97	0.97
Decision Tree	Top View images (Original)	0.945	0.94	0.94	0.94
Random Forest	Top View images (Original)	0.986	0.99	0.98	0.99
Linear SVC	Top View images (Augmented)	0.967	0.97	0.96	0.97
k-NN	Top View images (Augmented)	0.998	1.00	0.99	0.994
Decision Tree	Top View images (Augmented)	0.957	0.96	0.96	0.96
Random Forest	Top View images (Augmented)	0.977	0.98	0.98	0.98
Model Name	**Dataset**	**Accuracy**			
CNN	Top View images (Augmented)	0.98.9			

TABLE III. MODEL EVALUATION AND COMPARISON FOR SIDE VIEW CLASSIFICATION

Model Name	**Dataset**	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Linear SVC	Side View images (Original)	0.93	0.94	0.92	0.93
k-NN	Side View images (Original)	0.954	0.96	0.95	0.95
Decision Tree	Side View images (Original)	0.797	0.84	0.78	0.78
Random Forest	Side View images (Original)	0.856	0.90	0.84	0.85
Linear SVC	Side View images (Augmented)	0.967	0.97	0.96	0.97
k-NN	Side View images (Augmented)	0.998	1.00	0.997	0.998
Decision Tree	Side View images (Augmented)	0.859	0.88	0.85	0.85
Random Forest	Side View images (Augmented)	0.889	0.91	0.88	0.88
Model Name	**Dataset**	**Accuracy**			
CNN	Side View images (Augmented)	0.993			

The speed of each classifier was evaluated and compared in seconds as shown in Fig. 8 to 11. The speed is the computation time for classifying the test data. This is an important metric for the proposed system since the classifier must be able to identify whether or not an obstruction is present in the captured image in real-time. The computation times of the models were measured for top view classification and side view classification separately. Top view classification refers to classifying the images captured by the two cameras positioned at the top of the elevator. Side view classification refers to classifying the images captured by the 6 cameras positioned along the sides of the elevator car door. The speed of each model was also measured separately for the original dataset and augmented dataset. It can be seen that the Decision Tree model consistently performed with the lowest computation time followed by the Random Forest model. However, there is a trade-off between speed and correctness of predictions as the Decision Tree classifier had the poorest accuracy among the models considered.

In this study, a novel system design which leverages state-of-the-art machine learning techniques to detect obstructions between elevator car doors is proposed. Furthermore, an end-to-end analysis of its feasibility is presented which includes construction of a dataset of over 10,000 images, implementation of several modern binary image classifiers, and a thorough evaluation of model performances. This work advances the literature on improving elevator safety. The proposed system is designed to address the shortcomings of current elevator technology [2] by facilitating more effective detection of small and narrow obstructions through real-time image input. The use of machine learning techniques for elevator door control is – to the best of our knowledge – new. In contrast, a majority of existing works on improving elevator safety focus on detecting [9] [13] [14] [15] [20] and predicting [7] [8] elevator failure. Certain studies [6] focus on detecting only moving obstructions between elevator car doors whereas, the proposed system facilitates the detection of both moving and stationary obstructions.

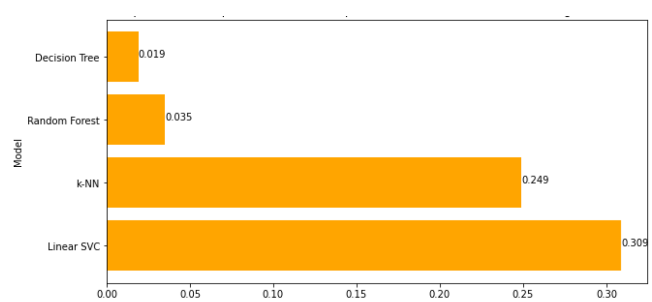


Fig. 8. Comparison of Computation Times for Top View Classification of Test Data (Original Dataset).

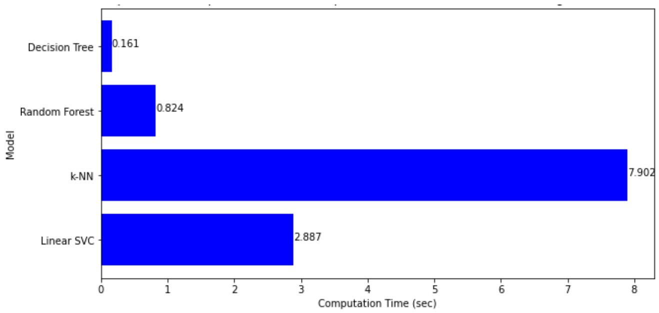


Fig. 9. Comparison of Computation Times for Top View Classification of Test Data (Augmented Dataset).

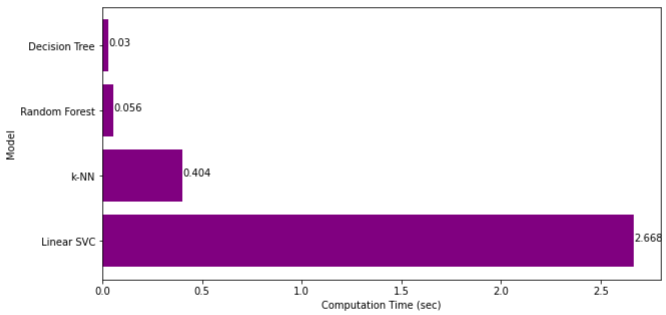


Fig. 10. Comparison of Computation Times for Side View Classification of Test Data (Original Dataset).

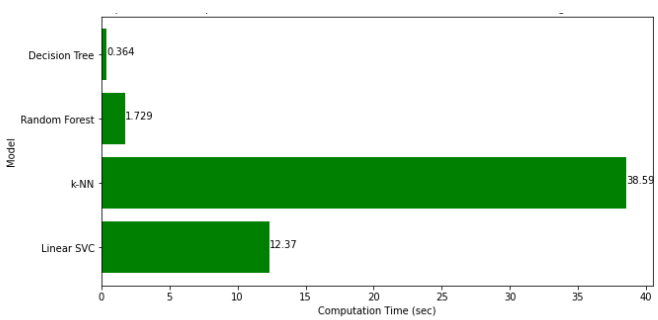


Fig. 11. Comparison of Computation Times for Side View Classification of Test Data (Augmented Dataset).

V. Conclusion and Future Work

Despite the widespread interest in the development of Smart Cities and Smart Buildings, very little advancement has been made in the technology used for elevator door operators. This paper proposes a novel elevator obstruction detection system that leverages modern technology such as binary image classification models, to improve the safety of elevators by effectively detecting smaller or narrower obstructions. This paper presents findings based on the evaluation and comparison of different binary image classification techniques for the purpose of obstruction detection. The considered models include Linear SVC, k-NN, Decision Tree, Random Forest, and CNN. A dataset was created through simulation of the proposed system using a functional elevator in a multi-story building. Brightness augmentation was used to improve the robustness of the models in different lighting conditions. The results of this paper show that it is feasible to develop the necessary models to perform image classification for the proposed system with high accuracies in real-time. All of the models performed well with respect to the considered evaluation metrics. The k-NN and CNN models had the highest accuracies for both datasets. However, the k-NN model had one of the highest computation times, indicating that the CNN model may be more suitable for the proposed system. While the Decision Tree model had the lowest computation time, this model also resulted in the lowest accuracy when compared with the performances of the other classifiers considered.

The work presented in this study can be extended to include the storage of the images captured to further improve the performance of the model. The use of different machine learning and image processing algorithms can be explored. Additional data can be collected to increase the size and improve the quality of the dataset.

Improving the safety and reliability of elevators presents itself as an underexplored research area with the potential to have a positive impact on millions of lives around the world. With the continuous advancement in the field of machine learning, computer vision and artificial intelligence, the scope of this paper is vast. The results of this paper can be used to improve the safety of elevators by leveraging the efficacy and immense capabilities of modern technology.

Acknowledgment

The authors would like to thank the institution Vellore Institute of Technology, Chennai for its constant support and for providing a conducive learning environment.

References

[1] A. Shrestha, “Safety considerations for the design of modern elevator systems,” University of Mississippi, 2019.

[2] M.S. Shivankar, H.P. Kasturiwale, D.T Ingole, “Development of industrial door operation system for elevator actuated by various drives,” Int. J. Sci. Eng. Res. 2012.

[3] J. O'Neil, G. Steele, C. Huisingh, G. Smith, “Elevator-related injuries to children in the United States, 1990 Through 2004,” Clin. Pediatr. 2007, 46(7), pp. 619-625.

[4] F. ShuangChang, C. Jie, Z. Yanbin, L. Zheyi, “Discussion on improving safety in elevator management,” 2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI). 2020.

[5] S. Magota, S. Kunimatsu, G. Yamamoto, R. Otsubo, T. Fujii, “The opening-and-closing speed control of an elevator door by frequency-shaping ILQ design method,” Transactions of the Institute of Systems, Control and Information Engineers. 2007, 20(8), pp. 331-337.

[6] X. Zeng, G. Zhao, Y. Wang, C. Wang, J. Wang, “Research on the elevator door control system based on the image processing technology,” 2010 International Conference on Electrical and Control Engineering, 2010.

[7] J. Hou, R. Qiu, J. Xue, C. Wang, X. Jiang, “Failure prediction of elevator running system based on knowledge graph,” Proceedings of the 3rd International Conference on Data Science and Information Technology. 2020.

[8] P. Wen, M. Zhi, G. Zhang, S. Li, “Fault prediction of elevator door system based on PSO-BP neural network,” Engineering. 2016, 08(11), pp. 761-766.

[9] Y. Guo, Y. Liu, X. Zhang, G. Wang, “The real-time elevator monitoring system based on multi-sensor fusion,” Journal of Physics: Conference Series. 2021, 2010(1), 012182.

[10] N. Amruthnath, T. Gupta, “A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance,” 2018 5th International Conference on Industrial Engineering and Applications (ICIEA). 2018.

[11] C. Hsu, Y. Qiao, C. Wang, S. Chen, “Machine learning modeling for failure detection of elevator Doors by three-dimensional video monitoring,” IEEE Access. 2020, 8, pp. 211595-211609.

[12] Z. Wan, S. Yi, K. Li, R. Tao, M. Gou, X. Li and S. Guo, “Diagnosis of elevator faults with LS-SVM based on optimization by K-CV,” J. Electr. Comput. Eng. 2015, pp. 1-8.

[13] K. Mishra, K. Huhtala, “Elevator fault detection using profile extraction and deep autoencoder feature extraction for acceleration and magnetic signals,” Appl. Sci. 2019, 9(15), p. 2990.

[14] J. Yu, B. Hu, “Influence of the combination of big data technology on the Spark platform with deep learning on elevator safety monitoring efficiency,” PLOS ONE. 2020, 15(6):e0234824.

[15] X. Yi, “Design of elevator monitoring platform on big data,” 2016 International Conference on Industrial Informatics - Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII). 2016.

[16] M. Zihan, H. Shaoyi, Z. Zhanbin, X. Shuang, “Elevator safety monitoring system based on internet of things,” Int. J. Eng. 2018, 14(08), p.121.

[17] I. Olalere, M. Dewa, “Early fault detection of elevators using remote condition monitoring through IoT technology,” S. Afr. J. Ind. Eng. 2018, 29(4).

[18] S. Niu, H. Liu, G. Chen, H. Zhang, “The research and development of preventing the accidental movement of the elevator car safety protection device,” Lecture Notes in Electrical Engineering. 2018, pp. 239-244.

[19] L. Kong, L. Zhao, S. Wang, J. Jiang, H. Wu, “Application discussion and inspection on the unintended car movement protection system,” Journal of Physics: Conference Series. 2020, 1549(4):042008.

[20] L. Chen, S. Lan, S. Jiang, “Elevators fault diagnosis based on artificial intelligence,” Journal of Physics: Conference Series. 2019, 1345:042024.

[21] G. Wang, C. Han, C. Zhang, Y. Xu, M. Zhang, “Research of elevator security data mining ased on Hadoop,” 2019 Chinese Control And Decision Conference (CCDC). 2019.

The SMILE, A Cyber Pedagogy based Learning Management System Models

Paula Dewanti[1]*, I Made Candiasa[2], I Made Tegeh[3], I Gde Wawan Sudatha[4]
Postgraduate, Education Science, Universitas Pendidikan Ganesha, Singaraja, Indonesia[1]
Department of Information Systems, Institut Teknologi dan Bisnis STIKOM Bali[1]
Department of Mathematics, Universitas Pendidikan Ganesha, Singaraja, Indonesia[2]
Department of Educational Technology, Universitas Pendidikan Ganesha, Singaraja, Indonesia[3]
Department of Educational Technology, Universitas Pendidikan Ganesha, Singaraja, Indonesia[4]

***Abstract*—The study's purpose is to create an LMS model that is adapted to the characteristics of university students to enhance the learning experience by utilizing various multidimensional learning resources in Cyber Pedagogy. This research and development study used the Analyze, Design, Develop, Implement, and Evaluate (ADDIE) instructional design framework as well as the Waterfall system development model to develop learning materials and infrastructure. The study involves 50 students from the Bali Institute of Technology and Business, as well as five lecturers and six judges at the expert test stage, namely learning media experts, learning design experts, and teaching experts, who were chosen through purposive sampling. The SMILE Model (Simple, Multidimensional, and Interactive Learning Ecosystem) is designed to meet the learning needs and expectations of today's largest market share of higher education, the millennial generation. The SMILE Model was developed successfully with ongoing assistance from researchers' students, particularly in the E-Tourism course. The implementation is accomplished through the combination of university E-Learning and the use of Microsoft Teams as a virtual learning platform alternative. During the COVID-19 pandemic, this was considered the new face-to-face norm.**

Keywords*—*ADDIE; cyber pedagogy; e-learning; higher education; learning management system

I. Introduction

The purpose of this research is to develop a web-based Learning Management System (LMS) model which is termed "SMILE" (Simple, Multidimensional, and Interactive Learning Ecosystem). SMILE can be accessed by utilizing a variety of browsers on mobile devices and desktop computers. SMILE is expected to help students understand learning in a more entertaining way, as well as provide a different approach to millennial generation learning activities to improve their performance.

The SMILE consolidates constructivism learning theory as well as cyber pedagogy. The principles of constructivism theory that underlie the SMILE model include how students construct their own knowledge, with the readiness of knowledge gained through real-world experiences, collaborative activities, reflection, and interpretation, with activities that allow students to have different understandings of knowledge and perspectives in interpreting it. Despite the fact that the SMILE model incorporates Cyber Pedagogy principles, each student has completed independence in accessing multiple learning resources from any location, at any time, and in a variety of ways. Learning occurs in a multi-directional, egalitarian, inclusive, and non-bureaucratic manner, with interaction and feedback made available only when requested, both formally and informally, and students have complete control over the learning and teaching processes [1]. There are three actors (users) of the SMILE application who have varying levels of access, namely administrators, lecturers, and students. They will get a username and password to access the SMILE application and its features online.

Pyöriä, Ojala, Saari, & Järvinen discovered that the Millennial Generation, which is the current market share of higher education, may not be as homogeneous in fundamental learning strategies and attitudes as is frequently proposed; regardless of their general character traits in learning, there is no evidence that these traits affect their fundamental learning process. Many curricular strategies have been implemented to address alleged changes in the learning styles of Millennial students. Nothing has clearly demonstrated superior results in academic achievement or skill development for graduates, and there are concerns about Millennial student engagement in the learning process [2]. Furthermore, Conklin's research emphasizes the importance of student experience by establishing classrooms that encourage millennial learner autonomy and teachers who value their students' perspectives and experiences [3]. Problem-Based Learning (PBL), Student-Centered Learning (SCL), and Classroom as Organizations (CAO) all require learners to recognize and respect the ideas, thoughts, feelings, and relationships of other learners during the learning process by participating in learning experiences that other learners have made available to them [3]. Self-evaluation of their ideas may be considered. Learning experiences that are not conducive can undermine our efforts to stimulate students.

II. Literature Review

A. Cyber Pedagogy

"Cyber pedagogy" refers to the science or art of teaching in an online environment. Cyber Pedagogy focuses on motivations and teaching methods that are appropriate for the technology being used, rather than those that are best suited to the traditional face-to-face classroom context [1]. The Principles of Cyber Pedagogy include the following:

*Corresponding Author.

- Each participant has access to an extensive set of resources that promote learning as an individual in a 24-hour environment.
- Procedures toward the source of learning are common, with many of them taking place solely, and in different ways.
- Internet resources are available dynamically and are constantly updated.
- The learning process is naturally spontaneous, contextual, and critical.
- Learning can come in different forms in an inclusive, equitable, and bureaucracy-free environment.
- Feedback and interaction are done formally and informally only when requested.
- Formal evaluations can be obtained at any time and in a number of ways.
- Participants in the learning process have complete control over the teaching and learning processes.

Indrajit also stated that universities should implement student-centered learning, where knowledge transfer is focused on students as part of Cyber Pedagogy and students have complete control over the learning process. In order for it to work, students must construct and reconstruct knowledge. When learning is integrated into an activity, it is most effective, and student experience produces outcomes that significantly improve learning. Student mentors direct, guide, and monitor students as they use all available Internet learning resources to ensure proper and optimal use [1].

B. Learning Management System in Higher Education

Each university's learning management system (LMS) is tailored to its specific needs. LMS development has made use of free (open source) production sites like Moodle, Sakai, and others, as well as systems developed by institutions themselves according to their own concepts. To name a few, below are some examples [4].

- Andrew Ng and Daphne Koller, both computer science professors at Stanford, founded Coursera, an online learning platform, in 2012. Coursera offers Massive Open Online Courses (MOOC), specialization programs, and degree programs in a variety of subjects in collaboration with universities and other organizations [5].
- edX, founded by Harvard University and MIT, is also one of the leading providers of open online courses (MOOCs). Similar to Coursera, edX is a global online course provider focused on students in higher education from a variety of disciplines. edX also offers a number of free courses [6].
- The Directorate General of Learning and Student Affairs of the Indonesian Ministry of Research, Technology, and Higher Education manages the Indonesian Online Learning System, known as SPADA Indonesia, which aims to promote equitable access to quality learning in higher education [7]. SPADA Indonesia enables students from one university to take specific quality courses from other universities, and their learning outcomes are recognized equally by the college where they are enrolled [7]. Furthermore, Arifin contended that SPADA Indonesia was formed to address some of higher education's challenges, such as the limited capacity of higher education institutions, the low affordability of universities due to uneven distribution, and the fact that many universities still lack adequate and high-quality educational resources.
- The SEAMOLEC Open Learning Center is presented by the Southeast Asian Ministers of Education Organization's (SEAMEO) education research and development center. SEAMOLEC focuses on open and distance education, training, and the exchange of knowledge and resources within and beyond Southeast Asia. The mission of SEAMOLEC is to establish itself as a center of expertise of open and distance education, as well as to assist SEAMEO member countries in identifying educational problems and solutions in human resource development through open and distance education systems [8].

The Learning Management System (LMS) developed is termed SMILE (Simple, Multi-dimensional, and Interactive Learning Ecosystem). It is a Web-based application that can be accessed on mobile devices and computers using various browsers. SMILE is expected to help students understand learning in a more fun way and provide a different environment and approach to maximize student potential.

III. Research Methods

This study utilized the ADDIE instructional design framework, which covers Analysis, Design, Development, Implementation, and Evaluation [9], and the Waterfall system development model to develop learning materials and infrastructure that support students' learning. The Waterfall model, such as the ADDIE model, uses a structured framework for design that relies on completing each stage before going on to the next one. Requirement Analysis, System Design, Implementation, Testing, Deployment, and Maintenance are all part of the process [10]. By combining the two models, the designer's instructional team can perform more simple steps by formulating a cohesive model.

IV. Result and Discussion

A. ADDIE Design Development

1) ADDIE Model: According to Eller, the ADDIE Model has a lengthy history dating back to the establishment of teaching system designs and has been continuously improved by a number of scholars and instructional designers. Although structural design and software design are two separate areas, they contain many similarities and intersect in ways that instructional designers, subject matter experts, and educators who utilize the final version might benefit from [9]. Fig. 1 illustrates the ADDIE Model of SMILE.

The Waterfall Method is used in the development of SMILE LMS, so as to organize and integrate all aspects of learning materials and activities. From there, it can be seen that the Waterfall Model intersects with the ADDIE Model. The Waterfall steps are indicated in Fig. 2.

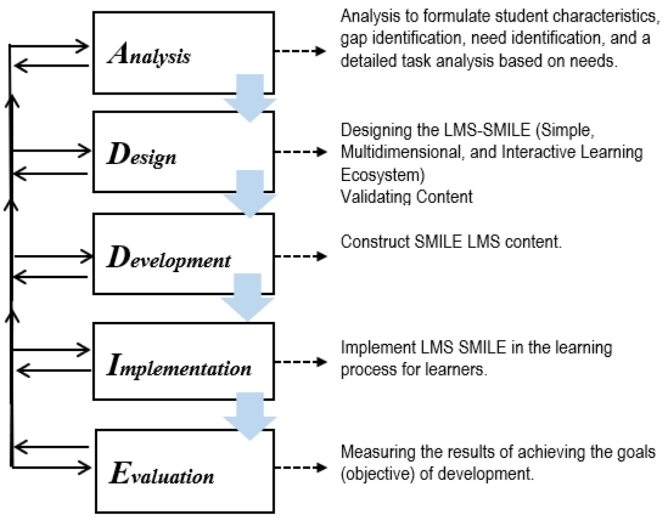


Fig. 1. ADDIE Model of SMILE.

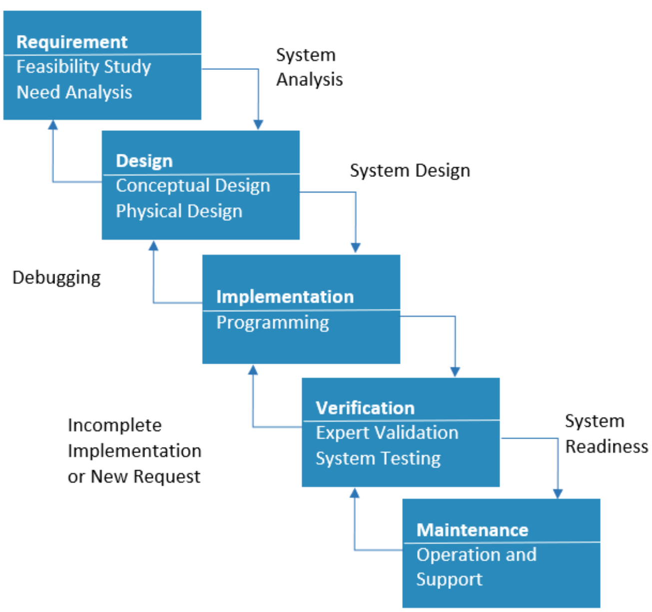


Fig. 2. Waterfall Model.

2) The waterfall model

a) Requirements: The needs analysis was included in the first stage of the waterfall model, and it involves both practical and non-functional requirements for the development of the SMILE LMS.

b) Design: The second stage is the design stage, in which the previous stage's system specification was explored and the system design was indeed prepared. In system design, Flow Diagrams, conceptual databases, and Entity Relationship Diagrams (ERD) are used to help with the overall system architecture concept. Fig. 3 describes an overview of the SMILE Design.

c) Implementation: Programming or coding is performed during the system's implementation stage, which is the design's interpretation into a computer-readable language. In addition, the generated program items will be subjected to verification.

d) Testing: At this stage, the system testing method of choice is Black Box Testing. It is a test method that involves observing the output results of data or conditions entered into the system without seeing how the data management process works. It is performed from the perspective of the user, enabling the identification of existing problems that can be resolved later.

e) Deployment: The SMILE is deployed into a live environment to be tested before being made available to users during this phase. This phase also includes user training programs to help users know how the system works.

f) Maintenance: This step involves providing support and maintenance for the SMILE to ensure that it runs smoothly. The main goal of this stage is to fix any errors, defects, or bugs that users encounter while using the Smile. Maintenance tasks include all processes needed to manage system continuity, enhancement, and improvement.

B. The SMILE Model

The LMS design framework is generally based on the ADDIE learning design model and the Waterfall system development model, which are used as guidance in designing learning materials and infrastructure that support student learning processes. The waterfall model is a method used in software development. The stages begin with analysis, which then proceeds to the stages of design, coding, testing, and support [11]. It can be seen that the waterfall model intersects with the ADDIE model. The ERD of SMILE is shown in Fig. 4, while the conceptual database SMILE is shown in Fig. 5.

Fig. 6 to 11 depict a few of the SMILE user interfaces, including the Login page, User Profile page, Admin's dashboard, Student's Detail page, Lesson Plan page, and the Syllabus page.

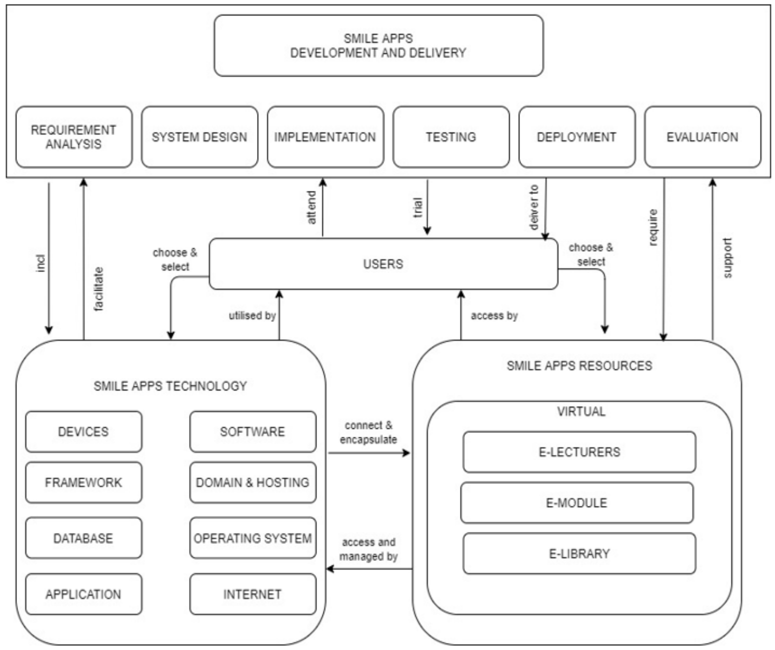


Fig. 3. SMILE Design.

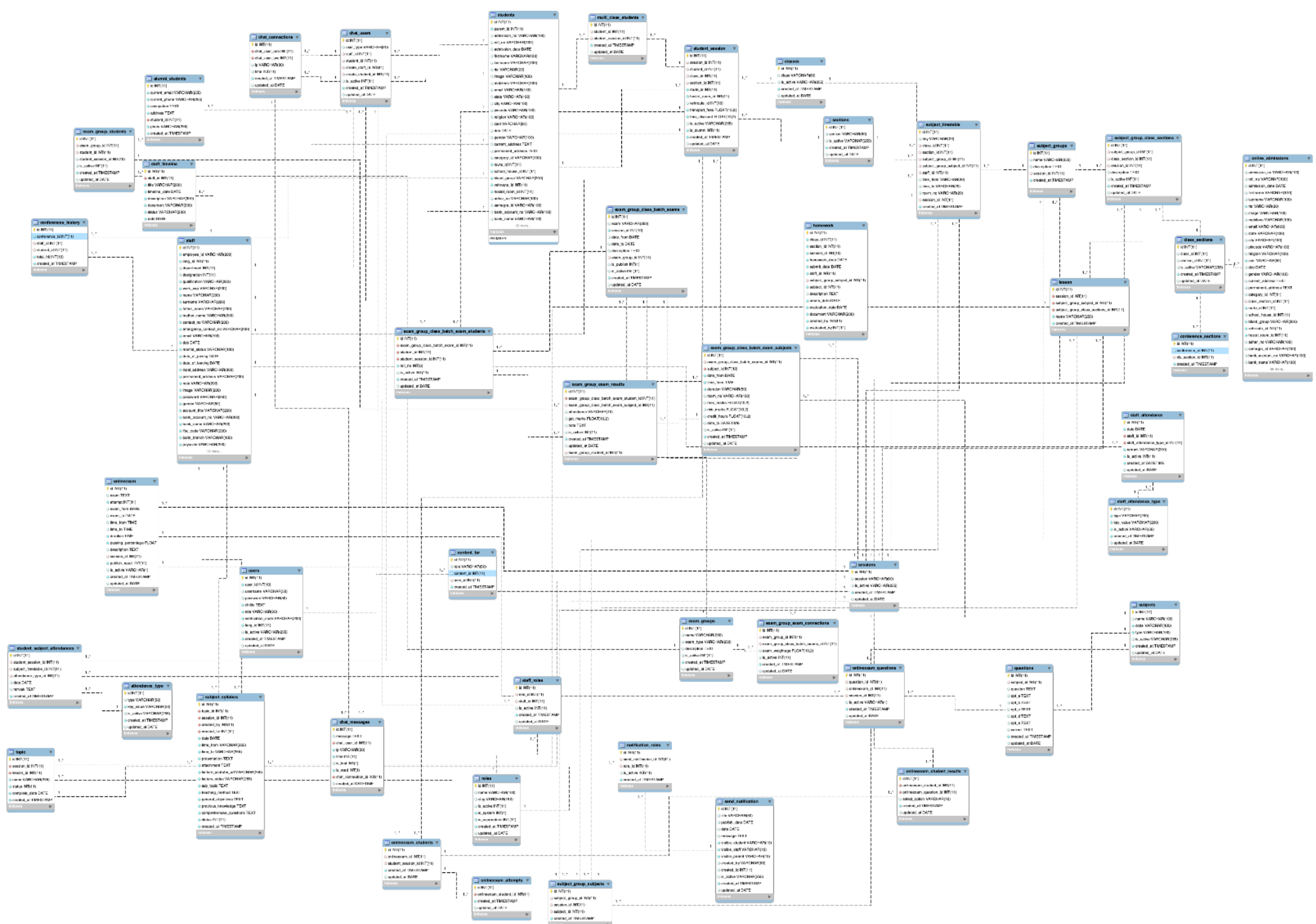


Fig. 4. SMILE's Entity Relationship Diagrams.

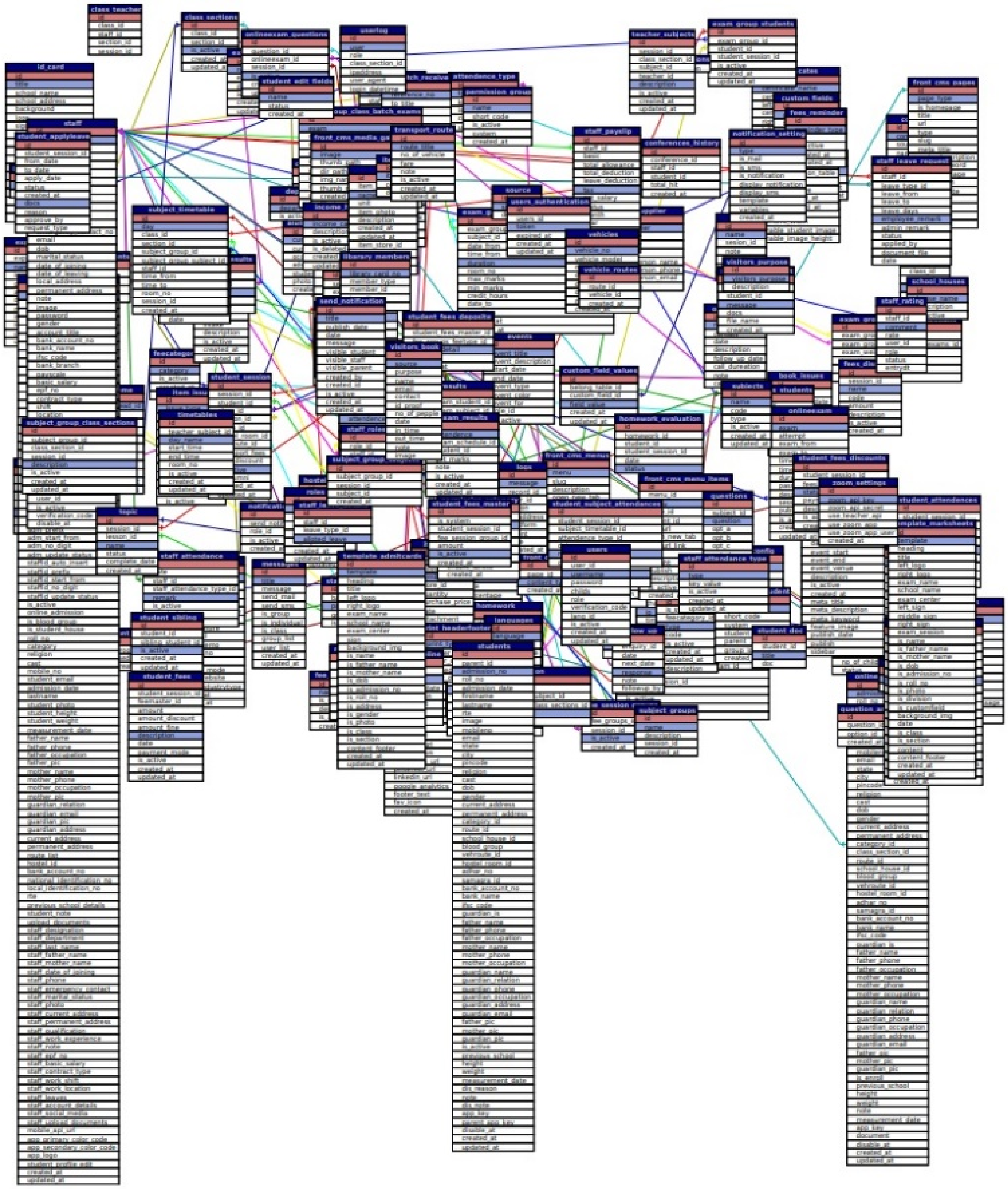


Fig. 5. SMILE Conceptual Database.

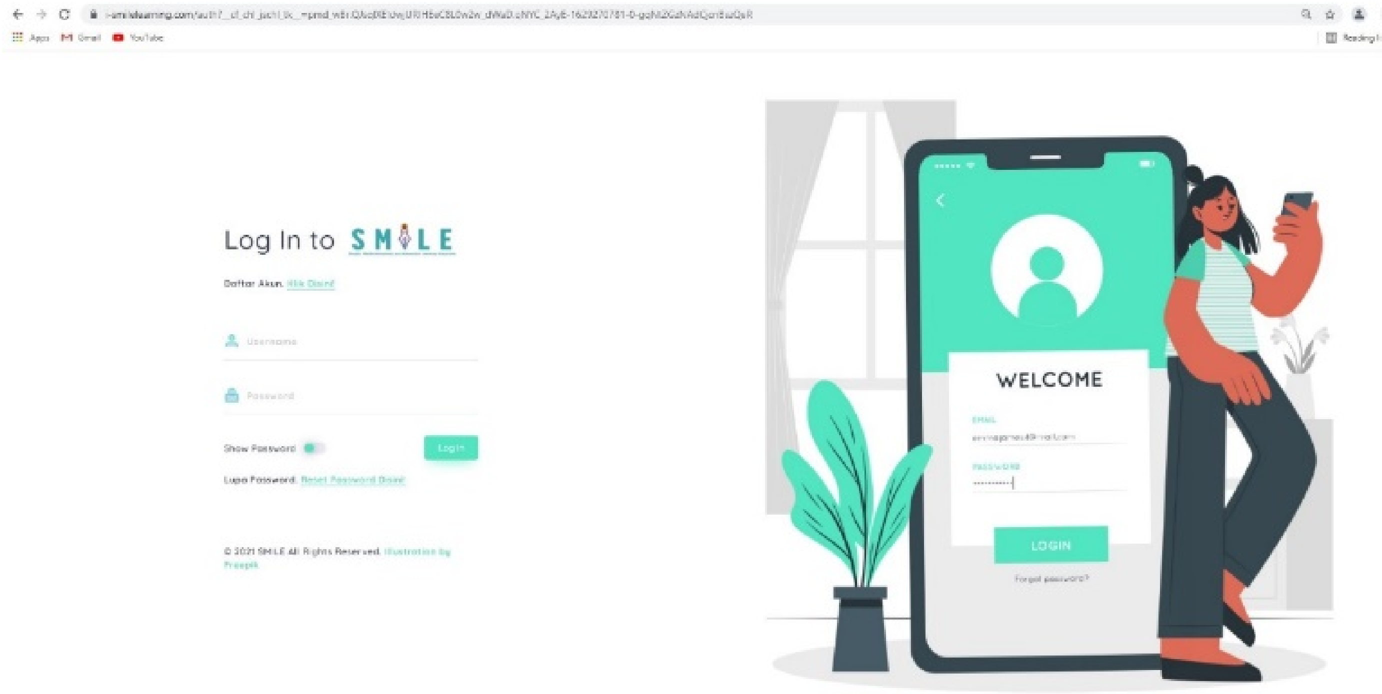


Fig. 6. Login Form for SMILE.

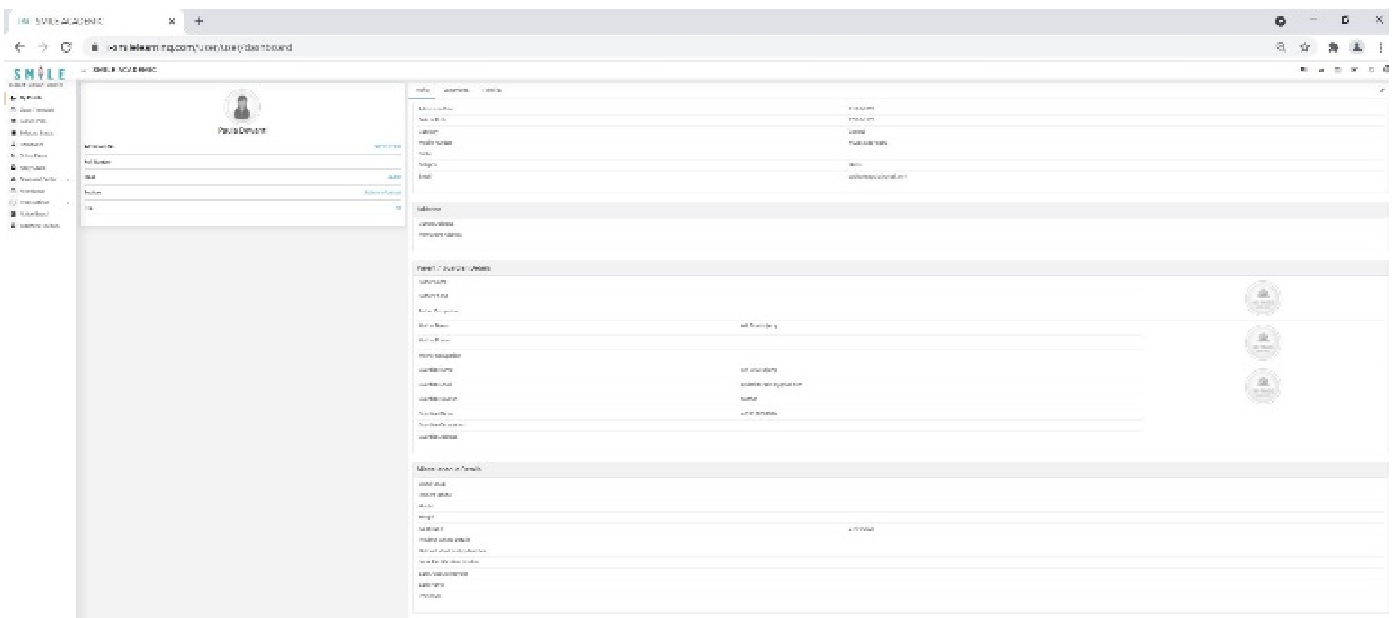


Fig. 7. User Profile Page for SMILE.

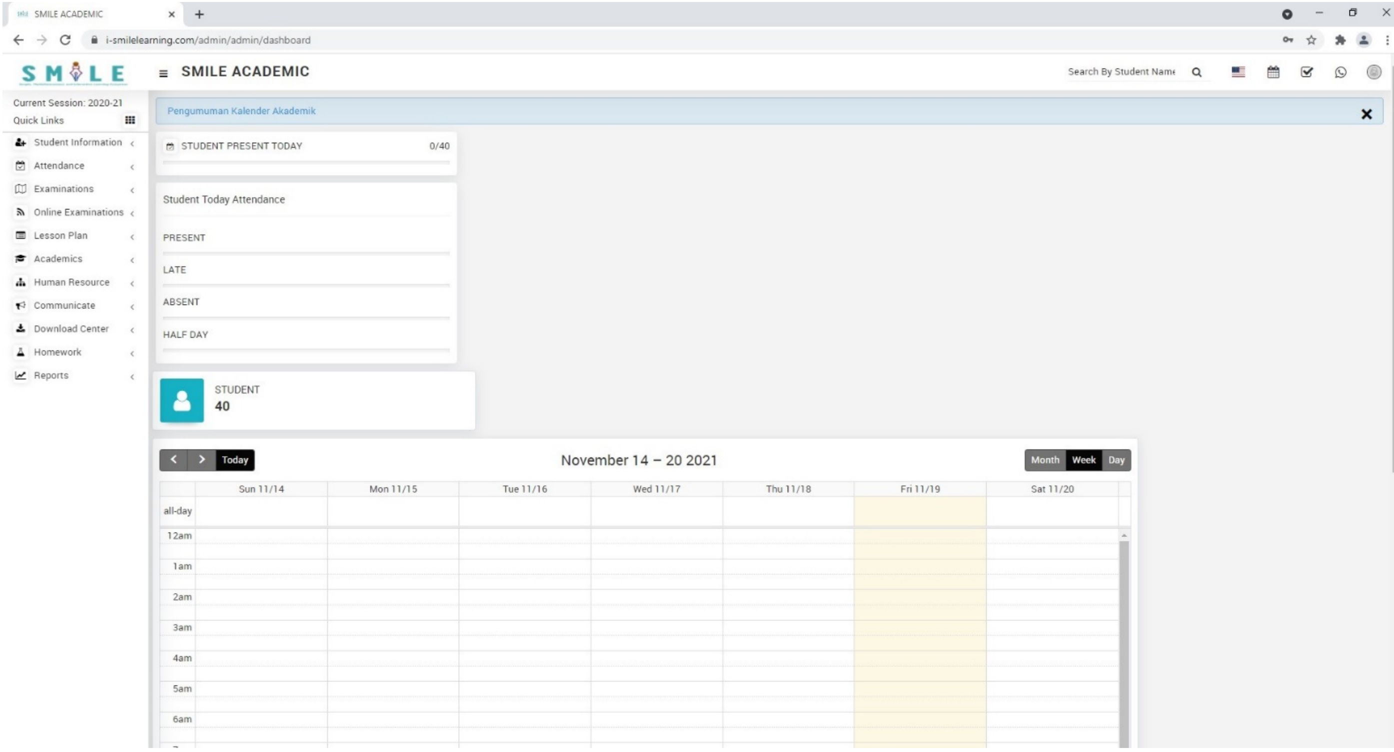


Fig. 8. Admin Dashboard Page for SMILE.

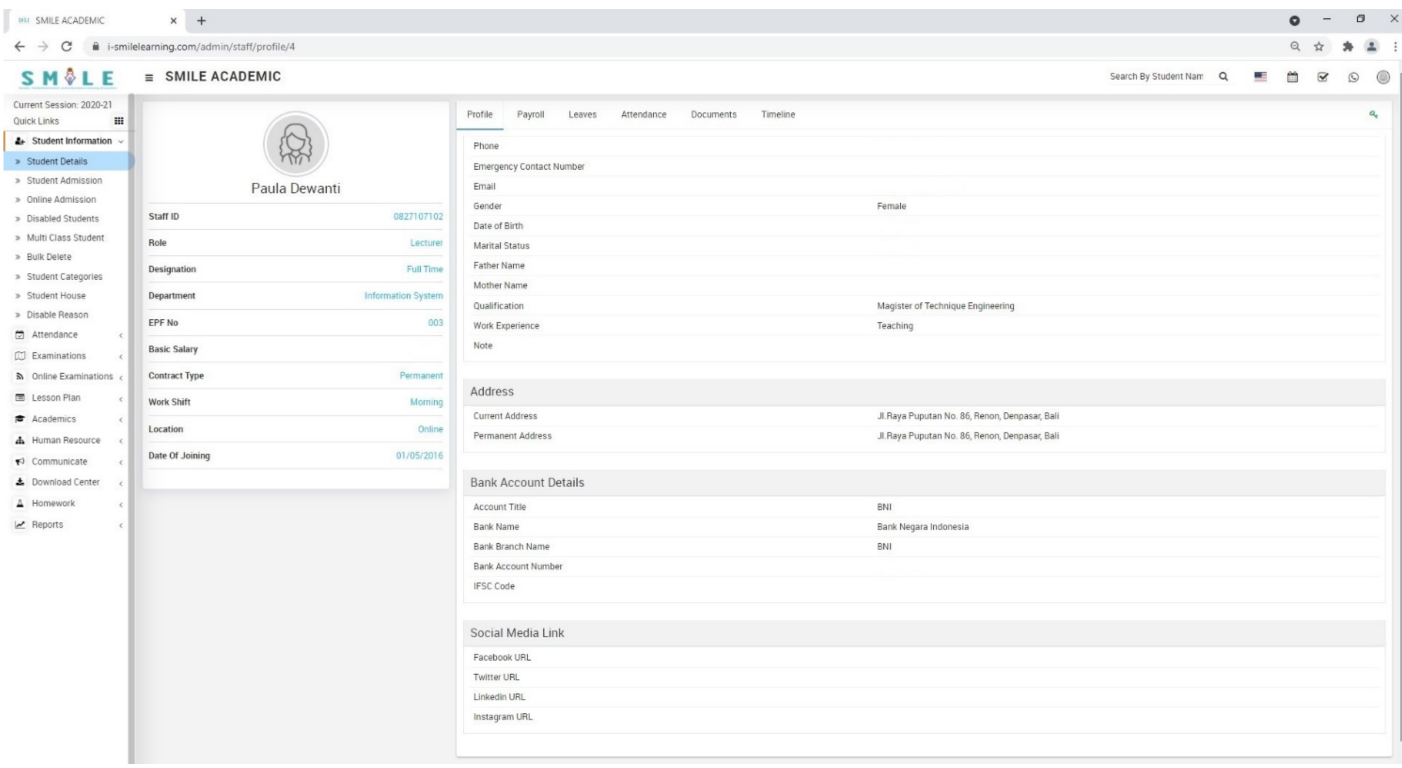


Fig. 9. Student Details Page for SMILE.

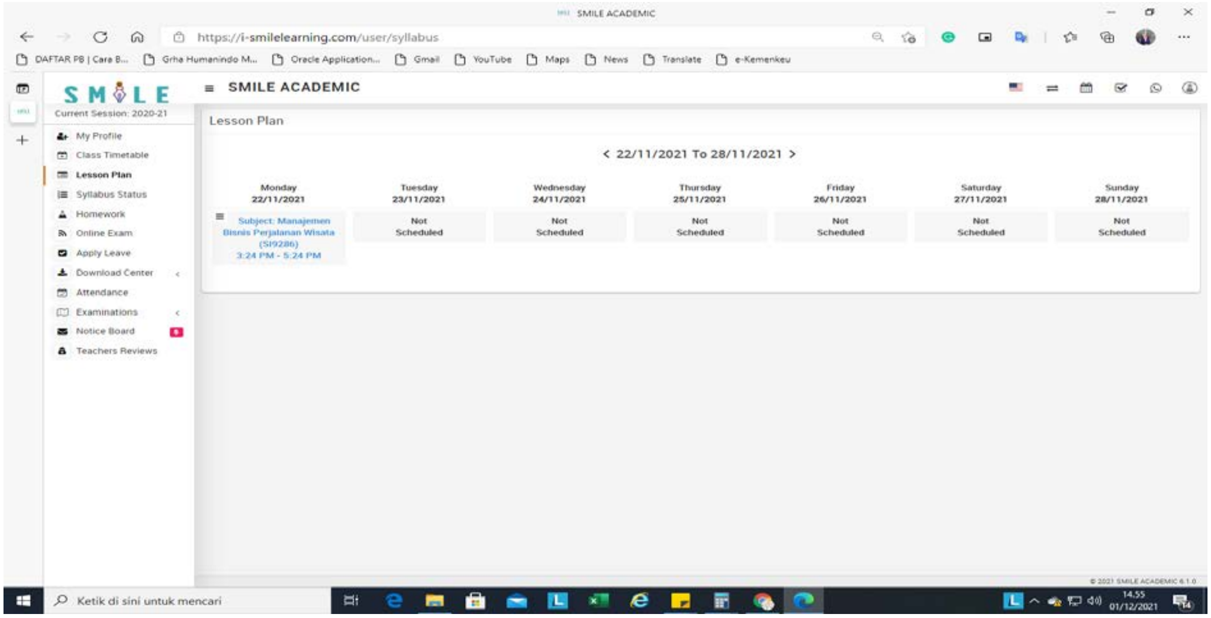


Fig. 10. Lesson Plan Page.

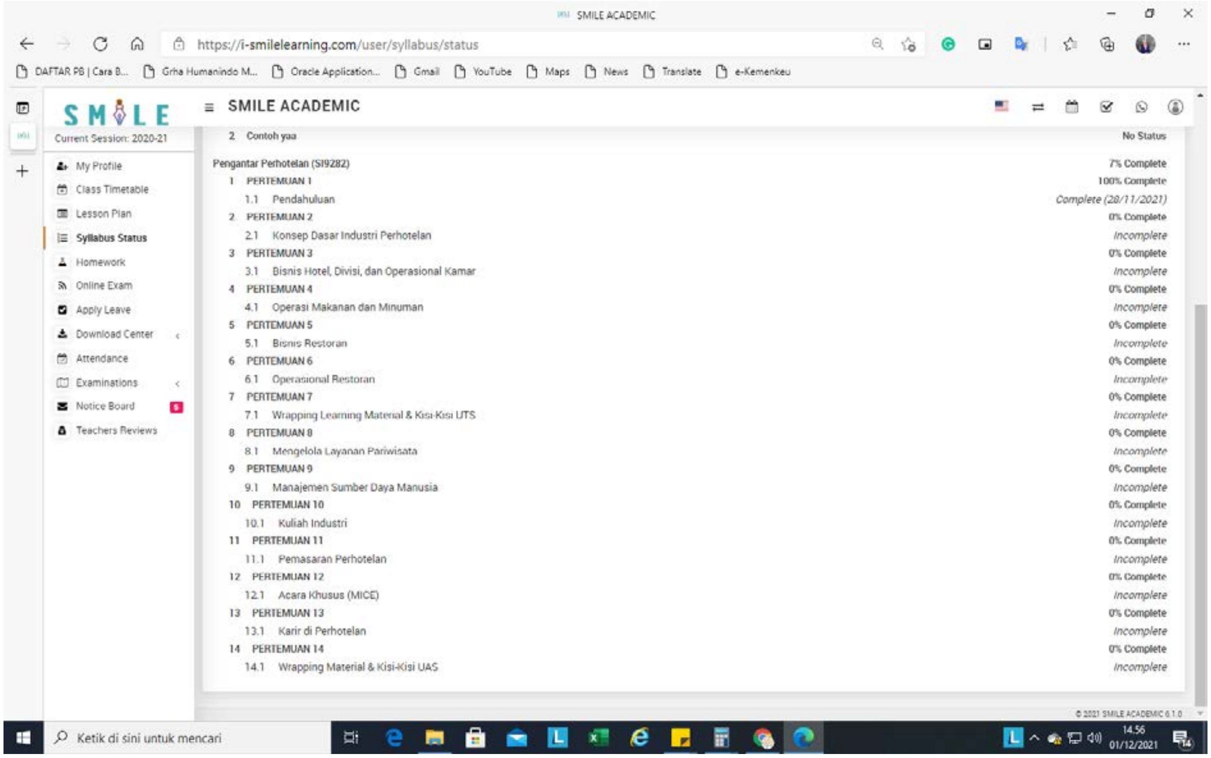


Fig. 11. Syllabus Page.

C. Experimental Findings

Three material experts and three media experts used an evaluation sheet to assess the content and media validity of SMILE, which was then tested using Gregory's cross tabulation. The evaluation results show that the SMILE Learning Management System meets the requirements for validity. The system's usability was evaluated using questionnaires distributed to students and teachers, as well as observation sheets. The results of the usability tests show that the developed SMILE Learning Management System meets the expectations of both students and teachers.

1) Expert findings: Tables I and II list the instruments and indicators that were measured.

The following steps were taken to obtain the validity data from the assessment by material experts and media experts [13].

TABLE I. LEARNING OBJECT REVIEW INSTRUMENT V. 1.4 [12]

No.	Rated Aspects
1	Content of Quality
	It is necessary to have veracity, accuracy, a well-balanced presentation of ideas, and an adequate level of detail.
2	Aligning Learning Objectives
	Learning objectives, activities, assessments, and student characteristics should all be interconnected.
3	Adaptation and Feedback
	Variability learner input or learner modeling motivates adaptive content or feedback.
4	Motivation
	Abilities to inspire and evoke interest or curiosity.
5	Design of Presentation
	Creating visual and auditory information to aid in learning and mental processing.
6	Usability of Interaction
	The user interface's predictability, ease of navigation, and the quality of the UI that supports features.
7	Accessibility
	Provide learners' support.
8	Reusability
	The ability to link various courses or learning contexts without modifying them.
9	Comply with Standards
	In accordance with international standards and requirements.

TABLE II. CYBER PEDAGOGY INDICATORS [1]

No.	Variable	Indicators
1	Mobility	The process method for learning resources is ubiquitous and can be from anywhere, anytime, and in various ways.
2	Location Awareness	The technology component is capable of providing information about the physical location of the device to other users or applications.
3	Interoperability	Users have the ability to exchange and use information. Learning takes place in a multi-way, egalitarian, inclusive, and non-bureaucratic way. Applications can interact with other applications through a mutually agreed protocol over various communication lines.
4	Seamlessness	Every student can access various learning resources freely and independently 24/7. Internet resources are dynamic, and are updated collectively every second.
5	Situation Awareness	The process of learning is intuitive, contextual, and critical in nature. a) Emotional self-awareness. b) Assertiveness. c) Self-esteem. d) Independence. e) Self-actualization.
6	Social Awareness	a) Effective communication. b) Effective learning.
7	Adaptability	Interaction and feedback can take place when and where it is desired–formally or informally. Evaluation of the learning procedure can be performed as many times as necessary.
8	Pervasiveness	Learners have full control over the learning and learning process.

a) To convert qualitative data to quantitative data, apply the Gregory content validity provisions as shown in Table III.

TABLE III. CROSS TABULATION OF THREE RATERS

Rater I	Irrelevant				Relevant			
Rater II	Irrelevant		Relevant		Irrelevant		Relevant	
Rater III	Ir-relevant	Relevant	Ir-relevant	Relevant	Ir-relevant	Relevant	Ir-relevant	Relevant
	A	B	C	D	E	F	G	H

[a.] Notes of the Three Raters' Cross Tabulation
A = The three raters disagree.
B = Rater I disagree, Rater II disagree, Rater II agree.
C = Rater I disagree, Rater II agree, Rater III disagree.
D = Rater I disagree, Rater II agree, Rater III agree.
E = Rater I agree, Rater II disagree, Rater III disagree.
F = Rater I agree, Rater II disagree, Rater III agree.
G = Rater I agree, Rater II disagree, Rater III agree.
H = The Three Raters are Agree

b) The Gregory Formula is then used to calculate the score to generate the content validity.

$$\text{Vi} = \frac{H}{(A+B+C+D+E+F+G+H)} \quad (1)$$

where,

Vi = Content Validity

c) Convert the average score to a qualitative value using the assessment factors listed in Table IV below.

TABLE IV. GUILFORD'S TABLE OF RELIABILITY COEFFICIENTS

Score Range	Classification
$\bar{X} > 0{,}80$	Very High
$0{,}60 < \bar{X} \leq 0{,}80$	High
$0{,}40 < \bar{X} \leq 0{,}60$	Moderate
$0{,}20 < \bar{X} \leq 0{,}40$	Low
$\bar{X} \leq 0{,}20$	Very Low

The SMILE LMS instrument has a coefficient of content validity of 1, which is classified as very high based on the conversion of the Guilford table.

2) User Usability Analysis: The Usability Testing method is applied using Nielsen's approach, which is classified into five instruments: learnability, memorability, efficiency, errors, and satisfaction. Using the 5-point Lickert Scale, from strongly disagreeing (1) to strongly agreeing (5). Tabel V shows the result from the lecturer's perspective.

TABLE V. USABILITY ANALYSIS

No.	Assessment Aspects	Likert Scale				
		1	2	3	4	5

A.	**Learnability**					
A1	Learning Management System SMILE can be studied easily.				33.3	66.7
A2	I am pleased with the ease and speed with which information is accepted in detail, as well as with the SMILE Learning Management System in particular.				50.0	50.0
A3	I can easily comprehend the content and information presented in the SMILE Learning Management System.			16.7	33.3	50.0
A4	I can easily understand and comprehend plots from the SMILE Learning Management System's navigation.				50.0	50.0
A5	I was able to learn how to use the SMILE Learning Management System without any written instructions or a manual book.				50.0	50.0
B.	**Memorability**					
B1	I have no trouble recalling how to use the SMILE Learning Management System.				50.0	50.0
B2	I can easily find and remember navigation directions and features on the SMILE Learning Management System.			16.65	16.7	66.65
B3	I find the SMILE Learning Management System easy to use.					100
C.	**Efficiency**					
C1	I can easily access the menu on the SMILE Learning Management System.				50.0	50.0
C2	I can easily obtain the available information on the SMILE Learning Management System.				66.7	33.3
C3	I can quickly find the information I need by starting my search in the SMILE Learning Management System.				66.7	33.3
D.	**Errors**					
D1	I haven't found an error when using the SMILE Learning Management System.			16.65	66.7	16.65
D2	I did not encounter any menu errors or mismatched functions while using the SMILE Learning Management System.				16.7	83.3
D3	In the SMILE Learning Management System, I can quickly locate the features and menus I'm looking for.			16.7	50.0	33.3
E.	**Satisfaction**					
E1	I am pleased with the SMILE Learning Management System's overall interface design.				50.0	50.0
E2	I feel comfortable using the SMILE Learning Management System.				33.3	66.7
E3	The color mix and layout of the content on the SMILE Learning Management System are convenient to see.				66.7	33.3
E4	When I saw the display on the system's dashboard, the SMILE Learning Management System exceeded my expectations.			16.7	33.3	50.0
E5	I found other supporting equipment such as file sharing, communication features such as chat and e-mail, as well as feature support.				50.0	50.0

The same approach was also used against the student's perspective. In the learnability aspect, 50.0% strongly agreed that the Learning Management System SMILE can be easily studied (A1), while the remaining 50% agreed. For (A2), 37.5% stated they strongly agreed, while 62.5% agreed. It follows with (A3) at 50.0% stated strongly agreed and 50.0% agreed. The (A4) got 37.5% strongly agreed, 50.0% agreed, while 12.5% tended to agree. I was able to learn how to use the SMILE Learning Management System without any written instructions or a manual book (A5); 37.5% of the students said they strongly agreed, while 62.5% said they agreed.

From the memorability aspect (B1), I have no trouble recalling how to use the SMILE Learning Management System. It is equally 50.0% between "strongly agreed" and "agreed". The results obtained from the student's perspective are in order: (B2) 37.5% strongly agreed, 50.0 agreed, and 12.5% tend to agree; and (B3) I find the SMILE Learning Management System easy to use; all the students stated that they strongly agreed.

The assessment continued with efficiency aspects; (C1) I can easily access the menu on the SMILE Learning Management System. It was strongly agreed upon by 50.0%, agreed upon by 37.5%, and tended to agree upon by 12.5%. I can easily obtain the available information on the SMILE Learning Management System; 37.5% strongly agreed, 50.0% agreed, with 12.5% inclined to agree. (C3) I can quickly find the information I need by starting my search in the SMILE Learning Management System; 37.5% strongly agreed, 50.0% agreed, and 12.5% tended to agree.

In terms of the errors' aspects, the result shows that 25.0% strongly agreed, 62.5% agreed, and 12.5% tend to agree on (D1). I haven't found an error when using the SMILE Learning Management System. It is followed by (D2). I did not encounter any menu errors or mismatched functions while using the SMILE Learning Management System; 37.5% strongly agreed, 50.0% agreed, and 12.5% tended to agree. On (D3), In the SMILE Learning Management System, I can quickly locate the features and menus I'm looking for, it stated that 37.5% strongly agreed, 50.0% agreed, and 12.5% tend to agree.

The assessment concludes with a satisfaction aspect, generating the following result: (E1). I am pleased with the SMILE Learning Management System's overall interface design. 25% of them strongly agreed, 62.5% agreed, and 12.5% agreed. (E2). I feel comfortable using the SMILE Learning Management System. 37.5% stated they strongly agreed and 62.5% agreed. (E3). The color scheme and layout of the content on the SMILE Learning Management System are easy to understand. 37.5% strongly agreed, 50.0% agreed, and 12.5 agreed with some reservations. (E4). When I saw the display on the system's dashboard, the SMILE Learning Management System exceeded my expectations; 25.0% stated strongly agreed and 75.0% agreed. (E5). I discovered additional assistive equipment such as file sharing, communication features such as chat and e-mail, and feature support received 50.0% strongly agreed, 37.5% agreed, and 12.5 percent tend to agree responses.

D. Relevant Work

The developed Learning Management System (LMS) is identified as SMILE, which stands for Simple, Multi-dimensional, and Interactive Learning Ecosystem. It is designed to assist students in understanding learning in a relatively simple, insightful, and unique way in order to maximize their potential and competencies. It is accessible via mobile devices, personal computers (PCs), and various browsers [4].

Mcgreal's research outlines the organizational context that motivates the use of Open Educational Resources (OER) in teaching and learning, including policies, practices, development processes, and resources [14].

Meanwhile, Wang, Woo, Quek, Yang, & Liu explored the use of Facebook Groups in LMS development based on the availability of prospective pedagogical, interpersonal, and technical capabilities that enable the exchange of ideas, resources, announcements, and online discussions. Directly uploading files in various formats presents challenges, but this discussion lacks meaningful structure. Students can communicate and interact with one another with ease. Despite this, there is a failure to provide a safe social environment due to a lack of privacy [15].

Shin and Kang used a mobile application-based LMS to conduct research on student acceptance and its impact on learning objectives. Students at online universities, according to their findings, are open to the use of mobile technology as a new learning tool. Acceptance has a direct or indirect impact on student achievement. These findings help to improve our understanding of the use of mobile learning systems in higher education and provide useful guidelines for the development and implementation of mobile application-based learning management systems [16].

Mtebe claims that the use of LMS in universities in Sub-Saharan Africa is expanding, and the majority of the universities' limited resources are being used to develop LMS. Mtebe conducted a review of the literature on LMS use in Sub-Saharan Africa, as well as proposed strategies to assist institutions in using LMS more effectively while saving money [17].

The ideal didactic curriculum, according to Toohey, Wray, Wiechmann, Lin, and Boysen-Osborn, should include a mix of asynchronous and synchronous learning. A Learning Management System (LMS) such as Schoology, Canvas, or Blackboard can manage a Learning Resource Center (LRC) program, which provides students with video menus, interactive education modules, articles, quizzes, didactic recordings, and other assignments. Implementing an LMS in a residency program can range from free (Schoology) to a per-student fee that may be covered by the university [18].

V. Conclusion and Future Work

In this study, we look at how to design, validate, and determine the efficacy of a learning management system that relies on cyber pedagogy and national educational standards in the higher education e-learning ecosystem. The SMILE Learning Management System model was developed by integrating various learning resources based on the characteristics of the current generation of learners.

Usability testing was used to evaluate the efficacy of the SMILE prototype. The results are consistent with the observed initial objective of delivering a productive learning process in simple, interesting, and interactive ways through the use of various multi-dimensional educational resources inside the corridor of Cyber Pedagogy and National Education Standards.

This research can be enhanced in a number of different ways. This type of prototype can be developed in other universities with features tailored to the needs of each campus, including lecture content that is wide and varied, as well as the use of learning media that is simple to understand and tailored to the needs of students, to support an effective and efficient learning process.

Acknowledgment

The researcher would like to express gratitude to all of the students and professors who participated in the study. This research is based on the first author's dissertation, which has been submitted since 2021 at Ganesha University of Education. The researcher sincerely thanks the supervisor, who is listed as a co-author, for their expertise.

References

[1] R. Indrajit, “The Ecosystem Framework for U-Learning Implementation Through Courseware Development and Management,” no. 2000, pp. 245–254, 2017.

[2] P. Pyöriä, S. Ojala, T. Saari, and K. M. Järvinen, “The millennial generation: A new breed of labour?,” SAGE Open, vol. 7, no. 1, 2017, doi: 10.1177/2158244017697158.

[3] T. A. Conklin, “Making It Personal: The Importance of Student Experience in Creating Autonomy-Supportive Classrooms for Millennial Learners,” Journal of Management Education, vol. 37, no. 4. 2013. doi: 10.1177/1052562912456296.

[4] P. Dewanti, I. M. Candiasa, and I. M. Tegeh, “Designing a Learning Management System based on Cyber Pedagogy for Higher Education,” International Journal of Computer Applications, vol. 183, no. 27, 2021, doi: 10.5120/ijca2021921656.

[5] H. Shafiq, Z. A. Wani, I. M. Mahajan, and U. Qadri, “Courses beyond borders: A case study of MOOC platform Coursera,” Library Philosophy and Practice, vol. 2017, 2017.

[6] P. S. Aithal and S. Aithal, “Impact of On-Line Education on Higher Education System,” 2016.

[7] M. H. Arifin, “The role of student support services in enhancing student persistence in the Open University Context: Lesson from Indonesia Open University,” Turkish Online Journal of Distance Education, vol. 19, no. 3, 2018, doi: 10.17718/tojde.445116.

[8] I. Jung, G. A. G. Mendoza, J. C. Fajardo, R. B. Figueroa, and S. E. Tan, “MOOCs in six emerging apec member economies: Trends, research, and recommendations,” MOOCs and Open Education in the Global South: Challenges, Successes, and Opportunities, no. January, pp. 199–211, 2019, doi: 10.4324/9780429398919-21.

[9] R. Eller, "Implementing Design into Instruction: Intersections Between the Waterfall Model and Addie Model," vol. 1, no. 1, pp. 63–69, 2015.

[10] Gharajeh MS, "Waterative model: An integration of the waterfall and iterative software development paradigms," Database Systems Journal, vol. X, no. 15, pp. 75–81, 2019.

[11] V. Rastogi, "Software Development Life Cycle Models- Comparison, Consequences," International Journal of Computer Science and Information Technologies, vol. 6, no. 1, pp. 168–172, 2015.

[12] A. Gordillo, E. Barra, and J. Quemada, "An easy-to-use open source authoring tool to create effective and reusable learning objects," Computer Applications in Engineering Education, vol. 25, no. 2, 2017, doi: 10.1002/cae.21789.

[13] H. Retnawati, "Proving content validity of self-regulated learning scale (The comparison of Aiken index and expanded Gregory index)," Research and Evaluation in Education, vol. 2, no. 2, 2016, doi: 10.21831/reid.v2i2.11029.

[14] R. Mcgreal, "A survey of OER implementations in 13 higher education institutions," International Review of Research in Open and Distance Learning, vol. 20, no. 5, 2019, doi: 10.19173/irrodl.v20i5.4577.

[15] Q. Wang, H. L. Woo, C. L. Quek, Y. Yang, and M. Liu, "Using the Facebook group as a learning management system: An exploratory study," British Journal of Educational Technology, vol. 43, no. 3, pp. 428–438, May 2012, doi: 10.1111/j.1467-8535.2011.01195.x.
S. Shin and M. Kang, "The Use of a Mobile Learning Management System at an Online University and Its Effect on Learning Satisfaction and Achievement."

[16] J. S. Mtebe, "Learning Management System success: Increasing Learning Management System usage in higher education in sub-Saharan Africa," International Journal of Education and Development using Information and Communication Technology, vol. 11, no. 2, pp. 51–64, 2015, [Online]. Available: http://ijedict.dec.uwi.edu/viewarticle.php?id=2005.

[17] S. L. Toohey, A. Wray, W. Wiechmann, M. Lin, and M. Boysen-Osborn, "Ten tips for engaging the millennial learner and moving an emergency medicine residency curriculum into the 21st century," Western Journal of Emergency Medicine, vol. 17, no. 3. 2016. doi: 10.5811/westjem.2016.3.29863.

Location-based Mobile Application for Blood Donor Search

Orlando Iparraguirre-Villanueva[1], Fernando Sierra-Liñan[2], Michael Cabanillas-Carbonell[3]
Facultad de Ingeniería y Arquitectura, Universidad Autónoma del Perú, Lima, Perú[1]
Facultad de Ingeniería, Universidad Privada del Norte, Lima, Perú[2]
Vicerrectorado de Investigación, Universidad Norbert Wiener, Lima, Perú[3]

***Abstract*—Technological advances and the massive use of mobile devices have led to the exponential evolution of mobile applications in the health sector. Blood donation centers frequently suffer blood shortages due to lack of donations, which is why blood donation requests are frequently seen on social networks for blood donors in urgent need of a transfusion of a specific blood group. Mobile applications for blood donation are crucial in the health sector, since it allows donors and blood donation centers to communicate immediately to coordinate with each other, minimizing the time to perform the donation process. The present work was to develop a location-based mobile application for the search of blood donors, with the objective of increasing the number of donors, having a greater population reach, and reducing the time to search for blood donors. The results obtained show a significant increase of 39.58% in the number of donors, a reduction of 53.2% in the search time, and a greater population reach.**

***Keywords*—Blood banks; location; mobile applications; donor; applicant; blood bank; mobile apps**

I. Introduction

Donating blood is helping to save the lives of those who need a transfusion in a selfless and supportive way. Worldwide, 118.5 million blood donations are collected, 40% of which are in high-income countries [30]. In Peru, approximately 270 thousand units of blood are needed per year to guarantee the normal supply of this resource in the blood donation centers of the Ministry of Health. In donation campaigns, thousands of units of blood are collected from donors, although much more are needed to meet the demand of applicants [1] [2]. Deterrents and motivations for voluntary blood donation may differ by age and ethnicity [3]. Blood donation centers often suffer from shortages, so it is the applicants who look for their blood donor, often without success, therefore, strategies for effective donor recruitment and retention are required. Blood donation centers still use traditional methods for donor recruitment, such as blood donation campaigns, word of mouth, social networks, etc., although they are still valuable, but they are becoming less and less effective [4]. In addition, the Ministry of Health's blood donation centers work in isolation, without integration with other health institutions, which affects the quality of service. In this context, information technologies and smartphones have become an ally for blood donation [5][29].

The present research proposes the development of a location-based mobile application for blood donor search (DONAPE), for which the mobile application provides a direct location-based channel between blood seekers and blood donation centers. Achieving to increase the number of donors, improve the place of origin (geographical location) of donors and improve the search time.

The article is organized as follows. Section 2 describes the main related studies that have been conducted in the field of location-based blood donation and mobile technology. Section 3 presents the location-based mobile application development methodology and research method. Section 4 develops the case study. Section 5 describes and analyzes the results. Section 6 presents a brief discussion. Finally, Section 7 presents the conclusions.

II. Review of the Literature

Currently, in order to guarantee the required supply of blood units, it is necessary to focus efforts on donor recruitment and retention. Therefore, it is important to know the characteristics of donors, their motives and deterrents to voluntary donation. In this context, technological progress has played a very important role in blood donation; technology has made it possible to streamline and automate processes. In this sense, a set of works related to the search for blood donors has been reviewed.

In [6] developed a mobile application for Android, which offers a simple and fast approach to searching for blood, where they can easily find donor and recipient data through their cell phones. While researchers at [7][8] developed cloud-based web applications to administer and manage the health system's blood units. These applications allow blood units to be allocated according to the severity of the case, availability of types and quantities. Along the same lines, the research work [9] proposed a method for allocating blood units, using a mobile application, which utilized a real-time hybrid algorithm to develop the blood unit allocation operation. In [10][11] [12][13] web and mobile applications were developed to manage blood donation, allowing to register, schedule, receive notifications and access information, synchronizing blood donation centers with emergency centers, to verify the availability of blood needed and to send a request to the nearest blood donation center.

Web applications for blood donation have also been developed using new technologies such as blockchain and learning algorithms. In [14][15] web applications were developed using blockchain technology, applying machine learning algorithms to develop and evaluate models for classifying blood donors as returnees and non-returnees. In [16] [17] [18] proposed web-based applications to manage blood donation campaigns, with the aim of collecting and organizing data. The researchers at [19][20] developed web applications to monitor and investigate the risks involved with donors before and after the donation process.

Although previous studies considered many aspects of the blood donation process, each of them exclusively addressed issues related to web or mobile applications as managers for blood donation, forgetting the importance of managing the location of the donor and applicant. The main reason for this research is that through the mobile application we can manage the location of both the donor and the applicant to make and improve the blood donation service.

III. Methodology

For the present work we chose to use the agile Scrum method to develop the project prototype. Scrum is an adaptive, iterative, fast, flexible and efficient framework. Scrum ensures transparency in communication and creates an environment of collective responsibility and continuous progress [21]. This method has 5 phases: initiation, planning and estimation, implementation, review and retrospective and launch, for the development of this project we will work with four processes, which are described below.

A. Sprint Planning

In this process, meetings are held to define the user stories, estimate the tasks and create the sprint backlog, and a time-box of eight hours is assigned during a month-long sprint [22]. In this process the Scrum team decides to complete the selected items in the prioritized product backlog to meet the sprint goal.

B. Sprint Development Work

In this process, the Sprint Backlog is implemented, no changes are made that affect the objective of the sprint, quality objectives are ensured, the scope is renegotiated with the product owner and the development team as it is developed. [23]. In this process it is possible to preview the inspection and the adaptation of the progress towards the goal according to the schedule of activities.

C. Sprint Review

The sprint review is conducted to demonstrate and validate the sprint, the Scrum Team presents the deliverables of the current sprint to the Product Owner. The Product Owner reviews the product increment against the agreed acceptance criteria and accepts or rejects the completed user stories [23].

D. Sprint Retrospective

In this process, a 4-hour meeting is held in a one-month sprint, and is developed as part of the sprint retrospective process. During this meeting, the Scrum team meets to review and reflect on the previous sprint in relation to the processes that were followed, the tools used, the collaboration and communication mechanisms, as well as other aspects of interest to the project [24].

A strength of Scrum lies in the use of cross-functional, organized and empowered teams that divide their work into short, concentrated work cycles called sprints.

The main characteristics of mobile applications are high performance and availability, which is why the client layer is developed offline, as shown in the architecture in Fig. 1.

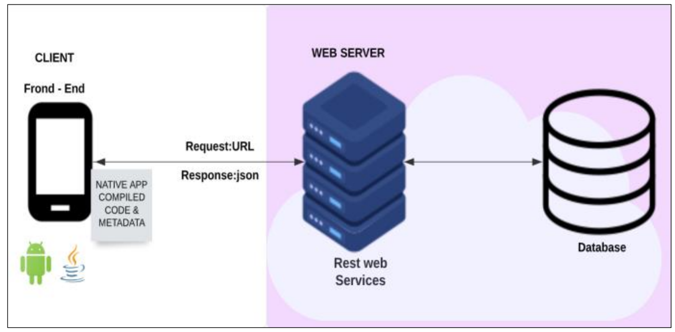


Fig. 1. Mobile Application Architecture.

IV. CASE STUDY

This section details the development procedure of the mobile application prototype, following the life cycle methodology mentioned above. Likewise, Table I describes the tools to be used in the development of the case.

TABLE I. Tools for Prototype Development

Software	Description
Android Studio	Development environment for the Android platform
Firebase	NoSql database
GitHub	Repository for version control
Figma	UI and UX design application
Play store	Online store service to distribute the mobile application.
Jira	Online software to manage project tasks.

The Scrum cycle starts with a stakeholder meeting, then the product owner develops the prioritized product backlog. Each sprint starts with a planning meeting where the prioritization of the stories is considered, and generally lasts from one to six weeks. Then, the sprint planning and time estimation for each of the user stories to be implemented is done.

A. Planification Del Sprint

In this section the analysis and identification of the requirements for the elaboration of the User Epics is performed, together with the descriptions of the functionalities that the mobile application will have, which are elaborated between the product owner and the Scrum team, and that can be improved during the project life cycle. This is an agile way to manage requirements without elaborating large amounts of documentation [25]. Table II describes the user epics that will be implemented for the development of the case study.

TABLE II. USER EPICS

#	Description
ADS-1	Manage the authentication module: The mobile application should have a login interface with security token for user registration and authentication
ADS-2	Manage the notification module: The mobile application must have the option of notifications to indicate to the user of requests.
ADS-3	Manage Google maps module: The mobile application must incorporate Google maps technology for real-time search of donors or applicants based on location.
ADS-4	Manage the health center location module: The mobile application, through geolocation, should locate the closest health centers for the blood donation process.
ADS-5	Manage the reporting module: The mobile application must show the available or connected donors, as well as the list of applicants..

Time estimation: In this phase the Scrum team decides to complete the prioritized product backlog, assigning the duration, priority and complexity for each of the user stories. There are many tools to estimate user stories, for this case Planing Poker is used to estimate and indicate the complexity whose value only makes sense to the development team [26]. The project estimate resulted in nine weeks, as shown in Table III.

TABLE III. PRODUCT BACKLOG

Description	*Duration Weeks*	*Story point estimate*	*Priority*	*Epic*
Sprint 1: User authentication module	***2 weeks***	***40***		
Implement the login interface		16	highest	ADS-1
Implement the interface for the registration of new users (donors and applicants).		14	highest	ADS-1
Implement the interface for password recovery		10	highest	ADS-1
Sprint 2: notification module	***2 weeks***	***35***		
Implement the notifications component in the startup interface		18	high	ADS-2
Implementing the tray where notification messages are received		17	high	ADS-2
Sprint 3: geolocation module	***3 weeks***	***40***		
Implement the google maps service in the application.		14	highest	ADS-3
Implement the real-time search option with the google maps service.		16	highest	ADS-3
Implement the search for the nearest health centers using geolocation.		10	highest	ADS-4
Sprint 4: reporting module	***2 weeks***	***33***		
Implement reports of available donors and applicants by geographic location.		17	medium	ADS-5
Implementing donor and applicant reports that have been attended to		16	medium	ADS-5

Project scope: understood as the sum of all product requirements and all work necessary to develop the final product/service [24]. The functionalities originate from the set of requirements established and analyzed together with the product owner. Therefore, in order to know the fulfillment of the project scope, the degree of progress of each of the user stories of the product backlog is verified, in order to know if the implementation is being carried out as planned.

Fig. 2 shows the hours that the team takes to develop the user story points, in order to verify in greater detail the degree of progress in the sprint estimates.

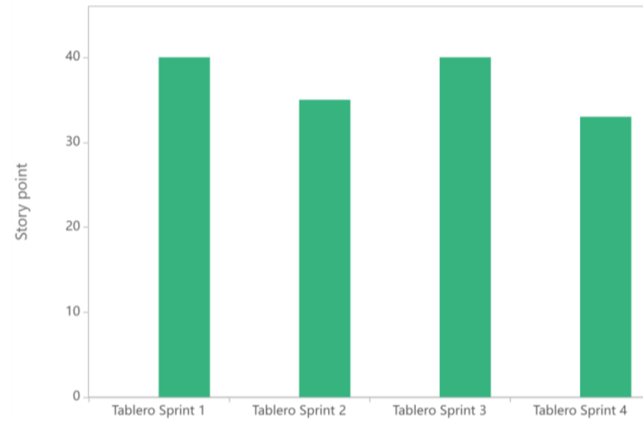


Fig. 2. Estimation of each Sprint.

B. *Sprint Development*

In this process, the following sprints are developed:

1) Sprint 1 (user authentication): In this Sprint the interface for the login Fig. 3, interface for the registration of new users (donor or applicant) and password recovery is developed. The user must have the access credentials to log in, otherwise, he/she will have to register his/her data as requested in the form. If the user forgets his/her password, the recovery process is performed.

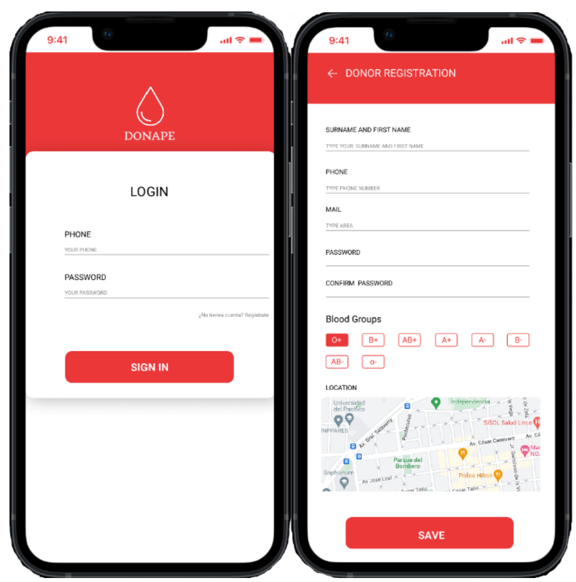


Fig. 3. Donor Login and Registration.

2) Sprint 2 (notification module): In this second Sprint, the interface for the notifications module was implemented. As shown in Fig. 4, this component has two sections: 1 for the inbox, and another one located in the upper right part of the startup interface.

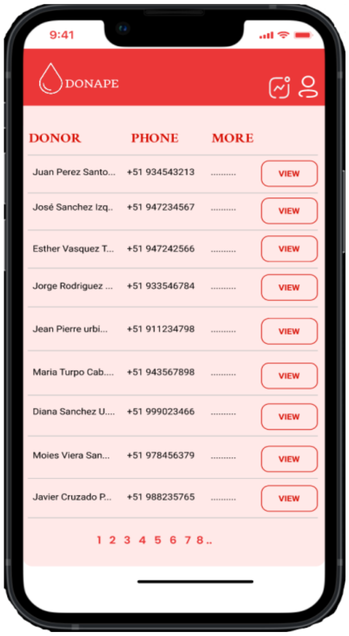


Fig. 4. Notifications and Tray.

3) Sprint 3 (geolocation module): In this sprint, the Google maps service is incorporated into the prototype, in order to search for donors and locate donation centers, as shown in Fig. 5. In the first interface of Fig. 5, the applicant searches for blood donors. In principle, the algorithm performs the search based on the location of the applicant, prioritizing the proximity of the donor. Once the donor has been located, the applicant searches for the health center that has a blood bank, using the second interface in Fig. 5.

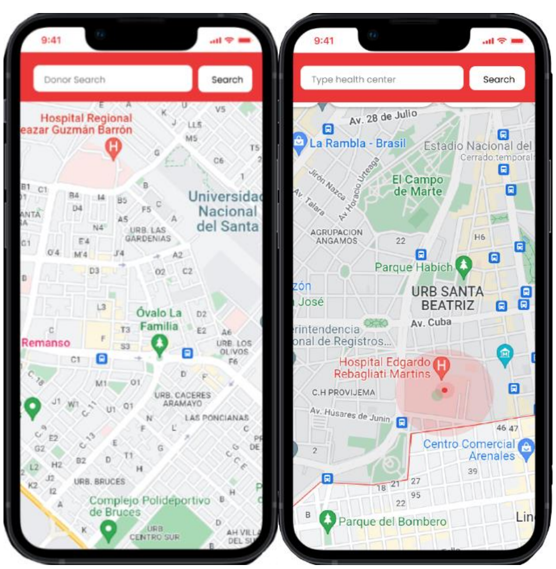


Fig. 5. Search for Donors (a) and Blood Donation Centers (b).

In the interfaces of Fig. 5(a), the first interface is where the applicant searches, chooses the donor and can contact immediately. (b) The second interface is used to search for the blood donation center closest to the interested parties, in order to carry out the blood transfusion process. The donor will be able to cancel the request through the application, in case they do not reach a mutual agreement.

4) Sprint 4 (reporting module): In this sprint the donor reports are developed, where the names, city of origin, cell phone number, date of last donation and blood type can be displayed, as shown in Fig. 6.

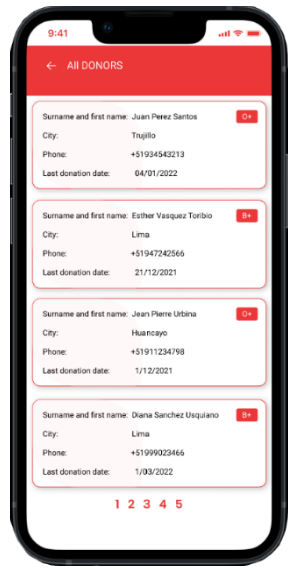


Fig. 6. Donor Report.

C. Sprint Review

After the development of each Sprint, we proceed with the review, for this Scrum Team members and Stakeholders participate in the review of the sprint to accept the deliverables, the review time is two to three hours. At this stage the Scrum Team demonstrates the achievements of the sprint, specifying each of the new features. This provides an opportunity for the product owner to inspect in detail what has been completed so far and determine if changes should be applied to the sprint.

D. Sprint Retrospective

The retrospective is the last step in a sprint. All Scrum team members attend this meeting, which is organized by the Scrum mater, this meeting together with the product owner, discusses important elements for future actions. This retrospective meeting covers both what went wrong and what went right. It also evaluates the process followed in the development of the sprints, tools, among other elements. It is also a space for new ideas or methods in order to continue improving the process and promoting good practices.

To demonstrate the results of the implementation of the mobile application, it was tested for 20 days. The design used was pre-experimental, with a pre-test and a post-test, since the online method was used (1). Table IV shows the elements of the design.

TABLE IV. DESCRIPTION OF PRE-EXPERIMENTAL ELEMENTS

Elements (1)	Description
Ge	Experimental group: the study is carried out in the Dlima laboratory for 20 days.
O_1	It is the measurement before implementing the mobile application. Pre-test data measurement: experimental group.
X	Mobile application = object to be tested
O_2	Measurement of post-test data: experimental group.

$G_e\ O_1\ X\ O_2$

V. RESULTS

This section describes the results related to the development of the research within the case study. In Peru there is no culture of voluntary blood donation, the need is permanent and the numbers required to cover the demand are high [1].

TABLE V. RESULTS OBTAINED IN THE PRE-TEST AND POST TEST

#Dia	KPI-1		KPI-2		KPI-3 (Horas)	
	Pre	Post	Pre	Post	Pre	Post
1	0	3	----	Piura – Olivos- Surco	8	4
2	0	2	----	Lince – Cercado	8	4
3	2	4	Lince - Surco	Trujillo – 2 Cuzco- SJL	6	3
4	1	2	Breña	Olivos – Huacho	8	4
5	1	2	Lince	Arequipa – Breña	8	4
6	2	2	Olivos -SJL	Breña – Tarapoto	4	4
7	1	1	Cercado	Olivos	8	3
8	2	2	SJL-Lima	Trujillo – Chimbote	4	4
9	2	2	Lince - olivos	Piura – SJL	4	3
10	2	3	Comas-Breña	Olivos – Breña – Piura	4	3
11	2	3	Comas-surco	Lince – olivos -Trujillo	4	3
12	3	5	SJL-olivos-lince	Arequipa – Cercado – Lince – Comas – SJL	4	2
13	2	2	Lince - breña	Chimbote - Olivos	4	2
14	2	3	Cercado - Surco	Cercado - Piura -Trujillo	4	2
15	3	3	Comas - olivos - SJL	Surco-Comas - Olivos	3	3
16	2	4	Lince - cercado	Cuzco - Surco – Arequipa – Chimbote	4	2
17	1	5	Cercado	Tarapoto – Breña – Surco-Piura- Comas	8	1
18	2	3	Surco - olivos	Trujillo-Comas - SJL	4	3
19	1	3	Lince	Breña – Olivos – Cuzco	8	2
20	2	3	SJL - Comas	Trujillo – Piura - Breña	4	2

For the case study, we worked with the Dlima laboratory, a blood bank that offers a modern and highly reliable blood transfusion system. Since the beginning of the COVID-19 state of emergency, blood donations have dropped by more than 70% due to fear of infection [27]. Since then, the laboratory implemented a set of strategies to search for donors and maintain the blood bank reserves and promote the culture of voluntary donation even in the context of a pandemic. One of the strategies was the proposal to develop a location-based mobile application to search for blood donors, with the following indicators: increase the number of donors (KPI-1), place of origin (geographical location) of the donors (KPI-2) and improve the search time (KPI-3). It is known that screening tests for hepatitis B and C, HIV, HTLV I and II, Chagas disease and syphilis are performed prior to blood transfusion. In our environment there are still prejudices regarding blood donation, therefore, the laboratory implemented a reward program that consists of performing a blood test known as "complete blood count" for any altruistic donor who decides to use the mobile application to donate blood. The data collection (Table V) for the pre-test measurement used a record type card and for the post-test the automated record of the mobile application.

Table VI shows the average results for each KPI; these results were derived from Table V. It should be noted that the districts of Lince, Breña, Cercado, Surco, Comas, Olivos and JSL belong to the city of Lima.

Fig. 7 shows that the average of KPI-1 has improved in the number of donors, in KPI-3 the average indicates that the time to search for donors has been reduced, in the same line, it is observed that in KPI-2 after the implementation of the mobile application, access to other regions of the country has been achieved.

To measure the indicators in Table VI, we used the scales of time in hours, quantity and distance.

TABLE VI. RESEARCH INDICATORS

Indicator		Pre-Test	Post Test
KPI-1	Number of donors	1.65	2.85
KPI-2	Donor origin	Lima	Lima + Provincia
KPI-3	Donor search time	5.45	2.9

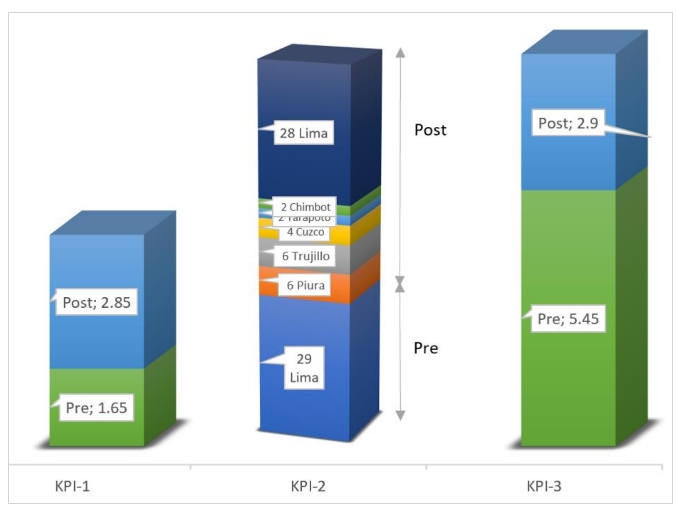


Fig. 7. Pre and Post Comparison of KPI's.

A. *Results of the First Indicator (KPI-1)*

Fig. 8 shows the normality test for KPI-1, where a standard deviation of 1.04 is observed with respect to the mean of 2.85 donors per day. Along the same lines, the data obtained show that the p-value is less than 0.05, which confirms that the information analyzed has a non-normal behavior.

Regarding the nonparametric Wilcoxon test (Table VII). It has an asymptotic significance level of 0.001, which is less than 0.05, which is the limit value for the acceptance of the hypothesis, therefore, it is stated that the implementation of the mobile application has a positive impact on attracting a greater number of blood donors.

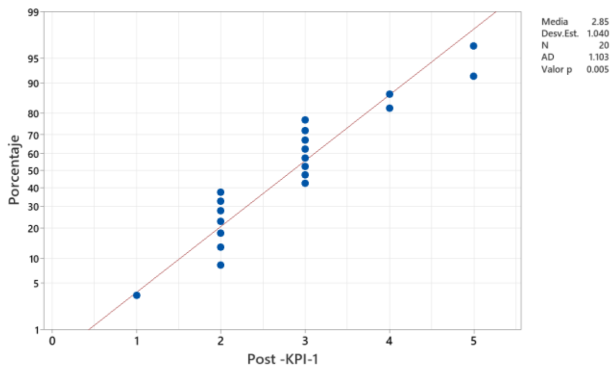


Fig. 8. Normality Test of KPI-1.

TABLE VII. WILCOXON´S TEST FOR KPI-1

Statistical test	
	PRE-TEST KPI-1 POST TEST KPI-1
Z	-3.359b
Sig. Asintótica(bilateral)	.001
Wilcoxon signed-rank test	
Based on positive ranges	

B. *Results of the Second Indicator (KPI-2)*

Fig. 9 shows the results of the second indicator, which evaluates the donors' place of origin. In the pretest, the origin of the donors belonged only to the city of Lima, including the districts (Lince, Surco, Breña, Los Olivos, SJL, Cercado and Comas), where a total of 29 blood donors were obtained. In the post test, with the use of the mobile application, a greater reach was achieved, obtaining blood donors from different cities in the country, such as: Piura, Trujillo, Cuzco, Tarapoto, Chimbote, among other districts of the city of Lima, totaling 48 blood donors. Therefore, we can affirm that after the implementation of the project, we have a positive impact on the population and a greater reach.

C. *Results of the Third Indicator (KPI-3)*

Fig. 10 shows the normality test for KPI-3, where a standard deviation of 0.9119 is observed with respect to the mean of 2.9 hours required to obtain a donor. In the same line, the data obtained show that the p-value is less than 0.05, which confirms that the information analyzed has a non-normal behavior.

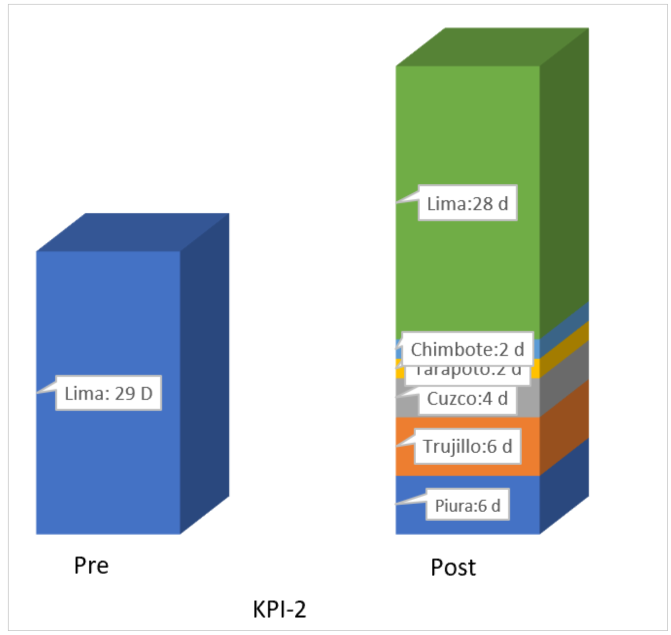


Fig. 9. KPI-2 Measurement Results.

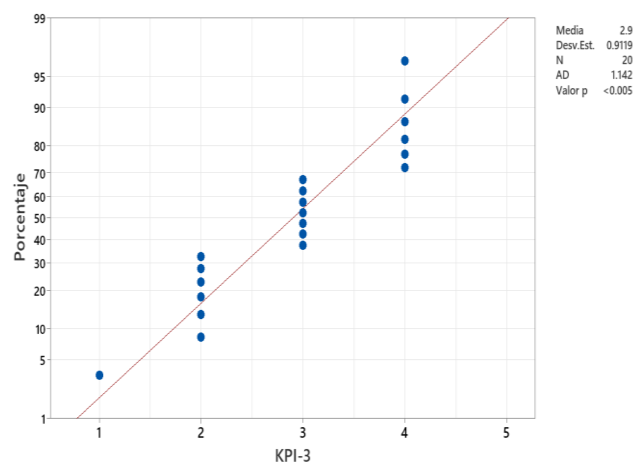


Fig. 10. KPI-3 Normality Test.

Regarding the nonparametric Wilcoxon test (Table VIII). It has an asymptotic significance level of 0.000, which is less than 0.05, which is the threshold value for the acceptance of the hypothesis, therefore, it is stated that the implementation of the mobile application has a positive impact with respect to measuring the time to recruit blood donors.

TABLE VIII. WILCOXON´S TEST FOR KPI-3

Statistical test	
	PRE-TEST KPI-3 POST TEST KPI-3
Z	-3.642b
Sig. Asintótica(bilateral)	.000
Wilcoxon signed-rank test	
Based on positive ranges	

VI. Discussion

About the case study: the mobile application was made with Android Studio, Firebase, among other tools such as Jira, which was used to manage user stories, allowing a controlled monitoring by both the team and the product owner. A set of works related to blood donation has been reviewed, where technology is used as an ally for this purpose. Works such as: [16][17] developed mobile applications for blood donation integrating cloud services with the purpose of promoting and facilitating blood donation. In the meantime [12][13] In their projects, they focused on the development of mobile applications to manage notifications, access donor information and synchronize with the blood banks, all of which have undoubtedly contributed to the development of this work.

About the methodology: Scrum is an agile methodology for software development, in this group there is a set of methodologies based on agility. However, the main reason why Scrum was chosen for the development of the mobile application is the short iterations known as sprints that allow early deliveries of two to four weeks and in turn a quick feedback from users and stakeholders [28]. Among other advantages and benefits of the Scrum methodology we can highlight: it helps to manage and minimize project risks, it improves the relationship between cost and benefit, it allows the development of team skills, and it works mostly with small teams.

With respect to the limitations that were found in the research work and that could have affected the results, in principle there was no list of blood donation centers authorized by the ministry, and to a lesser extent the time it took for the play store to review the mobile application.

VII. Conclusion

In this research article, a location-based mobile application was developed to search for blood donors, using the Scrum methodology, Google maps service to locate the donor and blood donation centers in real time.

After the implementation of the project, there was a significant increase in the number of blood donors by 39.58%, as well as a reduction of 53.2% in the time required to search for blood donors. The results show that donors from different cities in the interior of the country such as Piura, Trujillo, Chimbote, Cuzco and Tarapoto, representing 39.58%, however, it is important to note that the city of Lima continues to concentrate the largest number of blood donors, equivalent to 60.42% according to the case study.

The scientific contribution of this article is fundamental for the development of future work related to blood donation campaigns. With the data collected through the mobile application, it is possible to perform analytical processing, develop strategies with reward incentives to motivate blood donation, and add new functionalities so that doctors and patients can communicate.

Finally, it was demonstrated that the solution developed, manages to increase the number of blood donors, reduce the time to get donors, has greater population reach and in turn has access to the information provided by the donor, and with this to perform actions in favor of those concerned.

References

[1] Ministerio de Salud, “Donantes de sangre.” https://www.gob.pe/institucion/minsa/noticias (accessed Mar. 11, 2022).

[2] B. M. Shashikala, M. P. Pushpalatha, and B. Vijaya, Web based blood donation management system (BDMS) and notifications, vol. 801. Springer Singapore, 2018. doi: 10.1007/978-981-10-9059-2_12.

[3] S. Ouhbi, J. L. Fernández-Alemán, A. Toval, A. Idri, and J. R. Pozo, “Free Blood Donation Mobile Applications,” Journal of Medical Systems, vol. 39, no. 5, pp. 1–20, 2015, doi: 10.1007/s10916-015-0228-0.

[4] S. Moussa, “A location-based framework for mobile blood donation and consumption assessment using big data analytics,” Asian Journal of Information Technology, vol. 15, no. 22, pp. 4475–4481, 2016.

[5] M. N. Sahid Ramadhan, A. Amyus, A. N. Fajar, S. Sfenrianto, A. F. Kanz, and M. S. Mufaqih, “Blood Bank Information System Based on Cloud Computing in Indonesia,” Journal of Physics: Conference Series, vol. 1179, no. 1, Aug. 2019, doi: 10.1088/1742-6596/1179/1/012028.

[6] C. Saranya Jothi, V. Deepa, and V. Usha, “Effective on demand blood sharing app for medical emergency,” International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9, pp. 186–188, Jul. 2019, doi: 10.35940/IJITEE.I7708.078919.

[7] G. M. V. S. Sai Mahesh and C. Rajagopal, “Healthcare services on donate to save,” Research Journal of Pharmaceutical, Biological and Chemical Sciences, vol. 7, pp. 144–148, 2016.

[8] D. B. Abdullah and M. D. Younus, “Real-time cloud system for managing blood units and convalescent plasma for COVID-19 patients,” International Journal of Electrical and Computer Engineering, vol. 11, no. 4, pp. 3593–3600, Aug. 2021, doi: 10.11591/IJECE.V11I4.PP3593-3600.

[9] D. B. Abdulla and M. D. Younus, “Proposed Method to Allocate the Blood Units by Using a Hybrid of Real-Time Algorithms on the Cloud Environment,” International Journal of Computing and Digital Systems, vol. 11, no. 1, pp. 703–711, 2022, doi: 10.12785/IJCDS/110157.

[10] P. L. Wijayathilaka, P. H. Pahala Gamage, K. H. B. De Silva, A. P. P. S. Athukorala, K. A. D. C. P. Kahandawaarachchi, and K. N. Pulasinghe, “Secured, intelligent blood and organ donation management system - ‘Lifeshare,’” ICAC 2020 - 2nd International Conference on Advancements in Computing, Proceedings, pp. 374–379, Dec. 2020, doi: 10.1109/ICAC51239.2020.9357211.

[11] H. Hegedus, K. Szasz, K. Simon, T. Fazakas, A. Mihaly, and K. Nagy, “Blood notes: Software system for promoting and facilitating blood donation,” SISY 2019 - IEEE 17th International Symposium on Intelligent Systems and Informatics, Proceedings, pp. 77–82, Sep. 2019, doi: 10.1109/SISY47553.2019.9111536.

[12] R. Kumar, R. Kumar, and M. Tyagi, “Web Based Online Blood Donation System,” 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), pp. 1630–1632, Dec. 2021, doi: 10.1109/ICAC3N53548.2021.9725558.

[13] S. Ramya, S. Kayathri, and S. Meena, “Life blood contribution using android application to avoid blood donation problems,” International Journal of Scientific and Technology Research, vol. 9, no. 3, pp. 6480–6482, Mar. 2020.

[14] L. A. Dajim, S. A. Al-Farras, B. S. Al-Shahrani, A. A. Al-Zuraib, and R. Merlin Mathew, “Organ Donation Decentralized Application Using Blockchain Technology,” 2nd International Conference on Computer Applications and Information Security, ICCAIS 2019, May 2019, doi: 10.1109/CAIS.2019.8769459.

[15] A. S. Alkahtani and M. Jilani, “Predicting return donor and analyzing blood donation time series using data mining techniques,” International Journal of Advanced Computer Science and Applications, vol. 10, no. 8, pp. 113–118, 2019, doi: 10.14569/IJACSA.2019.0100816.

[16] L. Dutta, G. Maji, P. Ghosh, and S. Sen, “An integrated blood donation campaign management system,” Advances in Intelligent Systems and Computing, vol. 812, pp. 133–143, 2019, doi: 10.1007/978-981-13-1540-4_14.

[17] R. S. Ali, T. F. Hafez, A. B. Ali, and N. Abd-Alsabour, “Blood bag: A web application to manage all blood donation and transfusion processes,” Proceedings of the 2017 International Conference on Wireless Communications, Signal Processing and Networking, WiSPNET 2017, vol. 2018-January, pp. 2125–2130, Feb. 2018, doi: 10.1109/WISPNET.2017.8300136.

[18] A. M. Alicia et al., “YAWARweb: Pilot study about the usage of a web service to raise awareness of blood donation campaigns on university campuses in Lima, Peru,” International Journal of Advanced Computer Science and Applications, vol. 10, no. 7, pp. 402–407, 2019, doi: 10.14569/IJACSA.2019.0100755.

[19] M. Nabil, R. Ihab, H. el Masry, S. Said, and S. Youssef, “A Web-based blood donation and Medical Monitoring System Integrating Cloud services and Mobile Application,” Journal of Physics: Conference Series, vol. 1447, no. 1, Jan. 2020, doi: 10.1088/1742-6596/1447/1/012001.

[20] R. Patil, M. Poi, P. Pawar, T. Patil, and N. Ghuse, “Blood donor’s safety using data mining,” Proceedings of the 2015 International Conference on Green Computing and Internet of Things, ICGCIoT 2015, pp. 500–505, Jan. 2016, doi: 10.1109/ICGCIOT.2015.7380516.

[21] R. Arias-Marreros, K. Nalvarte-Dionisio, and L. Andrade-Arenas, “Design of a Mobile Application for the Learning of People with Down Syndrome through Interactive Games,” International Journal of Advanced Computer Science and Applications, vol. 11, no. 11, pp. 715–721, 2020, doi: 10.14569/IJACSA.2020.0111187.

[22] L. A. Garcia, E. OliveiraJr, M. Morandini, and S. Urbanowski, “Tailoring the Scrum framework for software development: Literature mapping and feature-based support,” Information and Software Technology, vol. 146, Jun. 2022, doi: 10.1016/J.INFSOF.2021.106814.

[23] G. Torrente, T. Q. De Souza, L. Tonaki, A. P. Cardoso, L. Manickchand Junior, and G. O. Da Silva, “Scrum Framework and Health Solutions: Management and Results,” Studies in Health Technology and Informatics, vol. 284, pp. 290–294, Dec. 2021, doi: 10.3233/SHTI210725.

[24] F. Hays, “A guide to the SCRUM BODYOF KNOWLEDGE (SBOKTM Guide): A comprehensive Guide to Deliver Project using Scrum,” De Rebus, pp. 9–11, 2016.

[25] N. Freedrikson Arifin, B. Purwandari, and F. Setiadi, “Evaluation and Recommendation for Scrum Implementation Improvement with Hybrid Scrum Maturity Model: A Case Study of A New Telco Product,” 2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS), pp. 178–183, Nov. 2020, doi: 10.1109 / ICIMCIS51567.2020.9354311.

[26] A. Mundra, S. Misra, and C. A. Dhawale, “Practical Scrum-Scrum team: Way to produce successful and quality software,” Proceedings of the 2013 13th International Conference on Computational Science and Its Applications, ICCSA 2013, pp. 119–123, 2013, doi: 10.1109/ICCSA.2013.25.

[27] UNMSM, “Situación de la donación de sangre durante la pandemia – Chiqaq News.” https://medialab.unmsm.edu.pe/chiqaqnews/situacion-de-la-donacion-de-sangre-durante-la-pandemia/ (accessed Mar. 21, 2022).

[28] S. Shafiee, Y. Wautelet, L. Hvam, E. Sandrin, and C. Forza, “Scrum versus Rational Unified Process in facing the main challenges of product conFiguration systems development,” Journal of Systems and Software, vol. 170, Dec. 2020, doi: 10.1016/J.JSS.2020.110732.

[29] K. Arai, “Mobile phone operations just by sight and its applications,” Adv. Intell. Syst. Comput., vol. 857, pp. 1366–1374, 2019, doi: 10.1007/978-3-030-01177-2_102.

[30] “Disponibilidad y seguridad de la sangre a nivel mundial.” https://www.who.int/es/news-room/fact-sheets/detail/blood-safety-and-availability (accessed Mar. 28, 2022).

Soil Nutrients Prediction and Optimal Fertilizer Recommendation for Sustainable Cultivation of Groundnut Crop using Enhanced-1DCNN DLM

Sivasankaran S[1], Dr. K. Jagan Mohan[2], Dr. G. Mohammed Nazer[3]
Research Scholar, Department of Computer Science and Engineering
Annamalai University, Chidambaram, Tamilnadu, India[1]
Associate Professor, Department of Information Technology
Annamalai University, Chidambaram, Tamilnadu, India[2]
Professor and Principal, Department of Computer Science
RAAK Arts and Science College, Villupuram, Tamilnadu, India[3]

Abstract*—Cultivation of crops and their parallel production yields hugely depend upon the fertility composition of the soil in which the crops are being cultivated. The prime fertility factors which contribute towards the health of the soil are the available soil nutrients. Varying climatic conditions and improper cultivation patterns have resulted in unpredictable growth and yield of the groundnut crops, one of the major cause for the fluctuation seen in groundnut pod growth patterns and production, is the differing soil nutrient compositions of the land which is under cultivation. The unnecessary usage of excessive artificial fertilizers to boost the soil strength, without properly diagnosing the exact nutritional need of the soil required for the conducive growth of the crop has led to the imbalanced distribution of the soil's major macro-nutrients constituents such as (Phosphorous (P), Potassium (K) and Nitrogen (N)). In this research article, we have made a detailed investigation for nutrient prediction mechanism of the soil nutrient datasets taken under investigation of a specific geographic location from one of the major groundnut cultivating districts (Villupuram) in the state of Tamil Nadu and have proposed a Soil nutrients prediction scheme and optimal fertilizer recommendation model for sustainable cultivation of groundnut crop using Enhanced-1DCNN DLM. This Investigation model utilizes the natural compact robust features of 1DCNN in classifying the major macro nutrients(N,P,K)on the basis of low, Medium and high values. Based on the generated heatmap results the correlation between certain macronutrients and their corresponding micronutrient presence is classified. This proposed model has been compared for its performance and error measures with existing SVM, Naïve Bayes and ANN models and has proved to be outperforming all the compared baseline models by preserving the original data distribution with an overall accuracy of 99.78%.

Keywords—Soil nutrients; Enhc-ID-CNN DLM; nutrient classification; fertilizer recommendation

I. Introduction

Farming activities across the globe, has started taking a different dimension in its approach serving to the changing socio-economical needs. The impact of technology has already started to make inroads in the agriculture sector, several nations have sensed it and are slowly in the process of adapting precision based agricultural activities. India being one of the world's largest agriculture-based country, the scope and possibility of adapting precision-based agriculture has slowly gained importance. Groundnut is one of the most predominant oil seed, which has been cultivated in our country. This nutrious nut has been cultivated across the year In India, in three seasons, namely, the monsoon or rainy season which is called as Kharif, the winter season which is called as Rabi and the Summer. In India, one of the most important groundnut-growing states is Tamil Nadu, where groundnut is grown in five seasons: Adipattam (June-July), Karthikaipattam (October-November), Margazhipattam (December-January), Masipattam (February-March), and Chithiraipattam (March-April) (April-May). The kharif season accounts for nearly 80% of the country's total groundnut production [25][26][27][28]. As of 2018-19, the area under groundnut cultivation in Tamil Nadu was around 3.38 lakh hectares. For the best results, groundnut cultivation requires sandy loam or clay loam soil with good drainage. The soil should be deep, the pH should be around 5.5 to 7, and the fertility index should be high. The heavy soil was found to be unsuitable for cultivation due to harvesting difficulties. Because these crops are salt-sensitive, the soil should not be salty. Groundnut crop soil should not contain any rocks or clay, as this will reduce the crop's yield during harvest. For healthy germination and growth, the temperature in the cultivation area should be around 27-30 degrees Celsius. The ideal annual rainfall for crops is between 450 and 1250 millimeters. Groundnut farming is not suitable for high altitudes, cold, and frosty climates. The cultivation of groundnuts benefits from a consistent warm climate. The cultivation and production of the groundnut have been affected by many factors, one of the important issues which leads the cultivation land to be less fertile is by the usage of excessive number of artificial fertilizers on to the soil, without properly predicting the exact soil nutrient deficiency, such cultivation patterns make the soil infertile over a period of time, by thus there is gradual decline seen in the production rates of the groundnuts. Growth of the groundnut crops and higher groundnut pod formations hugely depend upon the underlying soil nutrient composition on which the crop has been cultivated. There are many macro and micro nutrients which contributes towards the productive health of the crops. The major set of macro nutrients which include (Nitrogen, Phosphorous and Potassium) NPK, defines a important role in defining the productive nutrient composition of farm in which the groundnut is being cultivated. In addition to this basic macro nutrients, calcium, Sulphur and few other micro nutrients such as iron, zinc and manganese also contributes towards production of healthy groundnut pods, which rises the overall production yield of the groundnut crop [28][29][30]. Impact of artificial intelligence by the means of Machine learning and deep learning models have started creating a unique dimension in addressing complex problems through various quantifiable solutions across diversifying domains. In this research article, we have proposed a deep learning model (Enhanced 1DCNN DLM), for predicting the various soil nutrients constituents required for the sustainable cultivation of the groundnut crop. Mathematical models developed to predict soil nutrients composition must, in general measure possible physical quantities of the environment and provides formulas to describe the actual relationships between these soil-related parameters. Because of advances in computer modelling, empirical modelling methods have emerged as a dominant development model capable of extracting contextual physical quantities related to soil nutrient composition [19][20][31][34]. This empirical model is frequently used to choose a good model, correct physical quantities using the chosen model, and validate it to see how accurate it is at predicting soil fertility. These models must take into account relevant data, such as predefined input parameters and required output parameters. The input parameters, in particular, are chosen empirically with the goal of maintaining a minimum degree of correlation between them. To build predictive models, the majority of the major soil nutrient prediction schemes that have contributed to the literature have primarily used neural networks, linear regression, machine learning, and empirical formulas [2][3][5][10][18]. Soil nutrient prediction models based on deep learning have been found to be more robust and reliable in terms of prediction accuracy at this time. There are considerable amount of research directions defined through various data mining and machine learning methodologies, towards soil nutrient based crop prediction, soil fertility prediction and for soil fertilizer recommendation [24][23][22]. Crop based Soil nutrient prediction schemes towards identifying the exact nutrient composition of the land selected for agriculture and precisive fertilizer recommendations based on the prediction made seems to be a novel approach for implementation, and through deep learning model the prediction the performance metrics will be more and error measures will be comparatively less when compared with existing machine learning models [6][7][11][12].

II. Related Work

Soil states a pivotal role towards determining yield of the crops, in recent years there where many data mining and machine learning-based investigations made on the basis of the soil nutrients availability and its proposed fertilizer recommendation schemes. These variety of machine learning based schemes have opened up a wide scope towards reaching a more optimal soil nutrient prediction scheme more specific to particular crop cultivation patterns. Various base articles, pertaining to our research problem have been reviewed comprehensively in this context [13][14][15][16][17].

For classifications, Nikam et al.[1] defined a model involving J48/C4.5, knn, ID3, Artificial Neural Network, Support Vector Machine, and Naive Bayes. These classification methods are divided into three groups: statistical, machine learning, and neural networks. [1].

Three algorithms were used by J Gholap et al.[2] to define their work. J48(C4.5) and Jrip algorithms were used with 1988 soil instances in the J48: It is a very simple classifier that generates a decision tree with a 91.90 percent accuracy. The author also suggested that a future goal would be to develop a recommendation system that would suggest appropriate fertilizer based on the soil test sample and cropping pattern.

Dr. K. Arunesh et al. [9] investigated and experimented in 203 soil instances with 6 soil attributes from Virudhunagar District, Tamilnadu, India, and found that the Nave Bayes machine learning classification algorithm outperforms J48, random tree, JRip, and OneR,

Ramesh et al. [21] proposed a system that uses Naive Bayes, Bayes Net, Naive Bayes Updatable, J48, and Random Forest as classification algorithms. They used a dataset which comprised 1500 instances of Soil samples obtained from Department of Agricultural. In the classification of soil nutrients, J48 calculated 92.3 percent accuracy, while the Nave Bayes algorithm calculated 100 percent accuracy.

Using two algorithms, Nave Bayes and J48, Chiranjeevi M N et al. [5] proposed a system for analysing soil condition and nutrients which includes Potassium, pH, Nitrogen, EC, Phosphorus, OC, Sulphur, Iron, Zinc, Magnesium, Boron, and Copper at Belagavi Department of Agriculture in Belagavi. The Naive Bayes algorithm produced a better result than the J48 algorithm, correctly classifying the determined number of instances of the soil sample.

Naive Bayes Classifier had been applied to Tirupati, Andhra Pradesh soil, according to Bhargavi et al [8]. The soil data instances were all classified into different sand categories, such as loamy sand, clay, loam, sand, sandy loam, sandy clay loam, and clay loam.

Puno, J. C., et al. [18] proposed and developed a fully functional system using IP (IP enhancement, IP segmentation, and feature extraction) in MATLAB software. All 7 nutrient attributes are classified as L, M, H, S, D (Low, Medium, High, Sufficient, Deficient) values.

After experimenting with model 1 and model 2 for soil moisture estimation, Ahmad, S et al. [4] proposed a model based on five-year data with only one attribute for classification considered, namely VIC moisture. The author concluded that SVM model outperforms ANN and MLR models.

Juhi Reshma S R et al. [23] used Neural Networks to propose a recommendation system for predicting the number of fertilisers needed for a specific banana crop, as well as regression methods for upcoming plantations. Nitrogen (N), phosphorus (p), and potassium (k) are the three most important soil nutrients for crop growth. By default, soil contains a specific amount of NPK, which varies by location. Each crop has its own set of requirements. A model is constructed in this paper to recommend the number of fertilisers required for the banana crop.

A. Extracts Inferred from the Literature

The limitations of the available soil nutrient prediction models considered towards the literature study across the recent years is drafted below.

1) Most of the available soil nutrient predicts models have been approached by data mining and machine learning algorithms, which devised the approach towards nutrient prediction, but there exist limitations in terms of learning accuracy.

2) Mostly the available machine learning schemes have not defined well defined a sustained prediction accuracy to fit the need, opening a wide scope for improvement.

3) Though there is a considerable amount of research contributions done towards predicting the soil nutrients and fertilizer recommendation, there is only a limited class of investigation done for crop specific soil nutrient prediction and fertilizer recommendation.

4) The comparison of error measures and performance metrics reached by the available ML defined soil nutrient prediction schemes extend a clear scope for improvement if learning can be further be deeply enhanced.

5) On the basis of the above drafted limitations, it has been proposed to propose an Enhanced -1D Convolutional Neural Network based Deep Learning Model (Enhanced-1DCNN DLM) to facilitate accurate estimation of soil nutrient prediction and fertilizer recommendation for the sustainable cultivation of groundnut crops.

B. Proposed Work and its Scope of Contributions

The major aids of the proposed Enhanced-1DCNN based Deep Learning Model is listed as follows:

1) The proposed model works towards the soil nutrients composition prediction and its necessary fertilizer recommendations, which aids towards the better cultivation of groundnut crops in the specific geographical location present in the Villupuram district(Tamilnadu).

2) The proposed scheme has within the substantial qualities of 1DCNN to achieve consistent and automatic extraction features which contributes towards the optimal prediction of soil nutrient composition.

3) This proposed enhanced 1DCNN model tries to address the limits which prevail in the available methods used for the process of prediction in terms of performance towards prediction, multifeatured processing capability, generalization, and prediction.

4) The proposed enhanced 1DCNN introduced layer level inner optimization and fine tuning towards the attainment of accurate soil nutrient(N,P,K) prediction.

5) Experiments of the proposed Enhanced-IDCNN DLM based soil nutrient prediction and fertilizer recommendation scheme is performed on the basis of metrics pertaining to increase in the rate of performance metrics as well as decrease in the error metrics by thus evaluating the advantages it poses in par with the baseline schemes used for the purpose of investigation.

6) The statistical and stability analysis was performed over the proposed model, which confirmed the stability of the utilized deep learning model for nutrient prediction and its respective fertilizer recommendation.

The other sections of this article are organized as drafted below. In the second section A comprehensive review was made on the available data mining and ML modelled soil-nutrient prediction schemes which have aided to the study of literature over the recent few years. The third section describes an inclusive view over the proposed Enhanced -1D CNN-scheme which predicts the soil nutrient composition of the soil dataset and recommends the suitable fertilizer adapting the layer-based feature optimization approach with appropriate validations. In the fourth section the proposed model's investigational results and its corresponding discussion pertaining towards predicting soil nutrient composition based on its metrics of performance and error measures in comparison with the baseline models has been discussed.

III. Experiment Methodology

The proposed Enhanced-one dimensional CNN based DLM for predicting the soil nutrient composition serves towards achieving sustainable groundnut cultivation through optimal prediction of soil nutrient composition and recommendation of required fertilizer need of the crop under cultivation based on the (N, P, K) input given. This Enhanced-1DCNN DLM adopted the significant parameters of various macro and micro nutrients along with soil pH in the course of predicting optimal soil nutrient composition as depicted in Fig. 1. Most commonly CNN based deep learning models are used to analyse the images, Deep two Dimensional Convolution neural networks, which might have several hidden layers and heaps of parameters, can learn objects of complex dimensions and patterns on Being trained on a large visual dataset with labelled values. This unique ability, when properly trained, defines to be the prime tool for several applications involving Two Dimensional signals such as imageries and frames of videos. However, this might not be a feasible choice in many applications involving one Dimensional signals, particularly if the training data is infrequent or confined towards any particular application. To gap this problem, 1D CNNs are proposed and it has quickly achieved the desired optimal performance levels in a variety of applications. Another striking benefit of 1D CNNs is that its configuration is simple and compact, which only perform 1D convolutions, which drives the route towards the usage of on demand cost feasible implementation of the hardware. Considering the impact of 1D CNNs in analysing temporal data, we have chosen to perform the process of soil nutrient prediction based on the optimized 1D CNN, which performs flattened layer level enhancement of conventional 2D CNNs.

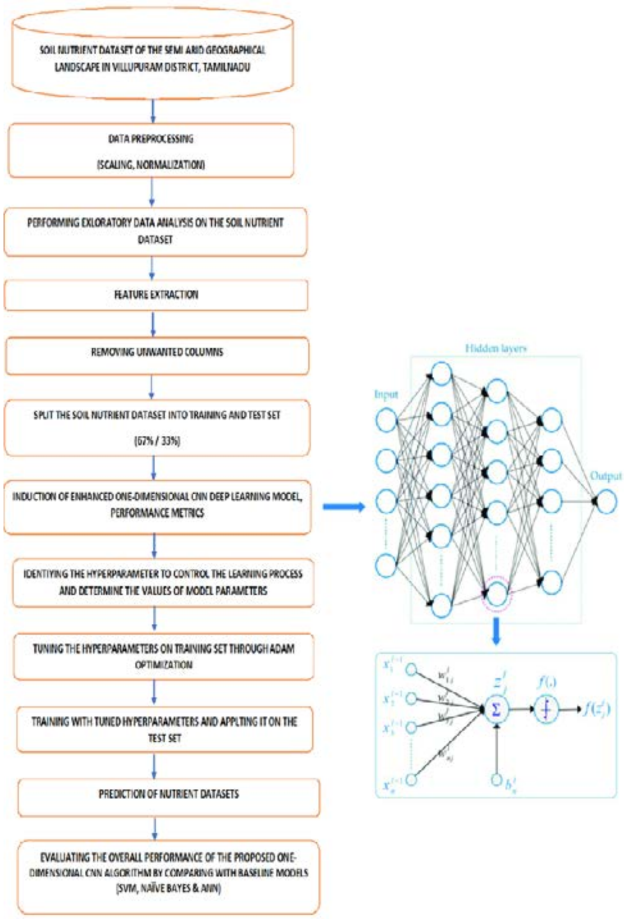


Fig. 1. Overall Process-flow Diagram for the Proposed Soil Nutrients Prediction Scheme using Enhanced 1D-CNN DLM.

A. Objective and Methodology

The cultivation of groundnut crop has been largely impacted by the improper or lack of proper assessment techniques used to exactly find the nutrient composition of the land under cultivation and needed percentage of correct fertilizers to be used based on the prediction scheme of the nutrients. The growth and the production yield of groundnut crops is largely impacted by the nutritious content of the soil, there are various set of macro and micro nutrients which contribute towards the proper growth of the groundnut crop starting from the sowing of seeds till towards the groundnut pods become mature enough to get harvested. The major set of nutrients present in the soil which contributes towards the crops growth are “Nitrogen[N], phosphorous[P] and Potassium [K]”, apart from other micro nutrient such as “Calcium (Ca), Sulphur[S], Zinc[Zn], Iron[Fe], Manganese [Mn] and Boron[B]”, soil parameters such as “pH, Soil electrical conductivity (EC), and Organic carbon (OC)” are also responsible in fixing out the overall fertility index of the soil. In this article we have chosen a semi-arid geographical landscape which is the most conducive soil pattern for the cultivation of groundnut and performed the model evaluation based on the nutrient datasets obtained from one such. The above given Table I, defines the parameters involved in the evaluation of the proposed Enhanced -1D CNN DLM.

TABLE I. PARAMETERS INVOLVED IN THE EVALUATION OF PROPOSED ENHC-1D CNN DLM

Geographical location of the dataset	Gingee (Taluk), Villupuram (District.), Tamilnadu (State), India
Latitude	12.2529° N
Longitude	79.4160° E
Source of Dataset	https://soilhealth.dac.gov.in/NewHomePage/NutriReport
Macro Nutrients	“Nitrogen[N], Phosphorous[P] and Potassium [K]”
other Nutrients	“Calcium (Ca), Sulphur[S], Zinc [Zn], Iron [Fe], Manganese [Mn] and Boron[B]”
Soil core parameters	“pH, Soil electrical conductivity (EC), and Organic carbon (OC)”
Nutrient classification classes on the datasets	Low, Medium & High

Excessive usage of artificial fertilizers, pollution variants and changing weather conditions make a huge impact by downgrading the quality of the soil under cultivation, and hence Prediction of the soil nutrient composition stays to be a vital factor towards framing up a well-defined precision farming prototype to uphold the soil fertility index and the crop production yield.

In this research work, a proposed 1D-CNN based Deep learning model has been chosen to train the nutrient datasets conducive for the sustainable cultivation of groundnut crop. This proposed 1D CNN deep learning model serves to the need of bridging the gap between the exact soil nutrient prediction for the cultivation of groundnut crop and its production rates through the improved optimal soil fertilizer recommendation schemes.

The one-dimensional CNN model can aggregate local features and lessen the data dimensions through convolutional learning and pooling operations. As a result, by repeatedly using convolution operations and pooling operations, a deep convolutional neural network can extract high-level features while significantly reducing the dimension of the output. In this article, raw nutrient dataset is used directly as the input to deep neural networks, as a result, the timing of the deep convolutional neural network output's high-level features is not disrupted.

B. Evaluation of the Proposed Model with Enhanced One-Dimensional CNN

A 1D CNN is a special case of a conventional neural network. Unlike traditional neural networks, in which the hidden layer is fully connected, a 1D CNN employs a unique network structure that alternates between the convolution and the pooling layers.

As shown in the above Fig. 2, the proposed 1D CNN has an input layer, three layers of convolutional (C_1, $C_{2,}$, C_3), fully connected two layers (F1, F2) & an output layer. This 1D CNN stays to be an alternate enhanced version of 2D CNN.

These enhancements have proven to be more effective in certain applications which deal with 1D temporal data.

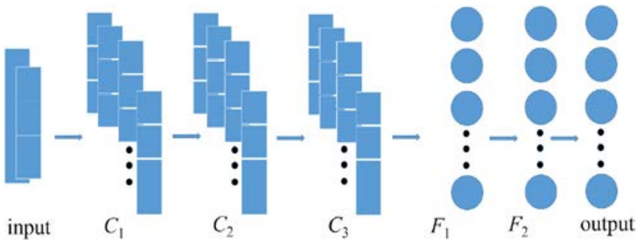


Fig. 2. Single CNN (1DS) Architecture.

Comparing with the 2D CNN, this 1DCNN are chosen to be more advantageous stating for the following reasons.

- The forward and the back propagation in one dimensional CNN require a simple array operation rather than a matrix operation.
- Due to its shallow Architecture, it involves in complex learning capability of ID temporal data comparing with 2D CNN which has a deeper architecture, 1D-CNN are much convenient to train and use.
- The Complexity of involving more advanced hardware setup which includes the involvement of GPUs and cloud infrastructure in 2D-CNN is hugely reduced with respect to 1D-CNN where, general CPU implementation with a relatively fast processing speed makes 1D-CNN more opt for the usage.
- Compact 1DCNN have proven to be efficient in terms of performance pertaining to concise datasets.

As shown in the above Fig. 2, the 1D CNN Architecture consists of two distinct type of layers they are the "CNN-layers" in one Dimensional convolutions and pooling, and Fully-connected layers. The arrangement of a 1D-CNN is designed by the following hyper-parameters:

- The total levels of hidden CNN and fully connected layers (in our proposed Enhanced 1D-CNN model as shown in the above Fig. 2, there are 3 CNN layers and 2 fully connected layers).
- Defining the size for Filters in each CNN layer.
- Subsampling feature in each CNN layer.
- The choice of pooling and activation functions.

The input layer in 1D-CNN stays to be an inert layer which accepts the raw 1D temporal data as that of the conventional 2D-CNN. The output layer is a fully connected layer consisting of equal number of neurons as that of the number of classes. Fig. 3 depicts three consecutive CNN layers of 1D-CNN. As shown in the figure, 1D filter kernel has a size of 3 and a subsampling feature of 2. Here, the kth neuron in the CNN hidden layer l first performs a series of convolutions, the sum of which passes through the activation function *f* by a subsampling operation. In fact, this is the main difference between 1DCNN and 2DCNN, where the 1D array replaces the 2D matrix in both the kernel map and the feature map**.** Processing further, the raw 1D datasets are processed by the CNN layers and it starts learning to extract the features potential for the purpose of classification to be performed by the Fully connected layers, as a result the process of feature extraction and classification are coupled together as a single process that which can be optimized so as to maximize the performance of the classification. This process of 1D CNN proves to be a major advantage as it results in the involvement of low computational complexity, since the only major operational cost is with the sequence of one dimensional convolutional which performs simply the linear weighted summing of two 1D arrays, which can be operated effectively during the forward and Back-Propagation operations in parallel.

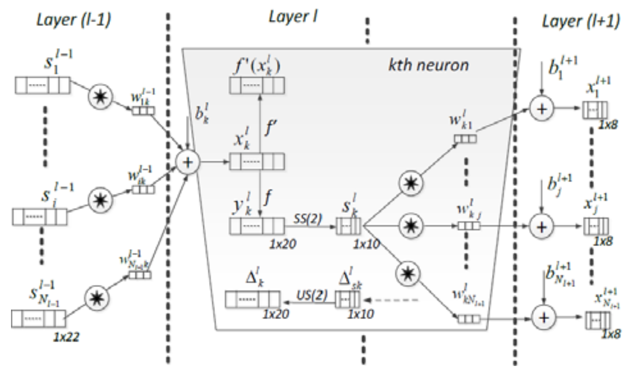


Fig. 3. Three Consecutive Hidden CNN Layers of a 1D CNN [33].

The procedure [33] of 1D Forward propagation (1D-FP) in each CNN-layer is shown in the equation (1) and the process of Back propagation is summed up as shown in the equation (2), is defined as follows:

$$x_k^l = b_k^l + \Sigma_{i=1}^{N_{l-1}} \mathrm{conv1D}(w_{ik}^{l-1}, s_i^{l-1}) \quad (1)$$

After the computation of weight and bias, they can be utilized to update the biases and weights with the learning factor, ε as,

$$w_{ik}^{l-1}(t+1) = w_{ik}^{l-1}(t) - \varepsilon \frac{\delta E}{\delta w_{ik}^{l-1}}$$

and

$$b_k^l(t+1) = b_k^l(t) - \varepsilon \frac{\delta E}{\delta b_{ik}^l} \quad (2)$$

The process of forward and back-propagation in hidden 1D CNN layers is depicted below in the Fig. 4.

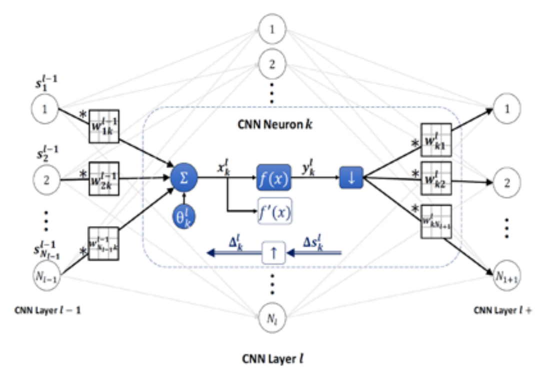


Fig. 4. CNN (Forward and Back-propagation in Hidden Layers).[33].

The process flow of the Back Propagation for the one-Dimensional temporal datasets in the training set can be stated as follows:

1) To initialize the weights and biases of the CNN.

2) For each Back Propagation iteration DO.

3) For each Nutrient Composition Value (Low, Medium, High) in the dataset, DO:

4) FP: Forward propagation from the starting input level layer through the output layer to find the outputs of each neuron.

5) BP: By Computing the delta error at the output level layer and to back-propagate it to first hidden layer to compute the delta errors,

6) PP: To Post-process by computing the weight and bias sensitivities.

7) Update: Updating the weights and biases by the (accretion of) sensitivities scaled up with the learning factor.

IV. Experimental Results and Discussion

The evaluation of the proposed model using Enhanced 1D-CNN scheme and the benchmarked SVM, Naïve bayes and ANN schemes have been conducted based on the nutrient datasets of the Villupuram district obtained for a particular geographical region of Gingee. The dataset used for the current investigation is available online and it pertains to the years starting from 2016 to 2020. “Precision, recall score, F1-Measure, and Accuracy”, as well as the error measures “RMSE, MAD, MAE, and R^2” is calculated, and a comparison to the baseline models is derived through graphical comparisons for the proposed Enhanced ID-CNN. The datasets are mounted on Google Drive, and the model is evaluated on the Google Colab platform.

The training process of the proposed Enhanced ID-CNN structure is attained on the basis of Villupuram district nutrient dataset which are obtained from the online GIS servers, which are referred to the India soil Health System (soilhealth.dac.gov.in) maintained by the Indian Council of Agricultural Research. This web portal plays a prime role for providing soil related parameters across all the states in the country district wise. The soil nutrient constituents taken for the training purpose is obtained from the Tamilnadu state wise data of Villupuram district which spans over four years from (2016-2020). Data pre-processing is carried out before introducing into the model. The dataset is then partitioned based on the ratio of 67:33 for Training and Testing, respectively. Then the pre-processed data is inducted with the model evaluation of Enhanced 1D-CNN, during the process the model is evaluated for its performance measure and error metrics. The process of identifying the hyperparameters to control the learning process and to determine the values of model parameters is done by tuning the hyperparameters on the test set.

The proposed Enhanced 1D-CNN Deep learning model for predicting Soil nutrient composition for the sustainable cultivation of the groundnut crop is a nascent Deep learning approach of this kind, hence for the purpose of evaluating the performance of the proposed Enhanced 1D-CNN model and its optimality consideration based on overall performance metrics and error measures it has been compared with standard and effective machine learning classifiers such as SVM, Naïve Bayes and ANN models and proved to be outperforming based on the evaluation results. The recommendation module evaluates the process of taking the given range of (N, P & K) values and generates the optimal fertilizers to be used for the sustainable cultivation of the groundnut crop. Since the general soil composition of the Villupuram district soil is Nitrogen Low by nature, the low predictions of Nitrogen below the minimum 17% of the basal need will have a nitrogen-based fertilizer recommendation for the sustainable cultivation of the groundnut crop, with respect to Phosphorous (P) based fertilizer recommendation, the higher percentage of phosphorous deposition found in Villupuram district soils along with the amount of needed phosphorous per hectare of groundnut cultivation demands a minimum of 35% of phosphorous requirement, which normally get satisfied due to its natural existence in the soil and hence rarely needs phosphorous based fertilizer recommendations. The requirement of Phosphorous is very much essential during the basal and flowering stages of the groundnut crops. The most important nutrient with respect to groundnut cultivation is the presence of Potassium(K), which is the most desirable nutrient required for the sustainable groundnut cultivation, starting from its early stage of growth till to its maturity, because it is responsible for making the crop disease resistive, regulates the water conditions within the plant cell, aids the crop in formation of proteins and chlorophyll and even often counterattacks the negative impacts of excess nitrogen supplements, supplied through fertilizers. To a minimum the percentage of Potassium will be nearly 55% in the total nutrient requirement for the sustainable cultivation of the groundnut crop, since the deposition of Potassium in the district of Villupuram seems to be in the Medium scale of level, fertilizers pertaining to Potassium are normally recommended when the predicted or the given values of Potassium is below the minimum scale required. The below Fig. 5, depicts the evaluation of the fertilizer recommendation module based on the given value of prediction.

Fig. 5. Fertilizer Recommendation Screen Shot.

A. *Consideration of Performance Measures*

- Accuracy: Adaptation of 1D-CNN has shown significant performance improvement in terms of the overall evaluation accuracy of the model in terms of classification. Comparative depiction of the accuracy of the proposed model is shown in Fig. 6.

$$ACCURACY = \frac{T_P+T_N}{T_P+T_N+F_P+F_N} \quad (3)$$

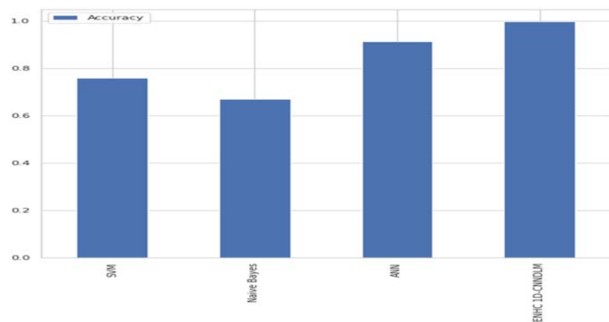


Fig. 6. Comparison of the Overall Prediction Accuracy of the Proposed Enhanced 1D-CNN.

- Precision: The overall Precision of the proposed Enhanced 1D-CNN described below is the ratio between the number of correctly identified positive predictions and the total number of positive predictions (True positive + False positive). Comparative depiction of the precision of the proposed model is shown in Fig. 7.

$$PRECISION = \frac{T_P}{T_P+F_p} \quad (4)$$

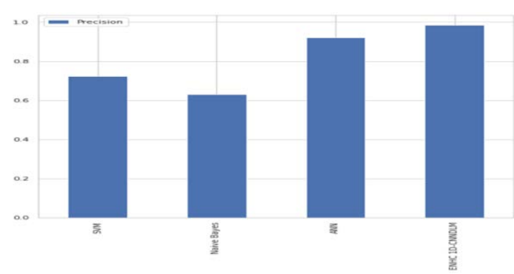


Fig. 7. Comparison of Precision Value for the Proposed Enhanced 1D-CNN.

Recall: As shown in Fig. 8, the recall value for the proposed Enhanced 1D-CNN is defined as the ratio of actually predicted true positive values to the overall sum of true predicted positive and false negative values.

$$\text{RECALL} = \frac{\text{TP}}{TP+FN} \quad (5)$$

F1-Measure: The harmonic mean of precision and recall is the F1 measure for the proposed Enhanced 1D-CNN, and its performance against the baseline model is depicted in Fig. 9 below.

$$F1 = 2 * \frac{PRECISION*RECALL}{PRECISION+RECALL} \quad (6)$$

e. Loss: This proposed Enhanced 1D-CNN Deep learning Model outputs a very minimal loss value by thus producing a more improved accuracy rate, The below Fig. 10 depicts it.

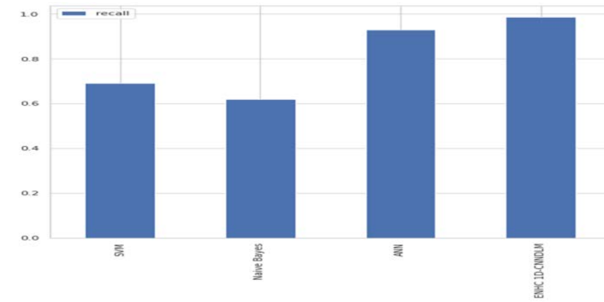


Fig. 8. Comparison over the Recall Score of the Proposed Enhanced 1D-CNN with Baseline Models.

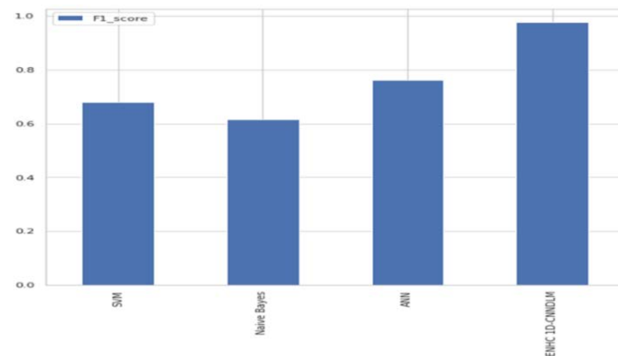


Fig. 9. Comparison over the F1-Measure of the Proposed Enhanced 1D-CNN with Baseline Models.

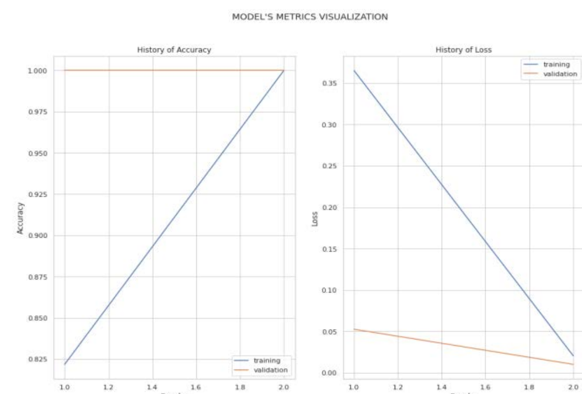


Fig. 10. Derived Accuracy and Loss of the Proposed Enhanced 1D-CNN.

The overall performance metrics of the proposed Enhanced 1D-CNN is depicted in Table II and it corresponding graphical representation is shown in the below Fig. 11.

TABLE II. PERFORMANCE METRICS OF THE PROPOSED ENHANCED 1D-CNN WITH THE BASELINE MODELS

	SVM	NB	ANN	ENHC ID-CNNDLM
LOSS	0.23450	0.33670	0.9890	**0.02347**
ACCURACY	0.75894	0.66894	0.9124	**0.99789**
PRECISION	0.72345	0.62971	0.9197	**0.98524**
RECALL	0.69080	0.61987	0.9298	**0.98741**
F1_SCORE	0.67890	0.61620	0.7620	**0.97745**

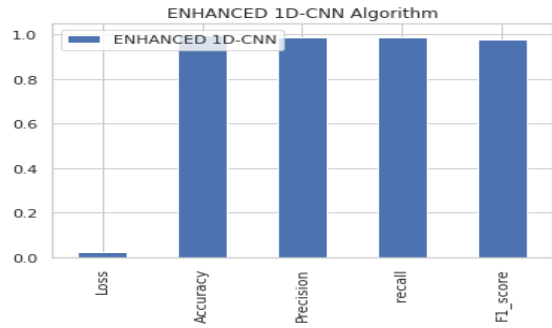


Fig. 11. Overall Performance Metrics of the Proposed Enhanced 1D-CNN.

Comparison of the Proposed Enhanced 1D-CNN with Baseline Models based on Performance measures, the below Fig. 12 depicts the optimality of the proposed model over other compared models.

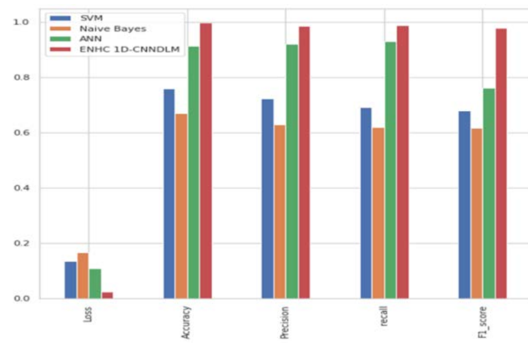


Fig. 12. Comparison of the Proposed Enhanced 1D-CNN with Baseline Models based on Performance Metrics.

B. Consideration of Error Metrics

The square root of the difference between the predicted values of the used model and the actual values associated with the study variable (soil nutrient) determined over the total number of observations is known as the Root of the Mean of the Squared Errors (RMSE) [32].

$$RMSE = \sqrt{\frac{ACT_{Val} - PR_{Val}}{Obs_{No}}} \tag{7}$$

MAE: The difference between the predicted values of the used model and the actual values associated with the study variable (soil nutrient) determined over the total number of observations is depicted by the mean of absolute values [32].

$$MAE = \frac{ACT_{Val} - PR_{Val}}{Obs_{No}} \tag{8}$$

Where, ACT_{Val} and PR_{Val} represents the actual and predicted values and Obs_{No} is the number of observations.

MAD: The average distance between each data point and the mean is the mean absolute deviation of a dataset. It gives us an idea of how variable a dataset is. [32].

$$\text{MAD} = \frac{\sum Absolute\ values\ of\ deviation\ from\ central\ measures}{Total\ .no.of.Observations} \tag{9}$$

The shown Table III depicts the calculated error metrics in par with all the baseline models with the Enhanced 1D-CNN and the corresponding graphical representation as shown in Fig. 13 shows the performance of the Enhanced 1D-CNN with a minimal loss in the process of prediction and hence optimality in terms of achieving low error metrics. The calculated Coefficient of Determination (R^2) Value of the proposed Enhanced 1D-CNN scheme stands ahead in terms of evaluation with all the baseline models, and thus denotes how well the coefficient fits with the values in the training dataset.

TABLE III. COMPARISON OF THE ERROR METRICS COEFFICIENT OF DETERMINATION OF THE PROPOSED ENHANCED 1D-CNN WITH THE BASELINE MODELS

	SVM	NAÏVE BAYES	ANN	1D-CNNDLM
RMSE	0.8092	0.8872	0.7950	**0.4839**
MAE	0.6549	0.7872	0.6321	**0.2342**
MAD	0.3238	0.6871	0.2871	**0.2270**
R^2	0.6123	0.5321	0.9312	**0.9872**

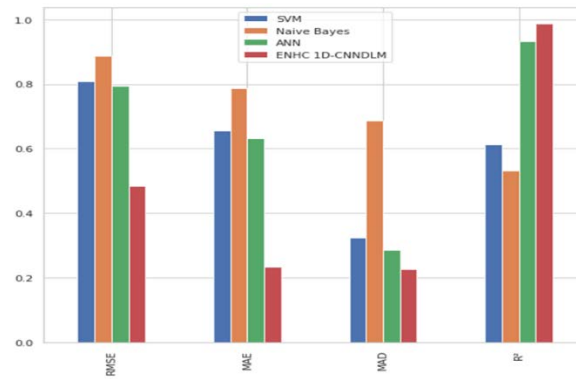


Fig. 13. Error Metrics and Coefficient of Determination (R²) Comparison of the Proposed Enhanced 1D-CNN with Baseline Models.

The proposed Enhanced 1D-CNN schemes' prediction efficiency is tested using the Villupuram district soil nutrient dataset by deriving the confusion matrix (see Table IV (A, B, C)). The computed confusion matrix for the proposed model, has shown evident prediction results by producing most optimal prediction accuracy rates for predicting the major macro nutrients of the soil Nitrogen (N), Phosphorous (P) and Potassium (K), based on the classification sets Low, Medium and High. This classification labels are defined on the basis of the nutrient dataset values, which represents the exact soil nutrient composition of these nutrients in the district of Villupuram.

The proposed Enhanced 1D-CNN model and the compared benchmarked schemes go through a three-step experimental process. Initially, the proposed Enhanced 1D-CNN model along with its benchmarked comparative models are compared for the performance measures such as (Accuracy, Precision, Recall, and F1-score) as well as error measures (RMSE, MAD, MAE). The model is also evaluated for its coefficient of determination value (R2) pertaining to the Villupuram district soil nutrient dataset. Figure 11 shows the performance of the proposed Enhanced 1D-CNN scheme and the benchmarked SVM, Nave Bayes, and ANN models in terms of mean accuracy and precision. The proposed Enhanced 1D-CNN deep learning architecture ensured maximum overall accuracy and precision, as the speed with which it deduced robustness features that could potentially influence the prediction of soil Nutrient Composition was way ahead when compared with other learning models. As a result, the proposed Enhanced 1D-CNN scheme has the potential to improve accuracy by23.89 per cent, 32.89 per cent, and 8.54 per cent over the baseline models SVM, Nave Bayes, and ANN, respectively.

TABLE IV. DERIVATION OF CONFUSION MATRIX DEPICTING THE PREDICTING EFFICIENCY OF (N, P, K) VALUES BY THE PROPOSED ENHANCED 1D-CNN, (A) CONFUSION MATRIX FOR PHOSPHOROUS CLASSIFICATION, (B) CONFUSION MATRIX FOR NITROGEN CLASSIFICATION, (C) CONFUSION MATRIX FOR POTASSIUM CLASSIFICATION

(A)

ACTUAL PHOSPOROUS (P)		**PREDICTED PHOSPOROUS (P) COMPOSITION**			
		LOW	**MEDIUM**	**HIGH**	**∑(LMH)**
	LOW	378	1	0	379
	MEDIUM	0	592	2	594
	HIGH	7	0	580	587
	∑(LMH)	385	593	582	1560

(B)

ACTUAL NITROGEN (N) COMPOSITION		**PREDICTED NITROGEN (N) COMPOSITION**			
		LOW	**MEDIUM**	**HIGH**	**∑(LMH)**
	LOW	1260	1	0	1261
	MEDIUM	0	260	1	261
	HIGH	7	0	31	38
	∑(LMH)	1267	261	32	1560

(C)

ACTUAL POTASIUM (K) COMPOSITION		**PREDICTED POTASIUM (K) COMPOSITION**			
		LOW	**MEDIUM**	**HIGH**	**∑(LMH)**
	LOW	624	1	0	625
	MEDIUM	0	770	2	772
	HIGH	7	0	156	163
	∑(LMH)	631	771	158	1560

Fig. 14. Heat Map Depicting the Correlations of the Soil Dataset.

The correlation between the variables is shown in Fig. 14 using a heat map. Correlation matrix that shows the relationship between two parameters in a dataset. The heat map is used in exploratory data analysis to check for correlations between the data. The heat map is used to visualize the parameter correlation matrices as well as to determine which parameters influence the output variable. The level of correlation between the various macro and micro nutrients, as well as pH, Electrical Conductivity (EC), and organic carbon (OC), has been absorbed in the heat map shown above. There is a strong correlation between the macro nutrient (Nitrogen) and the micro nutrient (Calcium), which is essential for the growth of groundnut crops during the time of pod formation, as shown in Fig. 14, but there are also instances of weak correlation. The value of heat map ranges from +1 to -1 where, Positive values indicate positive correlation, while negative values indicate negative correlation. A stronger linear association exists between data points that are closer together, whereas a weaker linear association exists between values that are closer to zero.

V. CONCLUSION

The prime reason for the loss of soil quality is due to the improper soil and crop management strategies deployed in the process of farming. Excessive usage of chemical fertilizers without the exact knowledge of required nutrients for the cultivation of crops has led to the gradual decline in the soil quality which has contributed towards the gradual decline in the production yield of the crops. Groundnut Crops being the most predominant oilseed crop cultivated in the state of Tamil Nadu, its cultivation has seen too such impacts due to the excessive fertilizer usage and improper cultivation patterns without understanding the actual nutrient of the soil, one of the major causes for the fluctuation seen in groundnut pod growth patterns and production, is the differing soil nutrient compositions of the land which is under cultivation.In this research article we proposed a novel deep learning based approach adapting Enhanced 1D-CNN scheme, towards predicting the soil nutrient composition for the choosen soil nutrient dataset pertaining to the geographical landscape of Villupuram district in the state of Tamil Nadu. The experimental results carried out has clearly shown the effectiveness of the proposed Enhanced 1D-CNN DLM, in terms of the overall performance measures (Accuracy, Precision, Recall, F1 Score, Recall and a very minimal Loss value), the optimality of the proposed model has been compared for its effectiveness with baseline models such as (SVM, Naïve Bayes and ANN) and proven to be outperforming all the baseline models in terms of increased performance measures resulting in the overall prediction accuracy of 99.78% and very minimal error measures. The fertilizer recommendation based on the predicted nutrient composition of the soil makes this proposed model more productive in terms of addressing the objective of this research work. This proposed scheme for the sustainable cultivation of groundnut crop may be considered as a reference scheme for crop specific precision farming moving further.

References

[1] S.NIKAM. (2015). A Comparative study of classification Techniques in Data Mining Algorithms. oriental journal of computer science & technology, 8(1), 13-19.

[2] Jay Gholap, Anurag Ingole,, Shailesh Gargade, Jayesh Gohil, & Vahida Attar. (2012). Soil Data Analysis Using Classification Techniques and Soil Attribute Prediction. ResearchGate-2012.

[3] Adhikary, T., Das, A. K., Razzaque, M. A., Chowdhury, M. E., & Parvin, S. (2015). Test implementation of a sensor device for measuring soil macronutrients. 2015 International Conference on Networking Systems and Security (NSysS). doi:10.1109/nsyss.2015.7042951.

[4] Ahmad, S., Kalra, A., & Stephen, H. (2010). Estimating soil moisture using remote sensing data: A machine learning approach. Advances in Water Resources, 33(1), 69-80. doi:10.1016/j.advwatres.2009.10.008.

[5] Chiranjeevi M N, Ranjana B Nadagoudar(2018)Analysis of Soil Nutrients using Data Mining Techniques. International Journal of Recent Trends in Engineering and Research, 4(7), 103-107. doi:10.23883/ijrter.2018.4363.pdt1c.

[6] Antonios Morellos a, Xanthoula-Eirini Pantazi a, Dimitrios Moshou a, Thomas Alexandridis c, Rebecca Whetton b, Georgios Tziotzios a, & Jens Wiebensohn d. (2016). Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectros- copy. Published by Elsevier. doi:10.7287/peerj.5714v0.2/ reviews/1.

[7] Balducci, F., Impedovo, D., & Pirlo, G. (2018). Machine Learning Applications on Agricultural Datasets for Smart Farm Enhancement. Machines, 6(3), 38. doi:10.3390/machines6030038.

[8] Bhargavi, P., & Jyothi, S. (2009). Applying Naive Bayes Classification Technique for Classification of Improved Agricultural Land soils. IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.8, August 2009, 9(8), 117-122.

[9] Dr. K. Arunesh, & V. Rajeswari. (2017). agricultural soil lime status analysis using data mining classification techniques. International Journal of advanced technology in engineering and science, 5(2), 27-35. Retrieved from www.ijates.com.

[10] Ingale, V., Vaidya, R., Phad, A., & Shingare, P. (2016). A sensor device for measuring soil macronutrient proportion using FPGA. 2016 Interna- tional Conference on Communication and Signal Processing (ICCSP). doi:10.1109/iccsp.2016.7754237.

[11] Kulkarni, K., & Vijaya, P. A. (2017). A comparitive study of land classification using remotely sensed data. 2017 International Conference on Computing Methodologies and Communication (ICCMC). doi:10.1109/iccmc.2017.8282720.

[12] Leisa J. Armstrong, Dean Diepeveen, & Rowan Madder. (n.d.). The application of data mining techniques to characterize agricultural soil profiles.

[13] Mathavi Parvathi, T. (2017). automated soil testing process using combined mining patterns. best: International Journal of Humanities, Arts, Med- icine and Sciences (BEST: IJHAMS), 5(8), 76-79.

[14] Mrs. N. Hemageetha, & Dr. G.M. Nasira2. (2016). Analysis of Soil condition Based on pH value Using Classification Techniques.

[15] Panchamurthi, S. (2019). Soil Analysis and Prediction of Suitable Crop for Agriculture using Machine Learning. International Journal for Re- search in Applied Science and Engineering Technology, 7(3), 2328-2335. doi:10.22214/ijraset.2019.3427.

[16] Patil, T. R., Shamshuddin K, Patil, R., & Sadanand P. (2016). Krushi Samriddhi: A decision support system for farmers to get high crop yield. 2016 International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT). doi:10.1109/ic- ctict.2016.7514548.

[17] Pawar, M., & Chillarge, G. (2018). Soil Toxicity Prediction and Recommendation System Using Data Mining In Precision Agriculture. 2018 3rd International Conference for Convergence in Technology (I2CT). doi:10.1109/i2ct.2018.8529754.

[18] Puno, J. C., Sybingco, E., Dadios, E., Valenzuela, I., & Cuello, J. (2017). Determination of soil nutrients and pH level using image processing and artificial neural network. 2017IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM). doi:10.1109/hnicem.2017.8269472.

[19] Rahman, S. A., Chandra Mitra, K., & Mohidul Islam, S. (2018). Soil Classification Using Machine Learning Methods and Crop Suggestion Based on Soil Series. 2018 21st International Conference of Computer and Information Technology (ICCIT). doi:10.1109/iccitechn.2018. 8631943.

[20] Rajeswari, V., & Arunesh, K. (2016). Analysing Soil Data using Data Mining Classification Techniques. Indian Journal of Science and Technol- ogy, 9(19). doi:10.17485/ijst/2016/v9i19/93873.

[21] Ramesh, V., & Ramar, K. (2011). Classification of Agricultural Land Soils: A Data Mining Approach. Agricultural Journal, 6(3), 82-86. doi:10.3923/aj.2011.82.86.

[22] Regalado, R. G., & Dela Cruz, J. C. (2016). Soil pH and nutrient (Nitrogen, Phosphorus and Potassium) analyzer using colorimetry. 2016 IEEE Region 10 Conference (TENCON). doi:10.1109/tencon.2016. 7848458.

[23] JuhiReshma S R, Dr. D. John Aravindhar (2021). Fertilizer Estimation using Deep Learning Approach. Nat.Volatiles&Essent.Oils,2021;8 (4):5745-5752.

[24] Anguraj.K, Thiyaneswaran.B, Megashree.G, Preetha Shri.J.G, Navya.S, Jayanthi. J Crop(2021). Recommendation on Analysing Soil Using Machine Learning, Turkish Journal of Computer and Mathematics Education ,Vol.12 No.6 (2021), 1784-1791.

[25] apeda.gov.in/apedawebsite/HACCP/2018_Groundnut_Survey_Report.p df.

[26] agritech.tnau.ac.in/agriculture/agri_seasonandvarieties_groundnut.html.

[27] apeda.gov.in/apedawebsite/SubHead_Products/Ground_Nut.htm.

[28] https://www.ikisan.com/tn-groundnut-history.html.

[29] Soilhealth.dac.gov.in/PublicReports/MacroNutrientsStateWiseNPK.

[30] Efficiency Of Groundnut Cultivation in Tamil Nadu, Dr. S. Sudhamathi, Science Technology and Development Volume VIII Issue XII DECEMBER 2019 ISSN: 0950-0707, Page No: 361.

[31] R. Jayalakshmi, M. Savitha devi, Predictive Model Construction for Prediction of Soil Fertility using Decision Tree Machine Learning Algorithm, INFOCOMP, v. 20, no. 1, p. 49-55, June, 2021.

[32] Sivasankaran, S., K. Jagan Mohan, and G. Mohammed Nazer. "Soil Moisture Quantity Prediction using Optimized Deep Learning Supported model for Sustainable cultivation of Groundnut plant." NVEO-NATURAL VOLATILES & ESSENTIAL OILS Journal| NVEO (2021): 13184-13205.

[33] S. Kiranyaz, A. Gastli, L. Ben-Brahim, N. Alemadi, M. Gabbouj, Real-Time Fault Detection and Identification for MMC using 1D Convolutional Neural Networks, IEEE Trans. Ind. Electron. (2018). doi:10.1109/TIE.2018.2833045.

[34] Dietterich, Thomas G. "Approximate statistical tests for comparing supervised classification learning algorithms." Neural computation 10.7 (1998): 1895-1923.

Performance Evaluation of Different Supervised Machine Learning Algorithms in Predicting Linear Accelerator Multileaf Collimator Positioning's Accuracy Problem

Hamed S. El-Ghety[1], Ismail Emam[2], AbdelMagid M. Ali[3]
Electrical Engineering Department, Faculty of Engineering, Aswan University, Aswan, Egypt[1, 3]
Oncology department, Faculty of Medicine, Ain Shams University, Cairo, Egypt[2]

***Abstract*—Radiation Oncology is one of the businesses that employs Machine Learning to automate quality assurance tests so that errors and defects can be reduced, avoided, or eliminated as much as possible during tumor therapy using a Linear Accelerator with MultiLeaf Collimator (Linac MLC). The majority of Machine Learning applications have used supervised learning algorithms rather than unsupervised learning algorithms. However, in most cases, there is a clear bias in deciding which supervised machine learning algorithm to use. And prediction findings may be less accurate as a result of this bias. As a result, in this study, an evidence is presented for a novel application of Logistic Regression technique to predict Linac MLC positioning accuracy, which achieved 98.68 percent prediction accuracy with robust and consistent performance across several sets of Linac data. this evidence was obtained by comparing the performance of various supervised machine learning algorithms (i.e. Logistic Regression, Decision Tree, Support Vector Machine, Random Forest, Naive Bayes, and K-Nearest Neighbor) in the prediction of Linac MLC's positioning accuracy problem using leaves' positioning displacement datasets with labelled results as training and test datasets. For each method, two parameters were used to evaluate performance: prediction accuracy and the receiver operating characteristics curve. Based on that evaluation, the right selection sequence was proposed for supervised Machine Learning algorithms in order to achieve near-optimal prediction performance for Linac MLC's leaf positioning accuracy problem. As a result, the selection bias, as well as the negative side effects (i.e. ineffective preventive maintenance plan for Linac MLC to avoid and solve causes of inaccurate leaf displacement such as motor fatigue and stuck problems) could have occurred were successfully avoided.**

Keywords*—Linear accelerators; logistic regression; performance evaluation; prediction methods; supervised learning

I. Introduction

Machine learning applications have been utilized in different industries including Radiation Oncology [6]. In Radiation Therapy, some researches summarizes potential various clinical applications such as head, neck, lung, and prostate cancer as well as radiation toxicity [1][2][3]. Other researches states that differences between planned and actual displacements of multi-leaf collimators (MLCs) are source of errors in dose distributions during radiotherapy [4]. However, Radiation Therapy is still considered niche area with big crude data that needs extensive use of machine learning applications. And since the precision medicine in radiation oncology, radiation toxicity and complication factors are inevitable conditions for oncology patients after radiotherapy [1][4][5] and since most of time the use of popular supervised learning algorithms (e.g. Support Vector Machine and Decision Tree) are supported by previous prediction accuracies in other industries regardless of differences in nature of the data itself which is considered a selection bias that may produce less accurate prediction results. So this paper focuses on performance evaluation of different popular supervised learning algorithms in the prediction of leaf displacement accuracy problem utilizing Linear Accelerator with Multi-Leaf Collimator (Linac MLC) by comparing two Criteria factors: the prediction accuracy of the algorithm and the corresponding receiver operating characteristics curve.

This work will help researchers tackling similar Linac MLC prediction problems with the same nature of displacement data to use logistic regression technique confidently to get near-optimal prediction. At the same time, this work will guide researchers in other business areas as well to follow the same evaluation process practice that is undertaken in this paper, prior to using a typical supervised learning algorithm with a typical data of certain nature, by this way, they can properly select the most suitable supervised learning algorithms that gives near-optimal prediction.

As follows, this paper will have seven remaining sections: Methods for supervised learning; Using supervised learning in Linac MLC; Methodology; Implementation; Results and discussion; Conclusion; Acknowledgment; References.

II. Methods for Supervised Learning

This section gives a brief background on different supervised machine learning algorithms.

A. Decision Tree (DT)

In machine learning, DT is one of the most useful and reliable classifiers. The decision tree has a hierarchical design that employs the divide-and-conquer strategy [7]. As a result, it can be used for classification. And reduced to a series of simple if-then statements [14].

B. Support Vector Machine (SVM)

SVM is a supervised learning algorithm. SVM is used for classification in many applications. Using the margin concept, the Linear Support Vector Classifier determines an optimal separating hyper-plane. This represents the distance between the hyper-plane and the nearest points to it on either side, and can be maximized for better generalization [8].

C. Random Forest (RF)

RF classifier is composed of several DTs, like how many trees build up a forest. Deep Decision trees frequently overfit the training data, which means that any minor change in the given data will produce a large variance in classification results. In other words, nature of training data makes them more likely to give wrong predictions with the test dataset. Random forest's decision trees needs to be trained using different portions of training dataset [9]. To classify a new sample, the sample's input vector must be passed down through the forest with each DT. Following that, each DT considers a different section of the input vector to determine the classification conclusion. The forest then decides whether to use the classification with the most 'votes' (for discrete classification outcomes, such as the MLC case study used in this research) or the average of all trees in the forest (for numeric classification outcome). Because the RF algorithm considers the results of multiple DTs, it can reduce the variance caused by considering only one DT for the same dataset [9].

D. Logistic Regression (LR)

LR is a normal type of regression where two state variables can be modelled easily. Thus, it helps to determine the likelihood that a new sample is associated with a typical class. And if it's used to classify binary samples, an input sample with a probability value greater than 0.50 is classified as 'class A'; otherwise, it is classified as 'class B [10].

E. Naïve Bayes (NB)

The NB classifier is a categorization strategy that computes the likelihood of an event based on prior knowledge of the event's conditions. Despite the fact that features in a class may be interdependent, so it considers that an item in that class is not directly associated to any other items [11].

F. K-Nearest Neighbor (KNN)

KNN classifier involves using a database to classify unknown cases. The observations are displayed in a three-dimensional space, with the number of qualities or features that each observation possesses indicated. Based on its similarity to other data points in the model, a new point is classified using some similarity measures [15]. KNN determines the new point's class by selecting the K closest points to the new example and voting for the most frequent class among them to be the new point's class, and so on, where K is the number of neighbors [8]. Fig. 1 illustrates the KNN method with k=1.

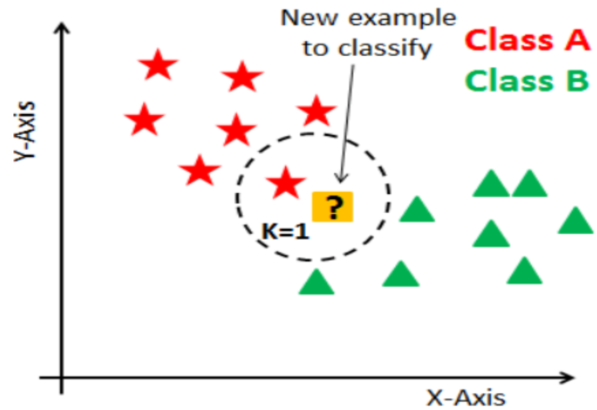


Fig. 1. Illustaration of K-Nearest Neighbor Method with k =1.

III. Using Supervised Learning in LINAC MLC

Because Modern radiotherapy procedures necessitate the use of high-precision beam shaping devices due to the reliance on administered dosage modulation. Random errors should be eliminated by paying close attention to the accuracy and performance of the MLC. Systematic errors must be identified and reduced [4][5].

Using Supervised Learning Algorithms to predict the problem of leaf displacement accuracy in a multi-leaf Collimator mounted in a Linear Accelerator Head will result in accurate positioning based on the shape of the tumor being treated while protecting other nearby body organs, thereby contributing to accurate radiation dose delivery to oncology Patients. Fig. 2 shows a photo of a Multileaf Collimator [12].

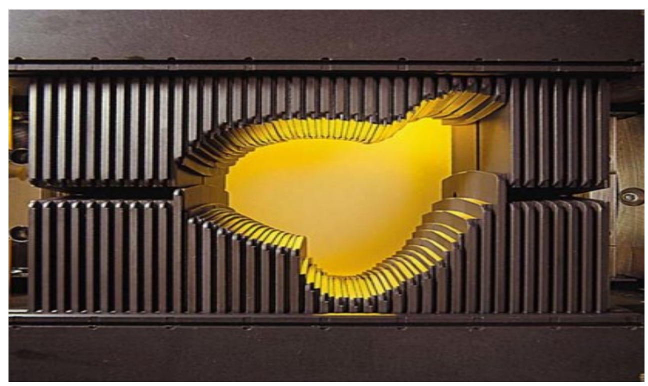


Fig. 2. Photo of Multi-leaf Collimator MLCi2 While A-Bank Leaves (Upper Leaves) and B-Bank Leaves (Lower Leaves) are positioned to take the Precise Shape of the Tumor so that the Treatment Beam can get through that Shape only [12].

IV. Methodology

As shown in Fig. 3, the application-related Data (i.e. Linac MLC's Leaves' displacement Dataset) will be used for Training while developing the Learning model provided by the supervised Machine Learning algorithm, as well as for testing to evaluate the developed learning model. As a result, we could finally compute accuracy, draw the receiver operating characteristic curve (ROC), and evaluate the various algorithms.

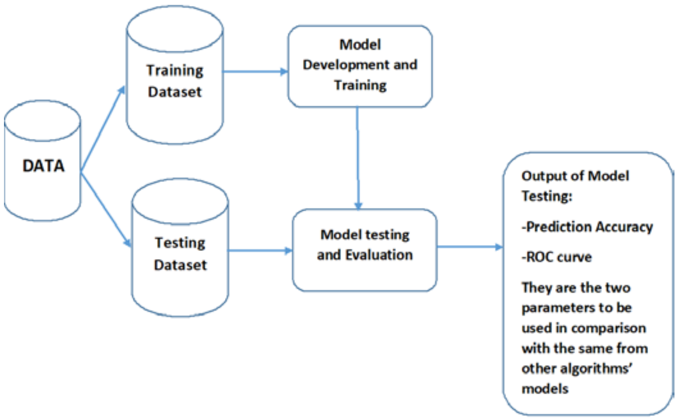


Fig. 3. Steps followed to Assess the Performance of SVM, Decision Tree, RF, Logistic Regression, NB and KNN Algorithms.

A. Confusion Matrix

The confusion matrix shown in Table I is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known [8].

TABLE I. CONFUSION MATRIX FOR TWO CLASSES

	PREDICTED CLASS	
	POSITIVE	*NEGATIVE*
ACTUAL CLASS	TRUE POSITIVE (*tp*)	FALSE NEGATIVE (*fn*)
	FALSE POSITIVE (*fp*)	TRUE NEGATIVE (*tn*)

The above confusion matrix taught us that there are two possible predicted classes: "True" and "False" [8]. In this paper, for example, we predict the absence of a leaf positioning accuracy problem for a typical leaf out of 40 pair leaves in Linac MLC, as a result, we can extract and define the following terms:

True positives (tp): These are cases where the prediction was correct and the leaves did not have positioning accuracy problem.

True negatives (tn): These are cases where the prediction was correct and the leaves have positioning accuracy problem.

False positives (fp): These are cases where the prediction was incorrect and the leaves actually did not have positioning accuracy problem.

False negatives (fn) are cases where the prediction was incorrect and the leaves actually have positioning accuracy problem. Table I can be used to extract the following rules [3]:

$$\text{Accuracy} = (tn + tp)/(tn + fn + fp + tp\ ) \quad (1)$$

$$\text{Error} = (fn + fp)/(tn + fn + fp + tp) \quad (2)$$

$$\text{Sensitivity} = \text{recall} = tp/(tp + fn) \quad (3)$$

$$\text{Specifity} = tn/(fp + tn) \quad (4)$$

B. Receiver Operating Characteristics (ROC)

The receiver operating characteristics (ROC) curve and confusion matrix are frequently used to evaluate the diagnostic ability of supervised machine learning algorithms [7]. The Y axis of receiver operating characteristic curves has the tp rate and the X axis has the fp rate. That is, the "ideal" point is at the top left corner of the plot, where fp rate is zero and tp rate is one. However, this isn't very realistic, but it does imply that a larger area under curve (auc) is usually preferable. Fig. 4 depicts an example of evaluation for various ROC curves. Whereas the blue curve has the lowest auc, indicating poor prediction performance, the red curve has the highest auc, indicating excellent prediction performance [13].

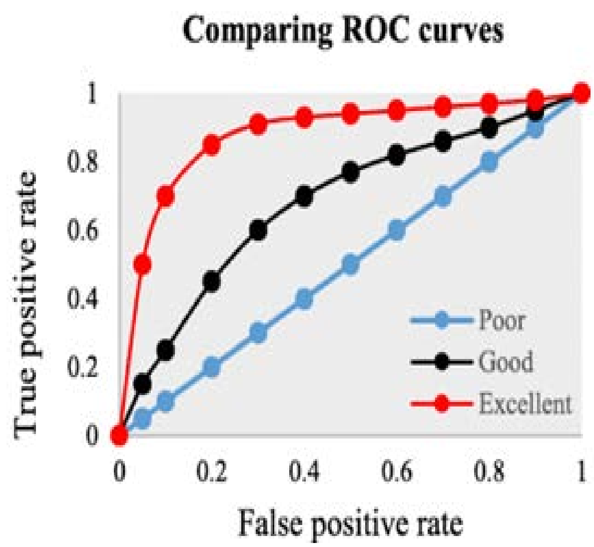


Fig. 4. Illustrates Example Evaluation for different ROC Curves; where Blue Curve is Poor because Area under Vurve is the Lowest and Red Curve is Excellent because Area under Curve is Highest.

C. Data Collection and Processing

Good test and training Dataset was collected over a working year (252 days) for an MLCi2 Multi-leaf Collimator (MLC) mounted in an Elekta Synergy Linear Accelerator. The MLC has 40 leaf-pairs Linac (80-leaves). The 40 leaf-pairs are numbered into two banks "A&B" as: (A1, A2,…..,A40 and B1,….,B40). The tolerance of the leaf positioning accuracy is 2 mm, while the action level is 3 mm. Table II shows a sample of collected data with labeled input features and its associated labeled result (i.e. answer to question is there isn't positioning accuracy problem?).

As shown in Fig. 5, 70% of MLC leaves displacement dataset was used for training the supervised machine learning models (e.g. x_train represents training features and y_train is the labelled result of training dataset) whereas 30% of the dataset was used to test and evaluate the trained models (e.g. x_test represents test features and y_test is the labelled result of test dataset).

Data is then processed using Python 3.8 using PyCharm IDE. Python package (i.e. Scikit-learn 0.23.2) was used to implement DT, SVM, RF, LR, NB and KNN Classifiers.

TABLE II. SAMPLE CASES FOR DISPLACEMENT DATA CASES FOR 80 LEAVES MULTI-LEAF COLLIMATOR (TRUE:T /FALSE:F)

LINEAR ACCELERATOR MULTI-LEAF COLLIMATOR MLC LEAVES' DISPLACEMENT (A1,A2,…A40,B1,B2,….B40)										**NO POSITIONING ACCURACY PROBLEM?**	
A1	*A2*	*A3*	*A4*	*A5*	…	*B37*	*B38*	*B39*	*B40*	*TRUE T/FALSE F*	
										A BANK LEAVES	*B BANK LEAVES*
0.24	0.99	0.42	0.38	1.17	…	0.24	1.28	0.62	1.04	T	T
										A	B
0.38	1.1	0.46	0.34	0.94	…	0.55	0.36	1.1	0.61	T	T
										A	B
0.42	0.84	0.84	0.92	0.24	…	1.09	0.96	0.75	0.77	T	T
										A	B
0.55	0.74	0.38	1.15	0.95	…	0.75	0.59	0.46	1.06	T	T
										A	B
1.13	2.12	0.98	0.96	1.29	…	1.32	1.35	1.89	0.74	F	F

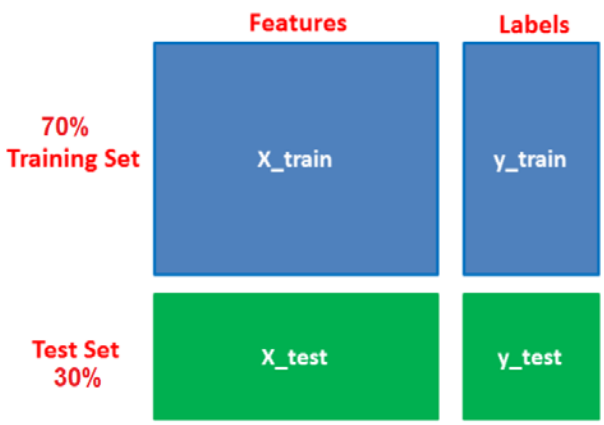


Fig. 5. Illustration of how Linear Accelerator Multileaf Collimator MLC Leaves' Displacement Dataset is Prepared for use by Machine Learning Algorithms.

V. IMPLEMENTATION

A. *Prediction for MLC's Positioning Accuracy*

Table III shows Measured Accuracy achieved by different Classifiers (i.e. DT, SVM, RF, LR, NB and KNN classifiers) in prediction of 40-pairs MLC's positioning accuracy problem.

TABLE III. EVALUATION OF SUPERVISED MACHINE LEARNING ALGORITHMS IN PREDICTION OF LINEAR ACCELERATOR MULTI-LEAF COLLIMATOR POSITIONING ACCURACY PROBLEM

		A-Bank Leaves	B-Bank Leaves
Decision Tree	Accuracy	94.74 %	93.42 %
Support Vector Machine	Accuracy	99.98%	97.37 %
Random Forest	Accuracy	97.36 %	98.68 %
Logistic Regression	Accuracy	98.68 %	98.68 %
Naïve Bayes	Accuracy	97.37 %	94.74 %
K-Nearest Neighbor	Accuracy	96.05 %	98.68 %

It shows that DT, SVM, RF, LR, NB and KNN classifiers were able to predict leaf positioning accuracy problem for Linac MLC successfully with average accuracy of 94.08%, 98.68%, 98.02%, 98.68%, 96.05%, and 97.37%, respectively. While Fig. 6 shows different ROC curves, one per each classifiers based on their prediction performance, so that we can compare between algorithms based on steepness of the curve and area under curve (auc). Where it's clear that red curve representing logistic regression algorithm has the highest ROC area under curve of 0.992, while green curve representing Decision Tree classifier has the lowest ROC area under curve.

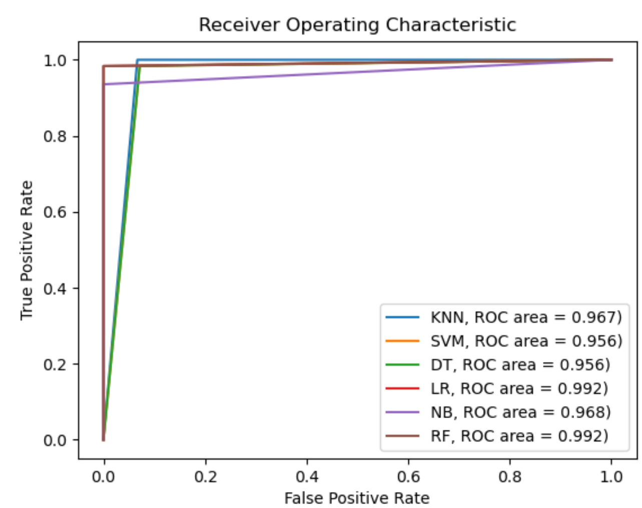


Fig. 6. Receiver Operating Characteristic Curve for Prediction of Linear Accelerator Multi-leaf Collimator's Positioning Accuracy per each Algorithm.

VI. RESULT AND DISCUSSION

DT, SVM, RF, LR, NB and K-Nearest Neighbor classifiers were able to predict leaf positioning accuracy problem for Linac MLC successfully with average accuracy of 94.08%, 98.68%, 98.02%, 98.68%, 96.05%, and 97.37%, respectively.

Although DT has reasonable average accuracy of 94.08% but it's the lowest one amongst other algorithms. And the ROC curve for DT performance has the lowest steepness and

ROC area under curve as well. So it's not recommended to select DT as first choice to use on such MLC's leaves' displacement dataset nature. On the other side, SVM performed well with average accuracy of 98.68% but it showed lower steepness in ROC Curve and area under curve.

Other classifiers RF, LR, NB and KNN showed better area under curve, and their average prediction accuracies were RF with 98.02%, LR with 98.68%, NB with 96.02% and finally KNN with 97.37%.

However, it's important to note that, Logistic Regression Classifier has the highest ROC area under curve of 0.992 and it showed exceptional performance by having the same classification prediction accuracy of 98.68% over two different datasets of same structure and nature but different values (i.e. MLC's A Bank Leaves' displacements and MLC's B Bank Leaves' displacement) which indicates more performance stability than other classifiers even SVM itself. In this paper, bias of algorithm selection have been successfully avoided and the recommended selection should consider the relevancy order based on classifiers evaluation as shown in Table IV, in context of application area related to MLC's Leaves' displacement Data. And accordingly we do recommend using LR as first choice because it has highest average prediction accuracy and the most stable performance across different MLC's datasets of the same nature.

TABLE IV. PROPOSED SELECTION ORDER FOR CLASSIFIERS TO USE IN PREDICTION OF MLC POSITIONING ACCURACY PROBLEM

Classifiers	**Average Accuracy**	**Selection Order**
Logistic Regression	98.68%	1
Support Vector Machine	98.68%	2
Random Forest	98.02%	3
K-Nearest Neighbor	97.37%	4
Naïve Bayes	96.05 %	5
Decision Tree	94.08%	6

On the other hand and according to application area perspective, the high accuracy of the prediction for MLC's Leaves' positioning problem would enable the physicist in oncology center to design customized service/preventive maintenance plans for each individual Linac MLC treatment machine particularly. And that could help to avoid MLC movement failure during radiation therapy sessions.

VII. CONCLUSION

This work is undertaken to avoid lower performance in prediction process of Linac MLC's positioning accuracy problem. In this paper, performance of DT, SVM, RF, LR, NB and KNN Classifiers is examined by measuring their prediction accuracies utilizing the same two sets of training and testing data for Linac MLC's leaves' positioning displacement data as well as receiver operating characteristic curves for the predicted outcomes per each algorithm. Findings in this study show that Logistic regression Classifier has exceptional performance by producing the same classification prediction accuracy of 98.68% over two different datasets of same structure and nature but different values (i.e. MLC's A Bank Leaves' displacements and MLC's B Bank Leaves' displacement) which indicates more performance stability than other classifiers even SVM itself. The findings show that values and structure of data affect the prediction accuracy of supervised learning algorithm applications across different industries and not necessarily the same performance. In order to increase the prediction accuracy in the same time, further research work is needed on more training and test datasets over longer periods (e.g. five years), and a multi-institutional study (e.g. different healthcare providers which uses the same model of Linac MLC).

ACKNOWLEDGMENT

The authors would like to thank the reviewers and editors for their constructive comments.

REFERENCES

[1] Suk Lee, Kwang Hyeon Kim, Choi Suk Woo, Jang Bo Shim, Yuan Jie Cao, Kyung Hwan Chang and Chul Yong Kim "Predictive Solution for Radiation Toxicity Based on Big Data", May 2017,ch.9,p175-191 Radiotherapy.

[2] El Naqa I, Bradley JD, PE L, Hope AJ, Deasy JO. Predicting radiotherapy outcomes using statistical learning techniques. Phys Med Biol. 2009;54(18):S9.

[3] Dean JA, Wong KH, Welsh LC, Jones AB, Schick U, Newbold KL, Bhide SA, Harrington KJ, Nutting CM, Gulliford SL. Normal tissue complication probability (NTCP) modelling using spatial dose metrics and machine learning methods for severe acute oral mucositis resulting from head and neck radiotherapy. Radiother Oncol. 2016 Jul;120(1):21–7.

[4] Joel N K Carlson, Jong Min Park, So-Yeon Park, Jong In Park, Yunseok Choi and Sung-Joon Ye,"A machine learning approach to the accurate prediction of multi-leaf collimator positional errors",Journal of Physics in Medicine & Biology,March 2016, Vol 61, Number 6,2514.

[5] El Naqa I, Li R, Murphy MJ. Machine learning in radiation oncology: theory and applications. Switzerland, Springer; 2015. 39.

[6] The R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, Machine learning: an artificial intelligence approach. Springer Science & Business Media, 2013.

[7] Fawcett T. An introduction to ROC analysis. Pattern Recogn Lett. 2006;27(8):861–74.

[8] Amr E. Mohamed, "Comparative Study of Four Supervised Machine Learning Techniques for Classification," International Journal of Applied Science and Technology Vol. 7, No. 2, June 2017, pp 5-17.

[9] Breiman L. Random forests. Mach Learn. 2001;45(1):5–32.

[10] Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied logistic regression. Wiley; 2013.

[11] I. Rish, "An empirical study of the naive Bayes classifier," in IJCAI 2001 workshop on empirical methods in artificial intelligence, 2001, vol. 3, 22, pp. 41–46: IBM New York.

[12] Elekta user manuals for MLCi2, Elekta© Crawley, UK.

[13] Jasmina Dj. Novakovic´,Alempije Veljovic´, Sinisa S. Ilic´,Zeljko Papic´,Milica Tomovic´. "Evaluation of Classification Models in Machine Learning" Theory and applications of Mathematics and Computer Science Vol 7 No.1, 2017,pp 39-46.

[14] Oded Maimon • Lior Rokach, Data Mining and Knowledge Discovery Handbook, NY: Springer. 2010, pp. 149-152.

[15] Jiawei Han and Micheline Kamber. Data Mining: Concepts and Techniques, NY: Elsevier. 2010, pp. 291-307.

Collaborative Course Assignment Problem to Minimize Unserved Classes and Optimize Education Quality

Purba Daru Kusuma, Ratna Astuti Nugrahaeni
Computer Engineering, Telkom University
Bandung, Indonesia

***Abstract*—This work proposes a collaborative course assignment model among universities. It is different from existing studies in educational assignment problems or course timetabling, where the scope is only within the institution or department. In this work, the system consists of several universities. A collaborative approach is conducted so that lecturers exchange is possible and conducted automatically. Each university shares its courses and lecturers. The optimization is conducted to minimize the unserved classes and improve education quality. The cloud-theory based simulated annealing is deployed to optimize the assignment. This model is then benchmarked with two non-collaborative models. The first model's objective is to minimize the unserved classes only. The second model's objectives are to minimize the unserved classes and improve education quality. The simulation result shows that the proposed assignment model is better in minimizing the unserved classes and improving education quality. The proposed model reduces 89 to 92 percent of the unserved classes ratio compared with the non-collaborative model.**

***Keywords*—Course assignment problem; simulated annealing; collaborative model; online teaching; combinatorial problem**

I. Introduction

Course assignment problem is a well-known study in operations research or optimization, especially in education. This popularity comes from the fact that course assignment is an important subject in educational operation. Moreover, circumstances in the educational operation are complex and diverse. First, the regulations in education among regionals or countries are different [1]. Furthermore, some institutions, especially universities, may have specific policies, needs, and objectives [1]. Although there is a generic form of course assignment model, many studies in this subject have specific circumstances that are transformed into objectives and constraints. The objectives include balancing workload [2], minimizing classes without lecturers [3], improving education quality [4], and so on.

Ironically, the scope of most existing studies in education timetabling and assignment problems is still within the university [5] or department [6]. On the other hand, lecturers are limited resources, like the timeslot and room. It means that these offered classes may not be conducted because no lecturers can handle these classes. The other condition is that these classes are still conducted, but the education quality may be dropped because less competent lecturers serve these classes.

On the other hand, the online class is common today. Due to the COVID-19 pandemic, schools, especially universities, are forced to swift from face-to-face to online interaction [7]. Although the online mechanism still has several weaknesses, such as the degradation of the learning outcomes and lack of interaction [7], the online class has several advantages. The location boundaries are not considered anymore if students and lecturers have reliable internet access [8], especially in rural or remote areas [9]. Moreover, physical rooms are not needed too.

This online class creates opportunities for inter-universities collaboration. Lecturer exchange becomes possible and easy. Some lecturers may teach courses that other universities offer. Through this collaboration, a university can offer courses to its students, although it does not have competent lecturers. A university also can open more classes even though its own lecturers are limited. On the other hand, a university also can provide its lecturers to handle courses or classes beyond their formal homebase. Moreover, assigning a course to a more competent lecturer can improve the education quality. Unfortunately, most studies in education timetabling or assignments were conducted based on physical classroom scenarios. These existing assignment models cannot be implemented directly to tackle this collaborative and online circumstance. Studies in operations research in the education area that promote collaboration and online learning are challenging and potential.

Based on this circumstance, this work proposes a collaborative course assignment model. In this work, the system consists of several universities. Each university shares its several classes and lecturers that can be assigned collaboratively. Each class has a specific course and timeslot. On the other hand, each lecturer can provide several courses and specific timeslots. The objective is to minimize the unserved classes and maximize the education quality.

This assignment model is then optimized by cloud-theory based simulated annealing (CSA). This algorithm is an improved version of simulated annealing. Rather than its original version, CSA is a population-based metaheuristic algorithm where every individual acts independently [10]. In the end, the best individual is chosen as the final solution.

This algorithm is chosen based on several reasons. First, the metaheuristic algorithm is popular in optimizing many operations research studies, especially assignment problems. This popularity comes from its approximate approach so that excessive computation can be avoided, although it may promise near-optimal or acceptable solutions [11]. Second, simulated annealing is a simple algorithm that can easily be implemented, improved, or modified to solve many optimization problems. Third, as a population-based algorithm, CSA is proven better than its original form in providing better solutions [10].

The contributions of this work are as follows:

1) This work proposes a course assignment model conducted collaboratively. The system consists of several universities rather than a single university or department, as it is common in most studies in education assignment or timetabling problems.

2) The proposed model is developed for online classes so that physical rooms are not needed. It is also different from most studies in education assignments or timetabling problems where their circumstance is the physical classroom.

This work is the continuation of our previous works in operations research in the education area. Both previous studies were conducted for joint course programs. The first study focused on the course timetabling [12], while the second focused on the faculty assignment [13]. The difference between this current work and the previous works is that the university's interest is considered in this work. In these previous studies, the existence of the universities as entities that provide lecturers and classes is not considered.

The remainder of this paper is organized as follows. The shortcoming studies in education assignment problems and course timetabling are explored in the second section. The proposed model that consists of both conceptual model, mathematical model, and the algorithm is explained in the third section. The simulation scenario and result are described in the fourth section. The more profound analysis conducted on the simulation result and findings are discussed in the fifth section. Finally, the conclusion and future research potentials are summarized in the sixth section.

II. Related Work

In general, the assignment problem can be defined as allocating or assigning a certain number of objects to a certain number of other objects in the most optimal way [14]. The assignment problem consists of two components. The first component is the assignment [14], and the second component is the objective function [14]. An assignment is a combinatorial structure that consists of the link between a set of objects and another set of objects. The relationship between these sets of objects can be one-to-one, one-to-many, or many-to-many. On the other hand, the objective is the purpose of the assignment. The objective represents a valid measurement to evaluate the assignment's performance. Based on this concept, assignment problem becomes a part of operations research, and it is widely used in many areas, such as transportation [15], manufacturing [16], logistics [17], and so on.

In the education area, there are two well-known assignment problems. The first is the course timetabling problem [1]. The second is the faculty assignment problem [1]. In the course timetabling problem, a course will be allocated to certain timeslots and rooms. In the faculty assignment problem, the focus is plotting lecturers to the courses in the system. Both problems are at the operational level [1].

To date, there are huge numbers of studies conducting the assignment problem in the education area. This circumstance shows that operation research studies in education are still interesting. Besides, assignment problem in education is widely studied since there are various circumstances in the education institutions. This variety comes from several aspects, such as regulation, institutional objective, and local challenge [2]. Moreover, these studies are usually proposed based on certain specific objectives.

Many studies in the education assignment problem used metaheuristic algorithms as the optimization method. The use of metaheuristic algorithms comes from several reasons. First, the metaheuristic algorithm is a proven method used in many optimization studies. Second, this algorithm is proven to achieve a near-optimal solution with reasonable computation resources [11]. Third, the metaheuristic algorithm is a popular algorithm that has been studied extensively until now. To date, there are hundreds of metaheuristic algorithms that have been developed. Several well-known algorithms are also used in many education assignment problems, such as genetic algorithm [18], simulated annealing [19], tabu search [20], variable neighborhood search [6], genetic programming [21], and so on. Several shortcoming studies in the education assignment problems are summarized in Table I. In the last row, the positioning of this work is stated.

TABLE I. Shortcoming Studies in Education Assignment Problem

Author	Scope	Physical Room	Optimization Method
[22]	department	needed	tabu search
[16]	department	needed	genetic algorithm, local search
[19]	department	needed	simulated annealing
[5]	department	needed	multi-agent, genetic algorithm
[20]	department	needed	variable neighborhood search, tabu search
[23]	university	needed	genetic algorithm
[6]	department	needed	tabu search, variable neighborhood search
[24]	university	needed	genetic algorithm
[25]	department	needed	tabu search, simulated annealing
[26]	department	needed	Monte Carlo search
[27]	department	needed	tabu search, iterated local search, simulated annealing
[21]	university	needed	genetic programming, genetic algorithm
this work	multiple universities	no needed	cloud-theory based simulated annealing

Table I shows that all shortcoming operation research studies in education still adopt a conventional approach. The scope of these studies is within a university or department. Several studies conducted certain classes, which is relatively small. Moreover, all these studies used physical rooms in the system.

This circumstance makes the existing assignment models cannot be implemented directly in the future education environment. The online collaborative system faces many different circumstances than the face-to-face noncollaborative system. In the future, online learning will become more popular. Moreover, the emergence of online learning makes collaboration among universities more possible. Based on this problem, proposing an assignment model that eliminates the physical boundaries and promotes a collaborative approach as in this work becomes very important and interesting.

The future online learning also promises more efficient teaching system. The university can provide more classes without creating more physical rooms or building so that it can save more capital expenditure for the development and operational expenditure related to the new buildings or rooms, such as electricity, water, cleaning service, furniture, maintenance, and so on. This cost reduction in the end can reduce the educational cost so that the institution will be more competitive. This cost reduction can also be used to tackle the cost increase in other posts, such as employee salary, internet, and so on so that the increase in the tuition fee can be avoided. The online learning also improves the teacher's movement time. In the face-to-face learning, the teacher must move from the current room to other room or from the current building to other building to teach other classes. Moreover, when the university has several separated locations, this movement wastes more time.

III. Proposed Model

This collaborative course assignment model consists of several entities: university, lecturer, and class. In the system, there are several universities. Each university has several lecturers and classes that will be shared in the system. It means a university still has a portion of its classes and lecturers that are not included in the system. These classes and lecturers will be managed exclusively by its homebase. Every lecturer can teach several courses, but their competence may be various among the courses. For example, a lecturer can teach database, algorithm, and object-oriented programming courses. His competence in teaching database is average but prominent in teaching algorithm and object-oriented programming. All lecturers have their available timeslots. A class is dedicated to a specific course and timeslot. In this model, students are already assigned to this class. Students are also abstracted and are assumed to take only one class per student. Fig. 1 illustrates the conceptual system. The blue dashed rectangles represent the universities, the blue circles represent the lecturers, and the yellow circles represent the classes.

The relationship between university, class, lecturer, course, and timeslot is as follows. The relationship between university and class is one-to-many. The relationship between university and lecturer is one-to-many. The relationship between lecturer and course is many-to-many. The relationship between lecturer and timeslot is many-to-many. The relationship between class and lecturer is one-to-many. The relation between classes, courses, and lecturers are shown in Fig. 2. In Fig. 2, red circles represent the lecturers, blue circles represent the courses, and green circles represent the classes.

This proposed model has several hard constraints that cannot be violated [1]. These hard constraints are as follows.

- The number of universities, lecturers, and classes is predetermined [13].
- A lecturer cannot teach courses beyond his competency [13].
- A lecturer cannot teach multiple classes with the same timeslot [12].
- A lecturer cannot teach beyond his possible timeslots [12].
- Timeslot for every class is predetermined [13].
- A class cannot be conducted beyond the provided timeslots [12].
- A class cannot be taught by multiple lecturers [13].

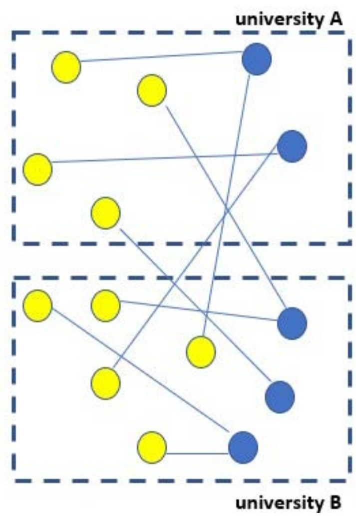


Fig. 1. Conceptual System.

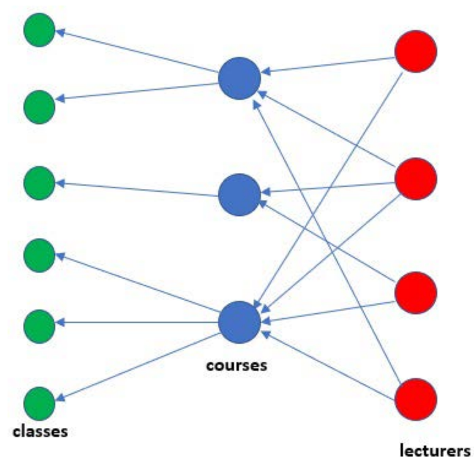


Fig. 2. Relation between Lecturers, Courses and Classes.

This proposed model has two objectives. The first objective is to minimize unserved classes. It becomes the primary objective. The second objective is to maximize the education quality, meaning that classes will be taught by the most competent lecturer wherever possible. The internal lecturer is prioritized to accommodate the university's interest.

There are two types of assignments. The first assignment is the intra-university assignment. The second assignment is the inter-university assignment. In the intra-university assignment, a class will be allocated to the possible internal lecturer exclusively. In the inter-university assignment, a class will be allocated to any possible lecturer without considering the lecturer's homebase. The collaboration is conducted in the inter-university assignment.

This process is then optimized by using cloud-theory based simulated annealing. As a metaheuristic algorithm, it consists of two phases. The first phase is initialization. The second phase is iteration. Both intra-university assignments and inter-university assignments are conducted in the initialization. In the intra-university assignment, several courses may remain unallocated to a lecturer or unserved due to the mismatch problem. These unserved classes will be assigned in the inter-university assignment. Meanwhile, the iteration phase consists of only the inter-university assignment. As in all simulated annealing algorithms, the iteration consists of external and internal loops. In the external loop, iteration runs from the initial high temperature to the end low temperature with a certain decrease rate [10]. In the internal loop, iteration runs from the first iteration to the maximum iteration [10].

Neighborhood search is conducted to improve the solution. The solution candidate is generated near the current solution. If this candidate is better than the current solution, then this candidate replaces the current solution immediately. Otherwise, this candidate may replace the current solution with a certain probabilistic calculation to avoid local optimal trap.

The mathematical model is then developed based on this conceptual model. Several annotations used in this mathematical model are as follows. Meanwhile, the process in the proposed model is shown in Algorithm 1.

u	university
l	lecturer
L	set of lecturers
l_a	available lecturer
L_a	set of available lecturers
t	timeslot
T	set of timeslots
c	course
C	set of courses
s	class
S	set of classes
S_s	set of served classes
S_u	set of unserved classes
st	status
st_a	timeslot availability status
st_c	same course status
st_i	inter-university status
n	number of entities
f	fitness
o	objective
U	uniform random
i	iteration
i_{init}	initial iteration
i_{max}	maximum iteration
e	temperature
e_{init}	initial temperature
e_{end}	end temperature
Δe	temperature decrease rate
k	Boltzman constant
a	individual / solution
a_{best}	best individual
A	population
d	candidate

algorithm 1: collaborative course assignment model

output: a_{best}
//initialization
for $x = 1$ **to** $n(A)$ **do**
begin
  a_x = intra-university-assignment (L, S)
  a_x = inter-university-assignment (L, S)
end
//iteration
$e = e_{init}$
while $e > e_{end}$ **do**
begin
  for $i = i_{init}$ **to** i_{max} **do**
  begin
    for $x = 1$ **to** $n(A)$ **do**
    begin
      d = neighborhood-search (a_x)
      if $f_1(d) < f_1(a_x)$ **then**
        if $f_2(d) > f_2(a_x)$ **then**
          $a_x = d$
        else
          if $U(0, 1) < \exp((f_2(d) > f_2(a_x))/e)$ **then**
            $a_x = d$
    end
  end
  $e = e - \Delta e$
end
a_{best} = min-sort ($f_1(A)$)

The explanation of Algorithm 1 is as follows. The algorithm's output is to find the best individual or solution that consists of the best assignment in meeting the primary and secondary objectives. The initialization consists of intra-university assignments and inter-university assignments that are conducted serially. The outer loop is the loop that runs from the initial high temperature to the end low temperature. The temperature decreases gradually based on the temperature decrease rate. Then, the inner loop consists of iteration from the first iteration to the maximum iteration. In the iteration process, a neighborhood search based on the inter-university assignment is conducted. This neighborhood search is conducted to produce a candidate. Then this candidate is evaluated by two fitness functions that represent the objectives. The first function is minimization, while the second function is maximization. The candidate will replace the current solution immediately only if its performance is better than the current solution in both fitness functions. If the first candidate's first fitness is worse than the current solution, it is rejected immediately. Suppose the candidate is better than the current solution only in the first fitness function. In that case, it may replace the current solution based on a probabilistic calculation where the fitness gap and the current temperature are considered. After all iterations end, the best solution is selected based on the primary objective.

As it is mentioned previously, the proposed assignment model has two objectives. The first objective is minimizing the unserved classes. The second objective is maximizing the education quality. The first objective is the primary objective, while the second objective is the secondary objective. These objectives are formalized by using (1) to (6).

$$o_1 = \min\left(f_1(a)\right) \quad (1)$$

$$f_1(a) = \frac{n(S_u)}{n(S)} \quad (2)$$

$$S_u = \{s \in S | st(s) = 0\} \quad (3)$$

$$o_2 = \max\left(f_2(a)\right) \quad (4)$$

$$f_2(a) = \frac{\sum_{n(S_s)} p(l,c)}{n(S_s)}, s \in S(l) \quad (5)$$

$$S_s = \{s \in S | st(s) = 1\} \quad (6)$$

The explanation of (1) to (6) is as follows. Equation (1) states that the first objective is to find a solution with minimum unserved classes. Equation (2) states that the first fitness function is obtained by dividing the number of unserved classes by the total number of classes. Equation (3) states that the unserved classes are classes that do not have lecturers. Equation (4) states that the second objective maximizes education quality. Equation (5) states that the education quality is obtained by dividing the summation of lecturers' competence related to the course and class with the number of served classes, and the class is taught by the lecturer. Equation (6) states that the served classes are classes that have lecturers.

The initialization process begins with the intra-university assignment process. Its mechanism is allocating every course to be taught by internal lecturers. This mechanism is conducted by collecting all internal lecturers who are available and competent to teach the selected course. Available means that the lecturer still has an available timeslot that is the same as the class timeslot. Competent means that the course taught in the class is on the lecturer's competence list. This mechanism is formalized by using (7) to (10).

$$st_i(s,l) = \begin{cases} 1, u(s) = u(l) \\ 0, else \end{cases} \quad (7)$$

$$st_a(s,l) = \begin{cases} 1, t(s) \in T(l) \land st\left(t(l)\right) = 0 \\ 0, else \end{cases} \quad (8)$$

$$st_c(s,l) = \begin{cases} 1, c(s) \in C(l) \\ 0, else \end{cases} \quad (9)$$

$$L_a(s) = \{l | st_i(s,l) = 1 \land st_a(s,l) = 1 \land st_c(s,l) = 1\} \quad (10)$$

$$l_{se}(s) = U\left(L_a(s)\right) \quad (11)$$

$$st\left(t(l)\right) = \begin{cases} 1, \exists s, t(s) = t(l) \land l(s) = l \\ 0, else \end{cases} \quad (12)$$

The explanation of (7) to (12) is as follows. Equation (7) states the internal status is 1 only if the class and lecturer are in the same university. Equation (8) states that the availability status is 1 only if the class timeslot is within the lecturer's timeslot and the related lecturer's related timeslot is still available (open). Equation (9) states that the competence status is 1 only if the course in the class is within the lecturer's course list. Equation (10) indicates the set of available lecturers for the class. A lecturer is available if it meets all three statuses. Equation (11) states that the lecturer is selected randomly within the set of available lecturers. Finally, the timeslot status of the lecturer is set 1 if there exists a class in which the timeslot is the same as the lecturer's timeslot, and it is taught by the lecturer as indicated in (12).

The second step is the inter-university assignment process. This step is conducted only for classes that have not been assigned yet after the first step ends. In this step, the university status is not considered anymore. It means that a class can be taught by any available lecturer in the system. This process is formalized by using (13).

$$L_a(s) = \{l | st_a(s,l) = 1 \land st_c(s,l) = 1\} \quad (13)$$

Equation (13) shows that only two parameters determine the availability of a lecturer. The first parameter is the availability status, which is determined by using (8). The second parameter is the competence status determined by using (9). Finally, the lecturer is selected by using (12), where the set of available lecturers is determined by using (12).

There are several notes due to the initialization phase. First, this phase does not guarantee that there are no unserved classes. Second, the education quality determined by using (5) has not been optimized.

These notes become the reason to conduct the optimization process through iteration by using cloud-theory-based simulated annealing. In this algorithm, neighborhood search is conducted to improve the current solution. This search follows the inter-university assignment. This search is conducted by selecting several classes randomly. If there exists classes within these selected classes assigned to certain lecturers, then the lecturer-class link will be reset. Finally, the inter-university assignment is conducted for all unserved classes.

IV. Simulation and Result

The proposed model is then implemented into a simulation to evaluate its performance. In this simulation, a certain number of universities are created. Then, a certain number of lecturers and classes attached to the universities are also created. Every class is conducted for a specific course within a specific timeslot. After these three entities are created, then the simulation runs to allocate these classes to a certain lecturer. The classes' course, lecturers' timeslot, lecturers' course, classes' university, and lecturers' university are generated randomly and follow a uniform distribution.

In this simulation, there are adjusted parameters and observed parameters. The observed parameters are the unserved classes ratio and the education quality ratio. The unserved classes ratio is a ratio between the number of unserved classes and the total classes. The education quality ratio is the average lecturer's competence score among the served classes. Meanwhile, the default value of the adjusted parameters is shown in Table II.

TABLE II. PARAMETERS' DEFAULT VALUE

Parameter	Default Value
$n(u)$	5
$n(C)$	20
$n(C(l))$	4
$n(S)$	200
$n(L)$	$n(S)$ / 2
$n(T)$	15
n($T(l)$)	7

There are three simulations. The first simulation is conducted to observe the relation between the number of classes and the observed parameters. The second simulation is conducted to observe the relation between the number of lecturer's timeslot and the observed parameters. The third simulation is conducted to observe the relation between the number of courses conducted by a lecturer and the observed parameters.

The reason of choosing these parameters is as follows. The number of classes, lecturer's timeslot, and lecturer courses are resources that are easily to improve. In the online education environment, university can add more classes without creating new rooms or building which are necessary in the face-to-face teaching environment. The number of lecturer's timeslot is also easy to manage. Even though the total timeslot in a week is fixed, the number of lecturer's timeslot can be improved in several ways if it does not surpass the total timeslot. First, timeslots for the collaborative teaching can be increasing by reducing timeslots assigned for the non-collaborative teaching. Second, several lecturer's non-teaching activities can be shifted outside the teaching timeslots window. The number of lecturers courses is also can be improved easily if the teacher has enough preparation for the new course assignment, especially the courses that are near the current assignment. For example, lecturer that teaches algorithm course can also be assigned to teach other programming related courses, such as object-oriented programming. On the other hand, lecturer that teaches artificial intelligence can also be assigned to teach machine learning and deep learning courses.

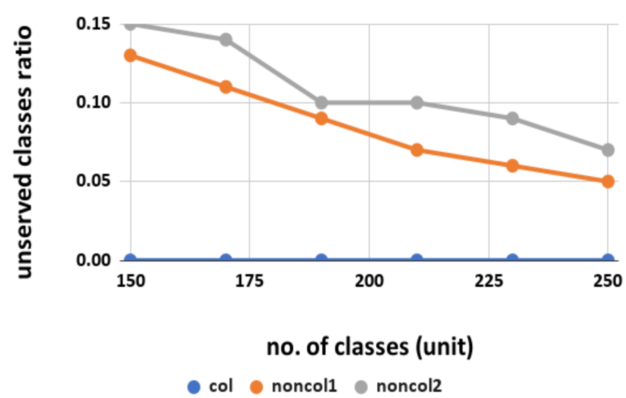
This proposed model is benchmarked with two non-collaborative assignment models, in which both models are concerned with minimizing the unserved classes. There is a difference between the first model and the second one. The first model's objective is only to minimize the unserved classes. The second model is not only concerned with minimizing the unserved classes but also with maximizing the education quality. The first model is adopted based on the model proposed by Arratia-Martinez et al [3].

Meanwhile, the second model is adopted based on a model proposed by Wicaksono and Wisesa [28], where education quality is prioritized. But the second model is improvised so that the unserved classes are considered too. Moreover, the circumstance is also modified to be comparable to the proposed model. In this simulation, both models use cloud-theory-based simulated annealing too. They are fairly compared because the purpose of this simulation is not to compare the metaheuristic algorithms but to compare the collaborative approach with non-collaborative ones.

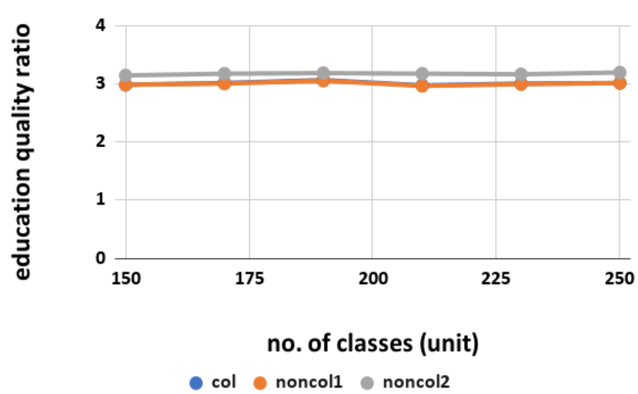
The first simulation is conducted to observe the relation between the number of classes and the observed parameters. The number of classes ranges from 150 to 250 with 20 step size. Other adjusted parameters are set to default. The result is shown in Fig. 3.

Fig. 3a shows that in general, the increase of the number of classes makes the decrease of the unserved classes ratio. It occurs in both non-collaborative models. On the other hand, the proposed model produces a zero unserved classes ratio due to this scenario. It occurs when the number of classes ranges from 150 to 250 units. It means that the proposed model outperforms both non-collaborative models in this simulation. Comparing both non-collaborative models, the first model is better than the second one in creating a lower unserved class ratio.

Fig. 3b shows that the number of classes does not affect the education quality ratio. The education quality ratio tends to be stagnant in all number of classes. This circumstance occurs in all models. Comparing among models, all models are competitive. Meanwhile, the second non-collaborative model is the best one with a very narrow gap. The performance of the proposed model and the first non-collaborative model is almost equal.



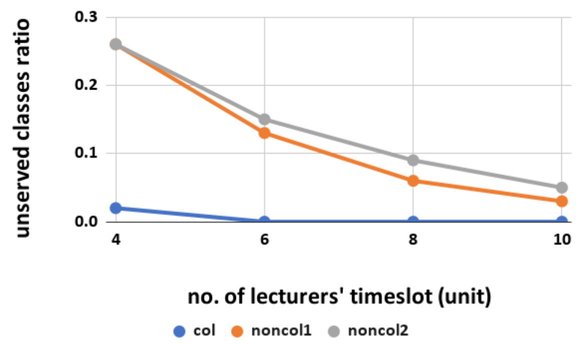
(a)



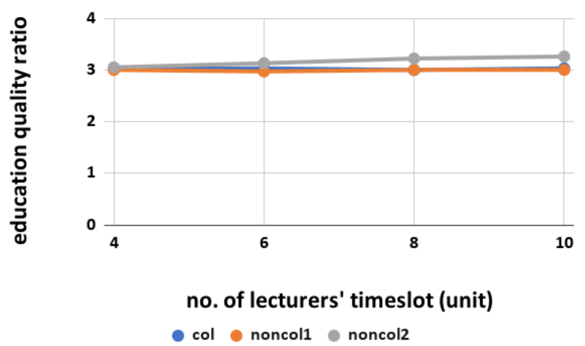
(b)

Fig. 3. Relation between the Number of Classes and Observed Parameters: (a) Unserved Classes Ratio, (b) Education Quality Ratio.

The second simulation is conducted to observe the relation between the number of lecturer's timeslots and the observed parameters. The number of lecturer's timeslots ranges from 4 to 8 with 2 step size. Other adjusted parameters are set to default. The result is shown in Fig. 4.



(a)



(b)

Fig. 4. Relation between the Number of Lecturers' Timeslot and Observed Parameters: (a) Unserved Classes Ratio, (b) Education Quality Ratio.

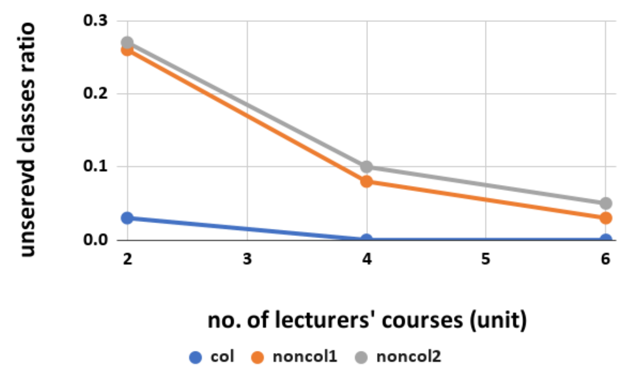
Fig. 4a shows that the increase of the number of lecturers' timeslot has made the unserved classes ratio decrease. It occurs in all models. Comparing among models, the proposed model performs as the best model. In the beginning, the unserved classes ratio is already very low. Then, the proposed model produces a zero unserved classes ratio when the number of lecturers' timeslot is higher than or equal to 6 timeslots. Meanwhile, the unserved classes ratio decreases significantly due to the decrease in the number of lecturers' timeslot but never reaches zero unserved classes ratio. In the beginning, the unserved classes ratio of the proposed model is only 8 percent of the non-collaborative models. It can be said that the proposed collaborative model reduces the unserved classes ratio by 92 percent relative to the non-collaborative models. The first non-collaborative model is better than the second non-collaborative model. The gap between these two non-collaborative models becomes wider due to the increase in the number of lecturers' timeslot.

Fig. 4b shows that the increase of the lecturers' timeslot creates different responses depending on the model. The education quality ratio tends to be stagnant for the proposed model and the first non-collaborative model. Meanwhile, the education quality ratio tends to increase less significantly due to the increase of the number of lecturers' timeslot for the second non-collaborative model. The performance is almost equal between the models, but the second non-collaborative model has little advantage.

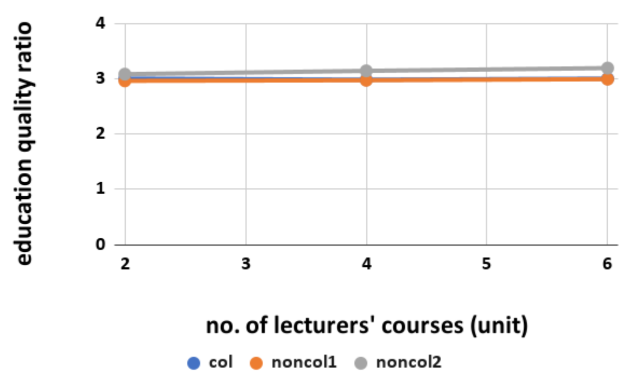
The third simulation is conducted to observe the relation between the number of lecturer's courses and the observed parameters. The result is shown in Fig. 5. The number of lecturer's courses ranges from 2 to 6 with 2 step size. Other adjusted parameters are set to default.

Fig. 5a shows that the number of lecturers' courses is inversely proportional to the unserved classes ratio. It occurs in all models. Comparing among models, the proposed model performs as the best model. In the beginning, the proposed model creates a very low unserved classes ratio. Then, the proposed model creates a zero unserved classes ratio when the number of lecturers' courses is higher than or equal to 4. Meanwhile, the first non-collaborative model performs better than the second non-collaborative model. In the beginning, the unserved classes ratio of the proposed model was only 11 percent of the non-collaborative model. It means that the proposed model reduces the unserved classes ratio of the non-collaborative model to 89 percent.

Fig. 5b shows that the increase in the number of lecturers' courses does not significantly affect the education quality ratio. The education quality ratio tends to be stagnant for the proposed model and the first non-collaborative model. Meanwhile, the education quality ratio increases less significantly for the second non-collaborative model.



(a)



(b)

Fig. 5. Relation between the Number of Lecturers' Courses and Observed Parameters: (a) Unserved Classes Ratio, (b) Education Quality Ratio.

V. Discussion

There are several findings due to the simulation result. The proposed collaborative model is better in minimizing the unserved classes. This proposed model becomes the best model compared with both two non-collaborative models. The proposed collaborative model is competitive enough in maximizing the education quality. Meanwhile, the education quality among models tends to be equal.

All three adjusted parameters are inversely proportional to the unserved classes. The reason is as follows. The higher number of classes with the same number of courses makes the matching process easier. It is because in this simulation, the number of lecturers is proportional to the number of classes. The increase of the lecturers' timeslot also minimizes the unserved classes. It is because the lecturers' availability increases too due to the class predetermined timeslot. The increase of the lecturers' courses also minimizes the unserved classes. But the number of lecturers' timeslot is more significant than the number of lecturers' courses.

The simulation result shows that all three adjusted parameters do not significantly affect education quality. The reason is that education quality is put as the secondary objective during the optimization process. The new solution can replace the existing solution only if its unserved classes ratio is lower. It means that the unserved classes ratio is more prioritized than education quality. It is different from the model that adopts other multi-objective methods, such as non-dominated sorting, as it is used in the non-dominated sorting genetic algorithm (NSGA II) [29] or weighted sum method [30]. In these two methods, all criteria are treated equally. The NSGA II promises pareto optimal [29]. In NSGA II, a solution is better than another if it meets two rules. The first rule is that this solution is better or equal to its opponent in all parameters [29]. The second rule is that this solution is better than its opponent, at least in a parameter [29]. Meanwhile, the weighted sum method is simpler. It is conducted by aggregating all weighted parameters [30]. The weight represents the priority.

Finally, the result shows that the collaborative model tends to be better than the existing non-collaborative models as it becomes the main reason for this work. The collaborative model is proven to improve the quality of service in the context of reducing the unserved classes. The unserved class has become the classic issue in many operations research studies in the education area. In general, despite the chosen optimization method, reducing unserved classes is conducted by increasing the resources (rooms, lecturers, timeslots, and so on). This work shows that the unserved classes can be minimized without increasing resources through collaboration. This collaboration allows idle resources to be transferred to the more needed demand. But this collaboration occurs due to the existence of online learning so that the class can be conducted without physical appearance. This result also strengthens the statement that collaboration or resource sharing can give comparative advantage [31], for example, in improving the utility rate of resources and efficiency [32].

This theoretical result can be used as basis for the practical use in the online collaborative education system. Every institution (university) can focus on the three aspects (number of classes, number of lecturer's timeslot, and number of lecturer's courses) to reduce the unserved classes. University can shift more classes provided in its own institution to be conducted in the collaborative system. It means that the opportunity of these classes will be conducted by lecturers from outside of the institution will be higher. Reciprocally, the institution can push more lecturers to join the collaborative system of more timeslots to be allocated in the collaborative system. Finally, every institution should encourage its lecturers to conduct more courses. In the current non collaborative system, a lecturer is difficult to teach other courses because these courses have been assigned to the colleagues. On the other side, in the collaborative system, the opportunity to teach beyond the lecturer's traditional courses is wider. This circumstance gives benefit for both parties. The institution will benefit by the reduction of the unserved classes. The lecturers will benefit by improving their skill, competence, and experience.

VI. Conclusion

This work has demonstrated that the proposed collaborative model has met the objective of minimizing the unserved classes and maximizing the education quality. The simulation result shows that the proposed collaborative model outperforms both non-collaborative models in minimizing the unserved classes significantly. The proposed model reduces 89 to 92 percent of the unserved classes ratio compared with the non-collaborative models. On the other hand, the proposed collaborative model performs equally with the non-collaborative models to maximize education quality. The reason is that in this work, the unserved classes are more prioritized than the education quality, so minimizing the unserved classes becomes the primary objective while maximizing the education quality becomes the secondary objective.

This work has several limitations so that it can become the baseline for future improvements. This work has not discussed the financial aspect due to the collaborative approach. In general, any proposed approach should give financial incentives. Without financial incentives, universities will hesitate to adopt any collaborative approaches. Based on it, it is potential and important to propose a financial model that follows the collaborative assignment model. This financial incentive can be obtained through the efficiency of reducing the unserved classes without additional resources, i.e., lecturers or rooms. It means that this saving can be distributed to the existing lecturers and institutions. Second, this financial incentive should be transferred from the institution who owns the class to the institution whose lecturer conducts this class. This financial model should give win-win solution for both institutions.

Acknowledgment

This work was financially supported by Telkom University, Indonesia.

References

[1] M. C. Chen, S. N. Sze, S. L. Goh, N. R. Sabar, and G. Kendall, "A survey of university course timetabling problem: perspectives, trends and opportunities", IEEE Access, vol. 9, pp. 106515-106529, 2021.

[2] M. Lindahl, A. Mason, T. Stidsen, and M. Sorensen, "A strategic view of university timetabling", European Journal of Operational Research, vol. 266, no. 1, pp. 35-45, 2018.

[3] N. M. Arratia-Martinez, C. Maya-Padron, and P. A. Avila-Torres, "University course timetabling problem with professor assignment", Mathematical Problems in Engineering, article ID: 6617177, pp. 1-9, 2021.

[4] C. Mallick, S. K. Bhoi, K. K. Jena, K. S. Sahoo, M. Humayn, and M. H. Shahd, "CLAPS: Course and lecture assignment problem solver for educational institution using Hungarian method", Turkish Journal of Computer and Mathematics Education, vol. 12, no. 10, pp. 3085-3092, 2021.

[5] J. Nourmohammadi-Khiarak, Y. Zamani-Harghalani, and M. R. Feizi-Derakhshi, "Combined multi-agent method to control inter-department common events collision for university courses timetabling", Journal of Intelligent Systems, vol. 29, no. 1, pp. 110-126, 2020.

[6] A. Muklason, R. G. Irianti, and A. Marom, "Automated course timetabling optimization using tabu-variable neighborhood search based hyper-heuristic algorithm", Procedia Computer Science, vol. 161, pp. 656-664, 2019.

[7] D. Hermawan, "The rise of e-learning in covid-19 pandemic in private university: challenges and opportunities", International Journal of Recent Educational Research (IJORER), vol. 2, no. 1, pp. 86-95, 2021.

[8] S. Dhawan, "Online learning: a panacea in the time of covid-19 crisis", Journal of Educational Technology Systems, vol. 49, no. 1, pp. 5-22, 2020.

[9] M. M. Zalat, M. S. Hamed, and S. A. Bolbol, "The experiences, challenges, and acceptance of e-learning as a tool for teaching during the covid-19 pandemic among university medical staff", PLOS ONE, vol. 16, no. 3, pp. 1-12, 2021.

[10] E. Torabzadeh and M. Zandieh, "Cloud theory-based simulated annealing approach for scheduling in the two-stage assembly flowshop", Advances in Engineering Software, vol. 41, no. 10-11, pp. 1238-1243, 2010.

[11] H. R. Moshtaghi, A. T. Eshlaghy, and M. R. Motadel, "A comprehensive review on meta-heuristic algorithms and their classification with novel approach", Journal of Applied Research on Industrial Engineering, vol. 8, no. 1, pp. 63-89, 2021.

[12] P. D. Kusuma and A. S. Albana, "University course timetabling model in joint courses program to minimize the number of unserved requests", International Journal of Advanced Computer Science and Applications, vol. 12, no. 10, pp. 121-127, 2021.

[13] P. D. Kusuma and D. Adiputra, "Lecturer-course assignment model in national joint courses program to improve education quality and lecturers' time preference", International Journal of Intelligent Engineering & Systems, vol. 15, no. 1, pp. 361-369, 2022.

[14] S. Faudzi, S. Abdul-Rahman, and R. A. Rahman, "An assignment problem and its application in education domain: a review and potential path", Advances in Operations Research, vol. 2018, article ID: 8958393, pp. 1-19, 2018.

[15] G. Homsi, J. Jordan, S. Martello, and M. Monaci, "The assignment and loading transportation problem", European Journal of Operational Research, vol. 289, no. 3, pp. 999-1007, 2021.

[16] M. R. Hassan, H. Hamdy, S. Alkhalaf, A. M. B. Eldin, M. Ahmed, and A. M. Hemeida, "Optimal components assignment problem for a flow network with node failure using fuzzy optimization", Ain Shams Engineering Journal, 2021, in press.

[17] A. Dhanapal, M. Jayalakshmi, K. Kavitha, and B. Sajad, "Solid assignment problem in manufacturing logistics", International Journal of Civil Engineering and Technology, vol. 10, no. 2, pp. 1846-1850, 2019.

[18] A. Rezaeipanah, Z. Abshirini, and M. B. Zade, "Solving university course timetabling problem using parallel genetic algorithm", International Journal of Scientific Research in Computer Science and Engineering, vol. 7, no. 5, pp. 5-13, 2019.

[19] M. H. Cruz-Rosales, M. A. Cruz-Chavez, F. Alonso-Pecina, J. C. Peralta-Abarca, E. Y. Avila-Melgar, B. Martinez-Bahena, and J. Enriquez-Urbano, "Metaheuristic with cooperative process for the university course timetabling problem", Applied Sciences, vol. 12, ID: 542, pp. 1-30, 2022.

[20] D. S. Vianna, C. B. Martins, T. J. Lima, M. F. D. Vianna, and E. B. M. Meza, "Hybrid VNS-TS heuristics for university course timetabling problem", Brazilian Journal of Operations & Production Management, vol. 17, no. 2, ID:e2020683, pp. 1-20, 2020.

[21] N. Pillay and E. Ozcan, "Automated generation of constructive ordering heuristics for educational timetabling", Annals of Operations Research, vol. 275, pp. 181-208, 2019.

[22] F. H. Awad, A. Al-kubaisi, and M. Mahmood, "Large-scale timetabling problems with adaptive tabu search", Journal of Intelligent Systems, vol. 31, pp. 168-176, 2021.

[23] B. Wang, Y. Geng, and Z. Zhang, "Applying genetic algorithm to university classroom arrangement problem", Journal of Physics: Conference Series, vol. 1325, 2019.

[24] M. Assi, B. Halawi, and R. A. Haraty, "Genetic algorithm analysis using the graph coloring method for solving the university timetable problem", Procedia Computer Science, vol. 126, pp. 899-906, 2018.

[25] S. L. Goh, G. Kendall, and N. R. Sabar, "Improved local search approaches to solve the post enrolment course timetabling problem", European Journal of Operational Research, vol. 261, pp. 17-29, 2017.

[26] S. L. Goh, G. Kendall, and N. R. Sabar, "Monte Carlo tree search in finding feasible solutions for course timetabling problem", International Journal on Advanced Science, Engineering and Information Technology, vol. 9, no. 6, pp. 1936-1943, 2019.

[27] S. L. Goh, G. Kendall, N. R. Sabar, and S. Abdullah, "An effective hybrid local search approach for the post enrolment course timetabling problem", OPSEARCH, vol. 57, pp. 1131-1163, 2020.

[28] I. E. Wicaksono and W. W. Wisesa, "An optimization model for teaching assignment based on lecturer's capability using linear programming", Indonesian Journal of Artificial Intelligence and Data Mining, vol. 3, no. 2, pp. 57-63, 2020.

[29] S. Verma, M. Pant, and V. Snasel, "A comprehensive review on NSGA-II for multi-objective combinatorial optimization problems", IEEE Access, vol. 9, pp. 57757-57791, 2021.

[30] S. Stoilova, "An integrated multi-criteria and multi-objective optimization approach for establishing the transport plan of intercity trains", Sustainability, vol. 12, ID: 687, pp. 1-24, 2020.

[31] T. Otcenaskova, K. Kolerova, and V. Bures, "Sharing of resources: theoretical background and a case study", Procedia – Social and Behavioral Sciences, vol. 109, pp. 698-705, 2014.

[32] T. Becker and H. Stern, "Impact of resource sharing in manufacturing on logistical key figures", Procedia CIRP, vol. 41, pp. 579-584, 2016.

A Novel Morphological Analysis based Approach for Dynamic Detection of Inflected Gujarati Idioms

Jatin C. Modh[1]
Research Scholar
Gujarat Technological University
Ahmedabad, India

Jatinderkumar R. Saini[2]*
Symbiosis Institute of Computer Studies and Research, Symbiosis International (Deemed University), Pune, India

Ketan Kotecha[3]
Symbiosis Centre for Applied Artificial Intelligence, Symbiosis International (Deemed University), Pune, India

***Abstract*—The Gujarati language is primarily spoken by Gujarati people living in the state of Gujarat, India. It is the medium of communication in the state of Gujarat. In the Gujarati language, 'rudhiprayog' i.e. idiom is a very much popular form of expression. It represents the real flavour of the Gujarati language. The idiom is a group of words saying one thing literally but means something else when we explored it in context. Like Gujarati verbs, idioms can be written in many forms. Due to different morphological forms of the same Gujarati idiom, Gujarati idiom identification is a challenging task for any machine translation system (MTS). Accordingly many forms also make idiom translation more complicated. In the current paper, Gujarati idioms in their different inflected forms are collected, analyzed and classified based on ending words. After classifying idioms, their base or root forms are identified. Base idiom form and their possible idioms forms are morphologically analyzed and rules are generated based on the relationship between base form and possible inflected forms of idioms. These rules are used to generate possible idiom forms from the base idiom form. Gujarati idiom in any valid inflected form can be dynamically detected from the Gujarati input text using the proposed novel morphological analysis based approach. The results are encouraging enough to implement the proposed model of rules for natural language processing tasks as well as a machine translation system for Gujarati language idioms.**

Keywords*—Gujarati; idiom; machine translation system (MTS); morphological analysis; natural language processing (NLP); rudhiprayog

I. Introduction

Gujarati is the official language of Gujarat state of India and also recognized by the constitution of India. Gujarati is the Indo-Aryan branch of the Indo-European language family. It is spoken by more than 46 million people. Most of the people speaking Gujarati live in the Gujarat state. Also Gujarati communities spread in the UK, USA and all around the world. Gujarati language is used in newspapers, magazines, television, education, business, communication and in all type of media. Gujarati is written using the Devnagari script. In Gujarati language, two numbers singular and plural are used. Three genders masculine, feminine and neuters are used. Three cases nominative, oblique and locative are used. A Gujarati verb corresponds with a person, number and gender. These all are marked by suffixes attached to the verb root. These make Gujarati inflection fairly complex [1][2][3].

A. Gujarati Idioms

Gujarati idioms are used in many forms for day-to-day communication in Gujarati language. Idioms are used by Gujarati people for expressing feelings and thoughts. Machine translation of Gujarati idioms is important for the communication with the other non-Gujarati people. Idiom is the phrase whose collective meaning is not the same as literal meaning of its individual words. Gujarati idioms like any other language idioms cannot be translated literally.

B. Gujarati Idioms with Static Forms

Static idioms are the idioms where its different inflected forms are not possible. Static idioms can have only single form in Gujarati language. For example, અક્કલનો ઓથમીર 'akkalano othamira' (i.e. stupid) is the idiom in which other inflected form of idiom does not make any sense. Another Gujarati idiom example is એદીનો અખાડો 'edino akhado' (i.e. very lazy) where other inflected and valid idiom forms are not possible.

C. Gujarati Idioms with Inflected Forms

Inflected idioms are the idioms where its different morphological forms are possible. Infected idioms can be used by its different forms in the sentence. Inflected idioms are generally ended with Gujarati inflected verb form. Inflection can be applied on the last word of idiom that is usually base verb form. For example, ફાચર માર 'phacara mara' (i.e. to disrupt) is the base form of idiom. Its various inflected forms like ફાચર મારવી 'phacara maravi', ફાચર મારી 'phacara mari', ફાચર મારીને 'phacara marine', ફાચર મારવામાં 'phacara maravamam' are possible in the Gujarati sentence. Inflection is applied on the last word of base form i.e. on word માર here. Another example, ખટકો રાખ 'khatako rakha' (i.e. keep in mind) is the base form and its possible forms used in the sentences are ખટકો રાખ 'khatako rakha', ખટકો રાખવા 'khatako rakhava', ખટકો રાખવામા 'khatako rakhavama', ખટકો રાખવો 'khatako rakhavo', ખટકો રાખી 'khatako rakhi', ખટકો રાખીને 'khatako rakhine', ખટકો રાખેલા 'khatako rakhela', ખટકો રાખેલો 'khatako rakhelo', ખટકો રાખો 'khatako rakho', ખટકો રાખ્યો 'khatako rakhyo'; Inflected forms are applied on the last word રાખ 'rakha' here.

*Corresponding Author.

D. Stemming of Idioms

Stemming is the important process in natural language processing. The purpose of stemming is to standardize words to their common base form. Stemming removes the affixes of the word to get the root word or base word or stem word [4][5][6]. For example, ફાચર મારવી 'phacara maravi', ફાચર મારી 'phacara mari', ફાચર મારીને 'phacara marine' are various inflected idiom forms used in the sentences; stem/base form of these idioms form is ફાચર માર 'phacara mara' (i.e. to disrupt). Another example, ખટકો રાખ 'khatako rakha' (i.e. keep in mind) is the stem/base form of various inflected idiom forms like ખટકો રાખવા 'khatako rakhava', ખટકો રાખવામા 'khatako rakhavama', ખટકો રાખવો 'khatako rakhavo', ખટકો રાખી 'khatako rakhi', ખટકો રાખીને 'khatako rakhine' etc. So ફાચર માર 'phacara mara' (i.e. to disrupt) and ખટકો રાખ 'khatako rakha' (i.e. keep in mind) are the stem form of idioms. This stem/base form is stored once in the idiom database to recognize the idioms from the input text. Using this stem form, other inflected idioms forms can be generated.

The rest of the paper is organized as follows: Section II represents the literature review related to Gujarati idioms translation and its identification from the Gujarati text; Section III covers the methodology including idioms data collection and the steps of proposed model; Results and analysis are discussed in Section IV; finally conclusion, limitations and future work are discussed in Section V.

II. Related Literature Review

Various machine translation system projects have been carried out, especially for the English idiomatic text. Few projects have been taken out for non-English languages containing idioms.

Microsoft Translator and Google Translate are machine translation systems support Gujarati language translation also. Microsoft Translator [7][8] and Google Translate [9][10] use Microsoft cognitive services and Google Neural machine translation system respectively. Both machine translation systems are very much accepted but failed in the correct translation of Gujarati idiomatic text.

Authors [11] employed a GIdTra for translating only Gujarati bigram idioms into English language. They used dictionary based approach for identifying Gujarati idioms from the input text. They focused bigram idioms to test their approach. Other forms of bigram idiom cannot be identified.

Modh and Saini [12] discussed various machine translation approaches for the Gujarati language. The researchers have implemented a context-based Machine Translation System (MTS) for translating Gujarati bigram and trigram idioms into the English language [13].

Researchers [14] experimented n-gram model and used IndoWordNet for getting synonyms of surrounding words of particular idiom. They also exercised various context windows sizes in the case of ambiguity in finding meaning. They also worked on diacritics and suffix based rules [15].

Muzny and Zettlemoyer [16] applied a supervised approach for the automatic identification of English idioms from the corpus of Wiktionary multi-word definitions. They claimed 65% precision level of accuracy.

Verma and Vuppuluri [17] experimented with the combination of dictionary knowledge and web knowledge for English idiom identification. Authors claimed their approach as language as well as domain-independent. They also accepted about non-availability of the meanings of idiom phrases.

Hubers et al. [18] studied on whether the age and emigration length affects the knowledge of idiomatic expressions in the Dutch language or not. They concluded that emigration length negatively affects emigrant idiom knowledge.

Kessler and Friedrich [19] experimented on 9-to-10-year old children whether children can predict idiom-final words. The authors concluded that children rapidly activate multi-word units for idioms and decompose them only after a short delay.

Bakir and Umbu [20] studied American Pie movie script and analyzed the typology of idiomatic expression and contextual meanings of English idioms. Ramadhan et al. [21] prepared a report on the adoption of English movies in learning English idioms for English undergraduate students.

Based on this literature review and study, researchers dealing with idioms face problems in accurate understanding, identification and translation of idioms. Very less work is done especially for the analysis and translation of Gujarati idioms. No researchers have tried to recognize Gujarati idioms from the Gujarati inputted text. No researchers have analyzed various idiom forms of Gujarati language.

The paper focuses on the study of Gujarati idioms and its various forms used in the sentences. The scope of this paper is to generate rules from the analysis of Gujarati idioms collection with their possible forms and implementing these rules in the algorithm for finding all types of Gujarati idioms within inputted Gujarati text. This implementation helps in the simplification and translation of Gujarati idioms to any language in the world.

III. Methodology

Different Gujarati idioms are collected and categorized on the basis of static idioms and inflected idioms. Inflected idioms are again sub-categorized on the basis of different ending words. On the basis of analysis of collected Gujarati idioms, rules and base forms of idioms are generated. All base forms of idioms are stored in the idiom database. Finally, the idiom database and these rules are used in the dynamic generation of different forms of the same Gujarati idiom. This dynamic idiom generation helps in identifying all Gujarati idioms from the inputted text.

A. Data Collection

Overall 3410 distinct Gujarati idioms are collected from different Gujarati language sources [22][23]. From these idioms, 6047 valid different Gujarati idiom forms are collected. For example, ખટકો રાખવો 'khatako rakhavo' idiom is collected; its distinct and base form is ખટકો રાખ 'khatako rakha' (i.e. keep in mind). From the base form, many inflection forms are possible by adding suffixes and diacritics but valid inflected forms used in Gujarati language are ખટકો રાખવા 'khatako rakhava', ખટકો રાખવો 'khatako rakhavo', ખટકો રાખી 'khatako rakhi', ખટકો રાખીને 'khatako rakhine', ખટકો રાખેલો 'khatako rakhelo', ખટકો રાખો 'khatako rakho', ખટકો રાખ્યો 'khatako rakhyo' etc. These inflected idioms are analyzed with their different possible forms for the derivation and finalization of the stemming rules for idioms. Base forms of all idioms are stored in the idiom database for further processing.

B. Idiom Classification

By analyzing the idioms collection, it is found that the idioms can be classified by three ways: (1) n-gram wise where n=1 to 9 (2) m-meaning wise idioms where m=1 to 7 (3) Static idioms and inflected idioms. In this paper, third classification is focused to generate the idiom identification rules. Static idioms are the idioms where only single idiom form is possible and therefore stemming is not appropriate. Inflected idioms are the idioms where various idiom forms from the base form can be generated. In the current paper, inflected idioms are analyzed. Various idioms with its inflected forms are analyzed to derive stem or base word form of particular idiom.

Table I shows the two types of idioms and its count. Static Gujarati idioms count is 215 and inflected Gujarati idioms count is 5832. Static idioms are the idioms where only single form of idiom is possible so stemming is not meaningful whereas inflected idioms are the idioms where stemming can be applied to derive base form or stem word. Inflected Gujarati idioms are usually ended with verb forms. This base verb form can generate other inflected verb forms of idioms by adding suffixes. For example, ફાચર મારવી 'phacara maravi', ફાચર મારી 'phacara mari', ફાચર મારીને 'phacara marine' are the inflected idioms; it is ended with the different verb forms of માર 'mara'; so base form of these idioms is ફાચર માર 'phacara mara'.

Inflected idioms can further be classified on the base of end words. Inflected idioms can be of four categories on the base of end character(s) or word. (1) Idioms end with વું 'vum' (2) Idioms end with વા 'va' or વાં 'vam' (3) Idioms end with વી 'vi' (4) Idioms end with વો 'vo'. Table II shows these four categories of idioms and their counts.

TABLE I. TYPES OF GUJARATI IDIOMS WITH REFERENCE TO STEMMING

Sr. No.	Types of Idioms	Count	Remarks
1	Static idioms	215	Not possible to derive stem/base word. Only single form of idiom is possible. No similar structure in idioms. Example એદીનો અખાડો 'edino akhado' i.e. very lazy, બારમો ચંદ્રમા 'baramo candrama' i.e. animosity, આડી વાટની ધૂળ 'adi vatani dhula' i.e. fruitless work, ઓઢવા આભ ને પાથરવા ધરતી 'odhava abha ne patharava dharati' i.e. very miserable situation, ગાંઠનું ગોપીચંદન 'ganthanum gopicandana' i.e. at own cost
2	Inflected idioms	5832	Stem word or base form can be derived. Other idiom forms can be generated from the base form. Idiom is usually ending with the base verb form. Example ફાચર મારવી 'phacara maravi'→ ફાચર માર 'phacara mara' ખટકો રાખવો 'khatako rakhavo'→ ખટકો રાખ 'khatako rakha' Here ફાચર માર phacara mara' is the base form of idiom ફાચર મારવી 'phacara maravi' and માર 'mara' is verb form.
	Total	6047	

TABLE II. INFLECTED IDIOMS WHERE STEMMING CAN BE APPLIED

Sr. No.	Inflected Idioms end with the word	Count	Example
1	વું 'vum'	2534	ગોથું ખાવું 'gothum khavum' i.e. make a mistake
2	વા OR વાં 'va' 'vam'	741	આકડા વાવવા 'akada vavava' i.e. planting the roots of animosity ઝાવાં નાખવાં 'jhavam nakhavam' i.e. to boggle
3	વી 'vi'	1370	આંખ ઠરવી 'ankha tharavi' i.e. to be satisfied
4	વો 'vo'	1187	ખોળો ખાલી હોવો 'kholo khali hovo' i.e. to be childless
	Total	5832	

Category 1 Idioms end with word વું : All idioms end with word વું are studied and can be further classified as shown in Table III; Idioms end with word એવું are 18, idioms end with word કવું are 68, idioms end with word કાવું are 2, idioms end with word ખવું are 133, idioms end with word ખાવું are 36, idioms end with word ગવું are 52 and so on. Corresponding base form of these idioms are derived as એવ, મૂક, સુક, રાખ, ખાવ, લાગ and so on. Table III shows a snapshot of partial data for the sub-categories of idioms end with word વું.

TABLE III. SNAPSHOT OF PARTIAL DATA OF IDIOMS END WITH THE WORD વું 'VUM'

Sr No.	Idioms end with the word	Count	Base or stem word derivation
1	એવું 'evum'	18	એવું 'evum'→ એવ 'eva'
2	કવું 'kavum'	68	મૂકવું 'mukavum'→ મૂક 'muka'
3	કાવું 'kaavum'	2	સુકાવું 'sukavum'→ સુક 'suka'
4	ખવું 'khavum'	133	રાખવું 'rakhavum'→ રાખ 'rakha'
5	ખાવું 'khaavum'	36	ખાવું 'khavum'→ ખાવ 'khava'
6	ગવું 'gavum'	52	લાગવું 'lagavum'→ લાગ 'laga'
7	ચવું 'chavum'	10	રચવું 'racavum'→ રચ 'raca'
8	ચાવું 'chaavum'	1	ગૂંચાવું 'guncavum'→ ગૂંચ 'gunca'
9	છવું 'chhavum'	6	પૂછવું 'puchavum' → પૂછ 'pucha'
10	જવું 'javum'	289	આંજવું 'anjavum '→ આંજવ 'anjava'

Category 2 Idioms end with word વા or વાં: All idioms end with word વા or વાં are studied and can be further classified as shown in Table IV; Idioms end with word ઝવા are 1, idioms end with word ટવા are 6, idioms end with word ટવાં are 2, idioms end with word ઠવા are 4 and so on. Corresponding base form of these idioms are derived as સૂઝ, પલટ, વાટવ, ઊઠ and so on. Table IV shows a snapshot of partial data for the sub-categories of idioms end with word વા or વાં.

TABLE IV. SNAPSHOT OF PARTIAL DATA OF IDIOMS END WITH THE WORD વા 'VA' OR વાં 'VAM'

Sr No.	Idioms end with the word	Count	Base or stem word derivation
1	ઝવા 'jhava'	1	સૂઝવા 'sujhava'→ સૂઝ 'sujha'
2	ટવા 'tava'	6	પલટવા 'palatava' → પલટ 'palata'
3	ટવાં 'tavam'	2	વાટવાં 'vatavam'→ વાટવ 'vatava'
4	ઠવા 'thava'	4	ઊઠવા 'uthava'→ ઊઠ 'utha'
5	ઠવાં 'thavam'	4	ઊઠવાં 'uthavam'→ ઊઠ 'utha'
6	ડવા 'dava'	57	ઊડવા 'udava'→ ઊડ 'uda'
7	ડવાં 'davam'	32	પડવાં 'padavam'→ પડ 'pada'
8	ઢવા 'dhava'	12	કાઢવા 'kadhava'→ કાઢ 'kadha'
9	ણવા 'nava'	11	ગણવા 'ganava'→ ગણ 'gana'
10	થવા 'thavam'	16	થવા 'thava'→ થવ 'thav'

Category 3 Idioms end with word વી: All idioms end with word વી are studied and can be further classified as shown in Table V; Idioms end with word તવી are 3, idioms end with word થવી are 33, idioms end with word દવી are 5, idioms end with word દેવી are 21, idioms end with word ધવી are 27 and so on. Corresponding base form of these idioms are derived as જીત, થવ, ખોદ, દેવ, બાંધ and so on. Table V shows a snapshot of partial data for the sub-categories of idioms end with word વી.

TABLE V. SNAPSHOT OF PARTIAL DATA OF IDIOMS END WITH THE WORD વી 'VI'

Sr No.	Idioms end with the word	Count	Base or stem word derivation
1	તવી 'tavi '	3	જીતવી 'jitavi'→જીત 'jita'
2	થવી 'thavi'	33	થવી 'thavi'→ થવ 'thava'
3	દવી 'davi'	5	ખોદવી 'khodavi'→ખોદ 'khod'
4	દેવી 'devi'	21	દેવી 'devi'→ દેવ 'deva'
5	ધવી 'dhavi'	27	બાંધવી 'bandhavi'→બાંધ 'bandha'
6	પવી 'pavi'	32	ચાંપવી 'campavi' →ચાંપ 'campa'
7	બવી 'bavi'	4	દાબવી 'dabavi' →દાબ 'daba'
8	મવી 'mavi'	12	રમવી 'ramavi'→રમ 'rama'
9	રવી 'ravi'	269	મારવી 'maravi'→માર 'mara'
10	લવી 'lavi'	58	ખોલવી 'kolavi'→ખોલ 'khola'

Category 4 Idioms end with word વો: All idioms end with word વો are studied and can be further classified as shown in Table VI; Idioms end with word ડવો are 138, idioms end with word પવો are 51, idioms end with word મવો are 19, idioms end with word રવો are 183, idioms end with word લવો are 31 and so on. Corresponding base form of these idioms are derived as પાડ, કાપ, જામ, ભર, ખેલ and so on. Table VI shows a snapshot of partial data for the sub-categories of idioms end with word વો.

TABLE VI. SNAPSHOT OF PARTIAL DATA OF IDIOMS END WITH THE WORD વો 'VO'

Sr No.	Idioms end with the word	Count	Base or stem word derivation
1	ડવો 'davo'	138	પાડવો 'padavo'→પાડ 'pada'
2	પવો 'pavo'	51	કાપવો 'kapavo'→કાપ 'kapa'
3	મવો 'mavo'	19	જામવો 'jamavo'→જામ 'jama'
4	રવો 'ravo'	183	ભરવો 'bharavo'→ભર 'bhara'
5	લવો 'lavo'	31	ખેલવો 'kheavo'→ખેલ 'khela'
6	લેવો 'levo'	27	લેવો 'levo'→ લેવ 'leva'
7	વવો 'vavo'	139	આવવો 'avavo'→આવ 'ava'
8	સવો 'savo'	11	ખસવો 'khasavo'→ ખસ 'khasa'
9	હોવો 'hovo'	39	હોવો 'hovo'→ હોવ 'hova'
10	ળવો 'lavo'	53	ઢોળવો 'dholavo'→ઢોળ 'dhola'

C. Rules Derivation [for Dynamic Inflected Idioms Generation from the base form]

By studying and analyzing all four categories of inflected idioms specified in Table III to Table VI, base forms of all idioms are collected. Further, by detailed morphological analysis and using reverse rules generation process, rules are defined to generate various idiom forms from the given base form of particular idiom as shown in Table VII.

Table VII shows the collection of rules for generating dynamic inflected idiom forms from the idiom base form by adding diacritics as well as different suffix characters. For example: ખટકો રાખ 'khatako rakha' is the base idiom form as per Rule 1. As per rule 2, second inflected form ખટકો રાખવા 'khatako rakhava' can be generated by adding suffix 'વ' and diacritics '◌ા' to base idiom ખટકો રાખ 'khatako rakha'. As per rule 3, third inflected form ખટકો રાખવાં 'khatako rakhavam' can be generated by adding suffix 'વ' and two diacritics '◌ા', '◌ં' to base idiom ખટકો રાખ 'khatako rakha'. similar way using remaining rules overall 43 different inflected idioms forms can be generated. All other inflected idioms forms can be generated by adding different character વ 'va' ય 'ya' ન 'na' લ 'la' ઈ 'i' as well as by adding different diacritics as shown in Table VII.

TABLE VII. RULES TO GENERATE POSSIBLE FORMS OF IDIOMS FROM THE BASE IDIOM

Rule No.	Rules Definition	Base form	Inflected idiom forms
1	Base idiom	ખટકો રાખ	ખટકો રાખ
2	Base idiom+'વ'+'◌ા'	ખટકો રાખ	ખટકો રાખવા
3	Base idiom+'વ'+'◌ા'+'◌ં'	ખટકો રાખ	ખટકો રાખવાં
4	Base idiom+'વ'+'◌ા'+'મ'+'◌ા'	ખટકો રાખ	ખટકો રાખવામા
5	Base idiom+'વ'+'◌ા'+'મ'+'◌ા'+'◌ં'	ખટકો રાખ	ખટકો રાખવામાં
6	Base idiom+'વ'+'◌ી'	ખટકો રાખ	ખટકો રાખવી
7	Base idiom+'વ'+'◌ુ'	ખટકો રાખ	ખટકો રાખવુ
8	Base idiom+'વ'+'◌ુ'+'◌ં'	ખટકો રાખ	ખટકો રાખવું
9	Base idiom+'વ'+'◌ો'	ખટકો રાખ	ખટકો રાખવો
10	Base idiom+'◌ા'	ખટકો રાખ	ખટકો રાખા
11	Base idiom+'◌ા'+'◌ં'	ખટકો રાખ	ખટકો રાખાં
12	Base idiom+'◌ા'+'ય'+'◌ુ'+'◌ં'	ખટકો રાખ	ખટકો રાખાયું
13	Base idiom+'◌ા'+'વ'	ખટકો રાખ	ખટકો રાખાવ
14	Base idiom+'◌ા'+'વ'+'◌ા'	ખટકો રાખ	ખટકો રાખાવા
15	Base idiom+'◌ા'+'વ'+'◌ા'+'◌ં'	ખટકો રાખ	ખટકો રાખાવાં
16	Base idiom+'◌ા'+'વ'+'◌ી'	ખટકો રાખ	ખટકો રાખાવી
17	Base idiom+'◌ા'+'વ'+'◌ુ'+'◌ં'	ખટકો રાખ	ખટકો રાખાવું
18	Base idiom+'◌ા'+'વ'+'◌ો'	ખટકો રાખ	ખટકો રાખાવો
19	Base idiom+'◌ી'	ખટકો રાખ	ખટકો રાખી
20	Base idiom+'◌ી'+'ન'+'◌ે'	ખટકો રાખ	ખટકો રાખીને
21	Base idiom+'◌ુ'	ખટકો રાખ	ખટકો રાખુ
22	Base idiom+'◌ુ'+'◌ં'	ખટકો રાખ	ખટકો રાખું
23	Base idiom+'◌ે'	ખટકો રાખ	ખટકો રાખે
24	Base idiom+'◌ે'+'લ'+'◌ા'	ખટકો રાખ	ખટકો રાખેલા
25	Base idiom+'◌ે'+'લ'+'◌ો'	ખટકો રાખ	ખટકો રાખેલો
26	Base idiom+'◌ે'+'વ'+'◌ા'	ખટકો રાખ	ખટકો રાખેવા
27	Base idiom+'◌ે'+'વ'+'◌ી'	ખટકો રાખ	ખટકો રાખેવી
28	Base idiom+'◌ે'+'વ'+'◌ુ'	ખટકો રાખ	ખટકો રાખેવુ
29	Base idiom+'◌ે'+'વ'+'◌ુ'+'◌ં'	ખટકો રાખ	ખટકો રાખેવું
30	Base idiom+'◌ે'+'વ'+'◌ો'	ખટકો રાખ	ખટકો રાખેવો
31	Base idiom+'◌ો'	ખટકો રાખ	ખટકો રાખો
32	Base idiom+'◌ો'+'ઈ'	ખટકો રાખ	ખટકો રાખોઈ
33	Base idiom+'◌ો'+'વ'+'◌ા'	ખટકો રાખ	ખટકો રાખોવા
34	Base idiom+'◌ો'+'વ'+'◌ા'+'◌ં'	ખટકો રાખ	ખટકો રાખોવાં
35	Base idiom+'◌ો'+'વ'+'◌ી'	ખટકો રાખ	ખટકો રાખોવી
36	Base idiom+'◌ો'+'વ'+'◌ુ'	ખટકો રાખ	ખટકો રાખોવુ
37	Base idiom+'◌ો'+'વ'+'◌ુ'+'◌ં'	ખટકો રાખ	ખટકો રાખોવું
38	Base idiom+'◌ો'+'વ'+'◌ો'+' +	ખટકો રાખ	ખટકો રાખોવો
39	Base idiom+'◌્'+'ય'+'◌ા'	ખટકો રાખ	ખટકો રાખ્યા
40	Base idiom+'◌્'+'ય'+'◌ા'+'◌ં'	ખટકો રાખ	ખટકો રાખ્યાં
41	Base idiom+'◌્'+'ય'+'◌ુ'	ખટકો રાખ	ખટકો રાખ્યુ
42	Base idiom+'◌્'+'ય'+'◌ુ'+'◌ં'	ખટકો રાખ	ખટકો રાખ્યું
43	Base idiom+'◌્'+'ય'+'◌ો'	ખટકો રાખ	ખટકો રાખ્યો

Rules definitions specified in Table VII are applied on the base idiom form to generate all possible idiom forms dynamically. These rules help in searching any inflected idiom form available in the input text.

D. Idiom Database Creation

Database of idioms is mainly required to store the distinct base form of idiom and its simplified Gujarati meaning. Idiom database is created with idiom column and other related columns like Gujarati meaning, English meaning etc. Static idioms are having single form so they are stored in idiom column as it is. For inflected idioms, their base form is stored once in the idiom column. Gujarati meaning column stores the meaning of particular Gujarati idiom in simple Gujarati words.

E. Proposed Model

Fig. 1 shows the algorithm steps for the proposed model.

Step-1: Input the Gujarati text OR Input the file with encoding format UTF-8

Step-2: Apply preprocessing steps on it

2.1:Remove extra spaces from the both end of text

2.2:Remove all whitespace spaces in between the text

2.3:Remove special characters (\ , " < > . ? ; : !)

Step-3: Tokenize the words of input text

Step-4: Search idioms from the input text by implementing rules generated in Table VII and comparing with idiom column of idiom database.

Step-5: Display result (all idiom forms detected and replaced with Gujarati meaning)

Fig. 1. Algorithm for the Proposed Model.

Input is the Gujarati idiomatic text that may contain any number of Gujarati idioms. Output will be the Gujarati text without any idiom. In the output text, all idioms will be replaced with the value of Gujarati meaning column. Other Gujarati text will remain as it is. Output is nothing but Gujarati text without any idioms. For example,

INPUT TEXT= કોઇ પણ કામમાં ફાચર મારવી એ તેનો સ્વભાવ છે. 'ko'i pana kamamam phacara maravi e teno svabhava che'.

FINAL OUTPUT= કોઈ પણ કામમાં વચમાં અડચણ નાખવી એ તેનો સ્વભાવ છે. 'ko'i pana kamamam vacamam adacana nakhavi e teno svabhava che'.

IV. Result and Analysis

Overall 7400 different idiom forms were entered as input text. Input text may contain one or more idiom phrases. Output results were obtained. The correctness of output results was verified by the two linguists with Gujarati mother tongue and doctorate degree in Gujarati language. The proposed model was able to detect all inflected idiom forms present within Gujarati text successfully. The input text is nothing but the Gujarati idiomatic text using any inflected idiom form and the output shows the replacement of Gujarati idiom with simple Gujarati meaning of the particular idiom. For example;

Example 1:

INPUT TEXT= ખટકો રાખવા ખટકો રાખી

'khatako rakhava khatako rakhi'

FINAL OUTPUT= કોઈ વાત બરાબર ધ્યાનમાં રાખવી કોઈ વાત બરાબર ધ્યાનમાં રાખવી.

'ko'i vata barabara dhyanamam rakhavi ko'i vata barabara dhyanamam rakhavi'.

Example 2:

INPUT TEXT= વિદ્યાર્થીઓએ અભ્યાસ માટે ખટકો રાખવો જરૂરી બને છે. 'vidyarthi'o'e abhyasa mate khatako rakhavo jaruri bane che'.

FINAL OUTPUT= વિદ્યાર્થીઓએ અભ્યાસ માટે કોઈ વાત બરાબર ધ્યાનમાં રાખવી જરૂરી બને છે. 'vidyarthi'o'e abhyasa mate ko'i vata barabara dhyanamam rakhavi jaruri bane che'.

For understanding, ખટકો રાખ 'khatako rakha' (i.e. keep in mind) is the base form of idiom. Its many inflection forms are possible by adding suffix વ ય ન લ ઈ and/or by adding different diacritics. But only valid different forms are given for testing like ખટકો રાખવા, ખટકો રાખવામા, ખટકો રાખવો, ખટકો રાખી, ખટકો રાખીને, ખટકો રાખે, ખટકો રાખેલા, ખટકો રાખેલો, ખટકો રાખો, ખટકો રાખ્યા, ખટકો રાખ્યો; Proposed algorithm detects all the inflected forms of base form idiom ખટકો રાખ and so algorithm displays the output as Gujarati meaning of idiom as કોઈ વાત બરાબર ધ્યાનમાં રાખવી. Other Gujarati text will remain as it is in the output. Table VIII displays Output text for the given Input text. Output text is the same Gujarati meaning for any inflected idiom form of the same base idiom.

TABLE VIII. Different Inflected Forms of Idiom as Input and Corresponding Output

Sr No.	Input text	Output text (Gujarati meaning)
1	ખટકો રાખ 'khatako rakha'	કોઈ વાત બરાબર ધ્યાનમાં રાખવી 'ko'i vata barabara dhyanamam rakhavi'
2	ખટકો રાખવા 'khatako rakhava'	
3	ખટકો રાખવામા 'khatako rakhavama'	
4	ખટકો રાખવો 'khatako rakhavo'	
5	ખટકો રાખી 'khatako rakhi'	
6	ખટકો રાખીને 'khatako rakhine'	
7	ખટકો રાખે 'khatako rakhe'	
8	ખટકો રાખેલા 'khatako rakhela'	
9	ખટકો રાખેલો 'khatako rakhelo'	
10	ખટકો રાખો 'khatako rakho'	
11	ખટકો રાખ્યા 'khatako rakhya'	
12	ખટકો રાખ્યો 'khatako rakhyo'	
13	ફાચર માર 'phacara mara'	વચમાં અડચણ નાખવી 'vacamam adacana nakhavi'
14	ફાચર મારવી 'phacara maravi'	
15	ફાચર મારી 'phacara mari'	
16	ફાચર મારીને 'phacara marine'	
17	ફાચર મારવામાં 'phacara maravamam'	
18	ભાંગરો વાટ 'bhangaro vata'	છૂપી વાત ખુલ્લી કરવી 'chhupi vata khulli karavi'
19	ભાંગરો વાટી 'bhangaro vati'	
20	ભાંગરો વાટવો 'bhangaro vatavo'	
21	ભાંગરો વાટ્યો 'bhangaro vatyo'	
22	ભાંગરો વાટીને 'bhangaro vatine'	

Features of proposed algorithm are as follows:

1) Applied algorithm is domain independent. Proposed implementation detects all Gujarati inflected idioms used anywhere in the input text and replaces all Gujarati idioms with simple Gujarati meaning.

2) Proposed model stores base form of inflected idiom in the database as single record. For example base form ભાંગરો વાટ 'bhangaro vata' (i.e. disclose a secret) is stored in the idiom database once, but it is used to generate all possible forms of the same idiom like ભાંગરો વાટી 'bhangaro vati', ભાંગરો વાટવો 'bhangaro vatavo', ભાંગરો વાટ્યો 'bhangaro vatyo', ભાંગરો વાટીને 'bhangaro vatine' etc.

3) Dictionary based approach is applied for searching static idioms from the input text because static idioms are having the single possible idiom form. No rules are applicable on static idioms as they are found in irregular forms like અક્કલની ખાણ 'akkalani khana' i.e. very intelligent person, આંખનો પાટો 'ankhano pato' i.e. disgusting, કાચું સોનું 'kacum sonum' i.e. very fertile land, ખાસડાને તોલે 'khasadane tole' i.e inferior, બારે દહાડા ને બત્રીસે ઘડી 'bare dahada ne batrise ghadi' i.e. persistent, માથે રાત જેવું ધાબું 'mathe rata jevum dhabum' i.e. utter darkness, વરઘોડાની વાડી 'varaghodani vadi' i.e. transient, વાડ તેવો વેલો 'vada tevo velo' i.e. people follows king etc.

4) Inflected idioms are generally ended with the words વું 'vum', વા 'va', વાં 'vam', વી 'vi', વો 'vo'. These idioms are collected and analyzed in Table III to Table VI. Considering limited number of idioms, base idiom forms of these idioms are collected manually. Using reverse process, algorithm is developed to generate all idioms forms from the base form. Only base forms of idioms are stored once in the idiom database.

5) Proposed algorithm generates all possible forms of idioms by applying all Table VII rules to base idiom form for detection of any idiom in the input text. So it sometimes rectifies minor spelling mistakes in the Gujarati idiom form automatically. For example, ભાંગરો વાટ્ય 'bhangaro vatya' and ભાંગરો વાટવ 'bhangaro vatava' both are erroneous spelling forms of base idiom ભાંગરો વાટ 'bhangaro vata' i.e. disclose a secret; but the proposed algorithm considers and corrects both as ભાંગરો વાટ 'bhangaro vata' i.e. disclose a secret.

6) The proposed model is the first approach in Gujarati language that able to find out any valid and possible forms of Gujarati idioms present in the Gujarati text and provides Gujarati simplification of the particular idiom.

V. Conclusion, Limitations and Future Work

The proposed rule-based model was successfully implemented and it successfully detected all the static and inflected Gujarati idiom forms from the Gujarati text. The proposed algorithm successfully detected all the idioms by implementing a dictionary-based approach as well as dynamic idiom form generation rule-based approach. The proposed algorithm generates all possible idiom forms dynamically to determine whether any inflected form of a particular base form idiom is present in the input text or not.

The proposed system can detect any form of the idiom from the Gujarati text but the thing is that the particular idiom base form must be present in the idiom database. The proposed system could not identify idiom that is not available in the idiom database. Future work is to collect all Gujarati idioms from possible sources to rectify this shortcoming. Also algorithm applies all the generated rules to all base form of idioms for generating possible idiom forms.

Based on the results obtained, it is advocated that the proposed system is worth implementation in the real world for machine translation of the Gujarati language. All pioneering machine translation systems for the Gujarati language including Google Translate and Microsoft Translator face the problem of idiom translation. The proposed system successfully identifies the Gujarati idioms available in the input text. The idiom identification method of the proposed model makes it easy for any machine translation system to deal with the Gujarati idiom. Additionally proposed system simplifies the idiom in terms of providing Gujarati meaning of that idiom. We believe that the output provided by the proposed system is the text without any Gujarati idiom will be further useful for the translation of Gujarati idiomatic text to any other language in the world.

References

[1] Wikipedia, "Gujarati language", https://en.wikipedia.org/wiki/Gujarati_language (accessed January 24, 2022).

[2] Wikipedia, "Gujarati grammar", https://en.wikipedia.org/wiki/Gujarati_grammar (accessed January 24, 2022).

[3] RitiRiwaz, "Gujarati Language Gujarati History and Facts", https://www.ritiriwaz.com/gujarati-language-gujarati-history-and-facts/ (accessed January 24, 2022).

[4] Wikipedia, "Stemming", https://en.wikipedia.org/wiki/Stemming (accessed January 24, 2022).

[5] GeeksforGeeks, "Introduction to Stemming", https://www.geeksforgeeks.org/introduction-to-stemming/ (accessed January 24, 2022).

[6] Bitext, "What is the difference between stemming and lemmatization?", https://blog.bitext.com/what-is-the-difference-between-stemming-and-lemmatization/ (accessed January 24, 2022).

[7] Bing Microsoft Translator, "Microsoft Bing", Microsoft Corporation Ltd.; Available Online: https://www.bing.com/translator (accessed January 24, 2022).

[8] Wikipedia, "Microsoft Translator", Available Online: https://en.wikipedia.org/wiki/Microsoft_Translator (accessed January 24, 2022).

[9] Google Translate, "Google Translate", Google Corporation Ltd.; Available Online: https://translate.google.co.in/ (accessed January 24, 2022).

[10] Wikipedia, "Google Translate", Available Online: https://en.wikipedia.org/wiki/Google_Translate (accessed January 24, 2022).

[11] Saini J. R. and Modh J. C., "GIdTra: A dictionary-based MTS for translating Gujarati bigram idioms to English," 2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC), 2016, pp. 192-196, doi: 10.1109/PDGC.2016.7913143.

[12] Modh J. C. and Saini J. R., 2018, "A Study of Machine Translation Approaches for Gujarati Language", International Journal of Advanced Research in Computer Science, Volume 9, No. 1, January-February 2018, pages 285-288; Available online: ijarcs.info/index.php/Ijarcs/article/download/5266/4497.

[13] Modh J. C. and Saini J. R., "Context Based MTS for Translating Gujarati Trigram and Bigram Idioms to English," 2020 International Conference for Emerging Technology (INCET), 2020, pp. 1-6, doi: 10.1109/INCET49848.2020.9154112.

[14] Modh J. C. and Saini J. R., "Using IndoWordNet for Contextually Improved Machine Translation of Gujarati Idioms", International Journal of Advanced Computer Science and Applications(IJACSA), 12(1), 2021. http://dx.doi.org/10.14569/IJACSA.2021.0120128.

[15] Modh J. C. and Saini J. R., "Dynamic Phrase Generation for Detection of Idioms of Gujarati Language using Diacritics and Suffix-based Rules", International Journal of Advanced Computer Science and Applications(IJACSA), 12(7), 2021. http://dx.doi.org/10.14569/IJACSA . 2021.0120728.

[16] Grace Muzny and Luke Zettlemoyer, "Automatic Idiom Identification in Wiktionary", The Stanford NLP Group; Available online: https://nlp.stanford.edu/~muzny/docs/mz-emnlp2013.pdf.

[17] Rakesh Verma and Vasanthi Vuppuluri, "A New Approach for Idiom Identification Using Meanings and the Web", Available online: http://www2.cs.uh.edu/~rmverma/ranlp.pdf.

[18] Ferdy Hubers, Catia Cucchiarini, Nicoline van der Sijs, "Knowledge of idiomatic expressions in the native language: Do emigrants lose their touch?", ScienceDirect, 2022, Available online: https://www.sciencedirect.com/science/article/pii/S0024384122000031.

[19] Ruth Kessler & Claudia K. Friedrich, "Delayed prediction of idiom constituent meaning points to weak holistic multi-word representation in children, Language, Cognition and Neuroscience", 2022, DOI: 10.1080/23273798.2022.2035781.

[20] Bakir and Ricky Umbu, "An Analyze Typology of Idiomatic Expression in American Pie Movie", 2022, Available online: https://eprints.umm.ac.id/83454/.

[21] Muhammad Azimi Ramadhan, Supiani, Angga Taufan Dayu, "The Adoption of English Movies in Learning English Idioms: The English Undergraduate Students' Perception of English Movies with English Subtitles", 2022, Available online: http://eprints.uniska-bjm.ac.id/9595.

[22] GujaratiLexicon, Gujaratilexicon.com, Available online: http://www.letslearngujarati.com/about-us (accessed January 24, 2022).

[23] Rudhiprayog ane kahevatsangrah, published by Director of Languages, Gujarat State, Gandhinagar. 2010.

Towards Employee Perceived Satisfaction in using Citrix Workspace Technology

Bandar Ali Al Fehaidi, Muhammad Asif Khan
Department of Information Systems, Taibah University, Madina al Munawwara, Saudi Arabia

***Abstract*—In view of increasing use of technologies and their impacts in businesses, companies are now using technology to accelerate business processes. However, some companies are still failing to achieve the desired promising results by using technology. Technology in any company succeeds only when the people working in the company accept the technology wholeheartedly and show diligence in learning and using it. In this research we have evaluated the perceptions about a technology called Citrix Workspace in a Saudi company to see whether workers are satisfied with this technology and how they find its usefulness at work. A theoretical framework Technology Acceptance Model (TAM) has been used in this research to study different variables. The data collected by employees within the organization has been analyzed in order to determine level of satisfaction with the technology and to know factors that keep the workers away from remote use of the technology. These results also help managers decide how Citrix Workspace technology or other such technologies can be used in remote sites of the company**.

Keywords—Citrix workspace; employee satisfaction; secure remote access; mobile worker; TAM

I. Introduction

These days working remotely has become an accepted practice in business organizations due to its ease, effectiveness and emerging technologies. Therefore, remote working and its applications need to be deeply understood from employees' perspectives. A Saudi company has been using Citrix Workspace platform to enable its employees to work anywhere and anytime. Citrix Workspace offers users a secure, single sign-on (SSO) channel to access all of the applications and content of an organisation's network so that it can be run across various appliances including laptops, smartphones and tablets, thereby enabling employees to work anywhere. It brings all employees applications, desktops and files into one place anywhere, anytime. Employees have access to everything they need in one location. This includes access to applications, desktop and their files with full desktop experience. Applications in Citrix Workspace can be collated in customized categories which make quick access to required application. Citrix Workspace integrates applications whether they are running in clouds or they are hosted on-premises. In order to use Citrix Workspace, employees use two factors authentication to access more than one-hundred applications and information systems to do their jobs anywhere. A reliable communication to the information systems, appliances and content workers has been founded to increase worker output and improve business performance. Technology has changed the way we interconnect with the world around us. It has modified the way of learning, shopping, socializing and working. In today's workforce, technology has changed the way companies interact with their employees and make online business interactions [1-2]. In recent Corona-19 pandemic situation organizations allowed employees for remote working using various applications. The management of the Saudi company decided to allow employees of the company to work remotely using Citrix Workspace platform following the regular office hours. A part of the vision of the Saudi company in information technology and digital transformation is to deliver best practice, transparent and secure information technology and communication services to customer. Because Citrix Workspace is a new remote work system, the matter of estimating employees' satisfaction and performance with and perceptions of Citrix Workspace has received little attention. This motivated us to assess the level of satisfaction and performance of the employees using Citrix Workspace as a remote platform. This research will attempt to fill this gap by addressing variables that influence employees' satisfaction and performance with and perceptions of Citrix Workspace. Researchers have been evaluating performance of different technologies to determine effectiveness [3] that also motivated us to conduct this study.

Since the Citrix Workspace has been introduced recently, a little investigation has been conducted to determine the employees' satisfaction and performance in terms of ease of use and effectiveness. In order to conduct our research following research question has been devised to which answer will be sought:

"What are perceptions of employees' satisfaction and performance about Citrix Workspace?"

In order to answer the question, Technology Acceptance Model (TAM) is the best model to measure employees' acceptance and satisfaction with Citrix Workspace in the Saudi company. The model is one of the most predominate and common models of research. It concludes that perceived usefulness and perceived ease of use are central to determining an individual's attitude to using the technology and, thus, eventually relating to actual use. Researchers and companies have consequently been attempting to realise variables that positively impact an individual's acceptance of information technology, thereby finally supporting its use. In the information technology field, TAM is one of the better-known models for clarifying why people use technology.

TAM is a model used in information to asses individual's attitude in adoption of technology in working environment. It concludes the perceived usefulness and perceived ease of use are central to this assessment. Researchers and companies have consequently been attempting to realize variables that positively impact an individual's acceptance of information technology, thereby finally supporting its use:

Based on the TAM, the following hypotheses have been developed:

H1: Perceived ease of use has a significant positive influence on perceived usefulness.

H2: Perceived ease of use has a significant positive influence on actual usage.

H3: Actual usage has a significant positive influence on user satisfaction.

H4: User satisfaction has a significant positive influence on performance.

This study aims to determine employees' satisfaction and performance with using Citrix Workspace and to address any factors that may lead to dissatisfaction.

II. Literature Review

Remote working and work from home policies were promoted during 1970s in response to the hike in oil prices, but recent Covid-19 pandemic has given executives a serious thought to shift major section of workforce remotely [4]. Although Covid-19 has forced people to work remotely, but the success of this practice shows many will continue to work remotely on permanent basis and number of remote workers will rise in future. Organisations must take great care when approaching, designing, implementing, and fine-tuning their remote work solutions. The current context of the remote work revolution not only transforms any company into a possible case study but also provides an opportunity to investigate remote work trends at a worldwide level. This could possibly give all of the necessary evidence to justify adjusting remote work initiatives to different nations with different backgrounds [5]. This provides the basis for choosing the Saudi company as a case study to explore new remote work initiatives by assessing employees' satisfaction and performance of using Citrix Workspace as a remote work solution.

With regards to remote work solutions, for so many years organisations have relied on virtual private networks (VPNs) to allow remote workers access to their networks. However, when actions to contain the coronavirus led many businesses to shift from traditional office workspace to remote working, many more employees became remote workers. When this occurred, VPN technology proved insufficient to cope with the increasing need for a larger number of employees to work remotely. Consequently, Some IT departments shifted from VPN solutions to Citrix Workspace using on-premises servers without sacrificing security. With Citrix Workspace, system administrators are able to tailor and distribute any number of workspaces effectively and easily, each with its required applications, appliances, data, and networking services to be aligned with an organisation's needs for different uses [6]. Citrix system is widely used as a mean for remote working which has strong position in the market of remote applications and desktop virtualization [7]. Therefore, user satisfaction and performance with Citrix Workspace need to be measured in the current study. In a study [8] it is reported that employees' productivity increased by 13% due to work at home and fewer breaks taken and better performance per minute. Furthermore, remote workers also reported extraordinarily high work satisfaction and their job exhaustion rate was approximately 50% decreased. Nevertheless, the study did not mention either technology used or relationship between technology and employees' satisfaction and performance which is core focus in the current study. In a separate study it is reported that in addition to saving money on traditional office workspace, employee satisfaction rankings were approximately 90 percent due to working remotely [9]. In another study it is reported that flexible work schedule has a positive impact on the life satisfaction of employees [10]. There is a positive relationship between flexible work schedules and worker satisfaction. A research study has addressed the impact of psychological happiness on employee satisfaction which can be linked to increased productivity resulting from flexible work schedules [11]. The research conducted by [10-11] has analysed the impact of flexible work schedules on worker satisfaction without determining which remote work solution was used as a variable which could affect employees' satisfaction.

A. Perceptions of Ease of Use

There is a considerable body of empirical research indicating that perceptions of ease of use significantly influence the information system context [12]. Perceptions of ease of use refer to the extent to which an individual anticipates being able to utilise a system with relatively little effort or struggle [13]. It is widely suggested in the empirical literature relating to IS that the greater is a system's ease of use, the greater will be the perceptions of how useful that system is [14]. Several studies in the empirical literature have established a positive relationship between perceptions of ease of use and perceptions of how useful a system is [15].

B. Perceived Usefulness

It is defined as "the degree to which a person believes that using a particular system would enhance his or her job performance" [13]. It is the perception of technology users which they build based on the results of their experience. In view of perceived usefulness, frequency of actual use of technology is suggested and its positive effect on individual's performance studied [16].

C. User Satisfaction

When considering the context of information systems, it is necessary to establish whether use of the system results in the satisfaction of users [17]. In the current study, the satisfaction of users is gauged in terms of the extent to which users of the Citrix Workspace are satisfied with their decision to use the technology and whether it lived up to their expectations.

D. Performance Impact

There are many reports in the empirical literature that higher levels of user satisfaction are associated with superior performance impact [16, 18-19]. Each of these investigations add to the existing body of knowledge and help to better understand the nature of the relationship between actual use and how individuals perform in the context of organizations.

Staff performance is influenced by the quality of the system and user satisfaction. If workers are satisfied with a computer system, it will add value and contribute to their performance and productivity [20]. The study also stated the impact of technology attributes on user satisfaction. The relationship between user satisfaction, performance and Citrix Workspace will be further investigated in the proposed research, as mentioned in hypothesis.

III. Method and Data Preparation

The study was devised as a quantitative correlational research design to examine the relationship between employee satisfaction and performance using Citrix Workspace to work remotely within the Saudi company. A quantitative method was deemed most suitable in this research because quantitative researchers concentrate on exploring measurable phenomena using statistical techniques [21]. Correlational designs are suitable when the objective of the study is to examine the relationship between variables of interest. The TAM is used to measure the employees' satisfaction and performance when using Citrix Workspace. An online survey questionnaire was designed using Google Forms to collect data from the respondents and that data is analyzed to draw out results and findings. Questionnaire is a best tool to determine user intention, expectation and perception [22-23].

Based on the four hypotheses stated above, relationships between five variables will be examined. These variables are perceptions of ease of use, perceptions of usefulness, actual use, the satisfaction of users, and performance.

A. *Survey Instrument and Data Collection*

The main instrument to conduct the study is an online self-administered questionnaire which was developed for this study. The population of the study is users of Citrix Workspace from all departments of the Saudi company. The purpose of the survey instrument was to determine employees' level of satisfaction in using Citrix Workspace technology remotely. The questionnaire was distributed among all work areas of the Saudi company (Central, Western, Eastern and Southern region) and 191 responses were received. 5 of the responses were rejected because the respondents did not use Citrix Workspace at all. Consequently, the sample of the research was 186 responses.

The questionnaire consisted of two main sections with 39 items/statements. It included various types of questions such as the following: (1) rating scale questions utilizing a Likert-type 5-point scale (i.e. 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, and 5 = strongly agree); (2) multiple choice single answer questions; and (3) an open question. The questionnaire was devised using Google Forms in the English and Arabic languages, as the workers mainly are native Arabic speakers. Google Forms is a website allowing users to create and disseminate surveys for business or individual purposes such as this research. Google Forms allows the survey creator to use a ready-made survey or create a customized one. The data results of surveys are collected and exported to Microsoft Excel so that they are ready to be used on SPSS software.

Participation in the research study was non-compulsory and participants were informed of the volitional nature and confidentiality measures at the beginning of the survey. Moreover, the data were collected anonymously in the survey and we analysed and interpreted it during the research period, before being deleted at the end of the research.

IV. Result and Discussion

In this study descriptive analytical method has been used to describe characteristics of the study respondents and their opinions around the components of the study. The hypotheses have been tested by means of Pearson correlation to explain the relationship between research variables at a significant level of 0.05. The internal consistency and reliability of the questionnaire were tested using SPSS version 23. After verifying the validity and reliability of the questionnaire test, a link of online questionnaire was distributed to the study population. Total of (191) questionnaires were retrieved, 5 participants were excluded from the study sample because they did not use Citrix workspace at all and the final study sample was 186.

Validity and reliability of the research questionnaire is an essential in the data collection phase, where these two quantifications ensure the procedure of getting precise and consistent data and results. Validity demonstrates whether the instrument or test measures what it is assumed to measure. It also refers to the ability of a study to measure what was set out to be measured or the validity of the resulting data. The measurement that needs to be valid is taken using instruments. In the case of the currents study, the instrument was taken from empirical literatures and then modified according to the stated research objectives.

Before dissemination of the final version of the questionnaire for the research, a pilot test was conducted using an online questionnaire on Google Forms. The pilot test was conducted via Google Forms by asking experts users of Citrix Workspace in the Saudi company to complete the questionnaire using a link that directed them to Google Forms to complete the survey. For the purpose of the pilot test, 15 surveys were completed. During the pilot test, the participants were given an optional comment area on each question to provide feedback regarding the clarity of the question or any spelling or linguistic errors. Therefore, any modifications required to the expression of the questions could be made prior to the data collection process for the current research.

In order to evaluate reliability of the survey instrument, Cronbach's alpha test of reliability and internal consistency was done on all questionnaire variables and items. Cronbach's alpha test of reliability and internal consistency was done on the four scales. The alpha values were explained through the guidance determined by George and Mallery (2016), in which $\alpha > .9$ Excellent, $\alpha > .8$ Good, $\alpha > .7$ Acceptable, $\alpha > .6$ Questionable, $\alpha > .5$ Poor, and $\alpha < .5$ Unacceptable. So, for reliability, this research investigated Cronbach's alpha coefficients to gauge the reliability of each of the five variables in the questionnaire. Those variables are ease of use, usefulness, actual usage, user satisfaction, and performance. We calculated the internal consistency of the questionnaire by calculating Pearson correlations between items-items in each variable. Pearson correlation coefficients measure the strength and direction of a linear relationship between two variables. Table I shows the Pearson correlation coefficients between items of each variable "ease of use, usefulness, user satisfaction, and performance impact" and the total score of the field. The Pearson correlation coefficients were ranged from 0.344 to 0.772 for Ease of use variable, from 0.872 to 0.907 for Usefulness variable, from 0.747 to 0.934 for User satisfaction variable, and from 0.780 to 0.867 for Performance impact variable, indicating that all the correlation coefficients shown are significant at ($\alpha \leqslant 0.01$), thus the fields are considered valid of what is being measured. The correlation is significant at 0.01 level.

In order to measure the stability of questionnaire, Alpha Cronbach method and split-half coefficients were used. The split-half reliability method requires dividing up the variables' items into two parts and calculating correlations between the first half and the second half of each variable's items using Spearman’s correlation test. It is clear from the results shown in Table II that the value of the Alpha Cronbach coefficient is high for each field, ranging from (0.739 to 0.960) which is greater than the recommended level of 0.7. As well as Spearman’s correlation coefficient which it was ranging (0.740 to 0.948), indicating very strong reliability of these items where their correlations were statistically significant at 0.01.

A total of 186 participants completed the questionnaires. Out of them, 172 (92.5%) were males and 14 (7.5%) were females. Most of the participants were from western region (49.5%). The age groups of the participants were as follows: 20 to 29 years (12.4%), 30 to 39 years (42.5%), 40 to 49 years (38.1%) and 50 years or above (7%). Most of them married and bachelor degree holders (67.7%). A large number of participants belonged to distribution and customer services (43.5%) followed by IT and digital transformation (25.8%). Table III depicts the statistical analysis of the variables under study that include mean, standard deviation and weighted percentage for the variables.

Participants’ responses toward the items for Perceived ease of use variable, Perceived usefulness variable, User satisfaction variable and Performance impact variable have been represented graphically in Fig. 1, Fig. 2, Fig. 3 and Fig. 4a and 4b, respectively.

TABLE I. PEARSON CORRELATION COFFICIENTS BETWEEN ITEM AND VARIABLE

Ease of use		**Usefulness**		**Use satisfaction**		**Performance impact**	
Item	Correlation	Item	Correlation	Item	Correlation	Item	Correlation
1	0.772	1	0.879	1	0.747	1	0.803
2	0.344	2	0.900	2	0.82	2	0.867
3	0.782	3	0.905	3	0.934	3	0.780
4	0.709	4	0.892	4	0.880	4	0.838
5	0.757	5	0.907			5	0.806
6	0.761	6	0.872			6	0.86
						7	0.832
						8	0.815
						9	0.860
						10	0.858

TABLE II. ALPHA’S CRONBACH AND SPLIT-HALF COEFFICIENTS

Variable	**Cronbach’s Alpha**	**Split-half reliability coefficients**		
		Correlation coefficient	Length corrections	
			Spearman coefficient	Guttman coefficient
Perceived ease of use	0.739	0.587	0.740	0.733
Perceived usefulness	0.947	0.851	0.920	0.919
User satisfaction	0.863	.0761	0.864	0.834
Performance impact	0.960	0.901	0.948	0.947

TABLE III. STATISTICAL ANALYSIS OF THE VARIABLES

Variable	**Item**	**Responses**					**Mean**	**SD**	**%**
		Strongly agree	Agree	Neutral	Agree	Strongly disagree			
Perceived ease of use	1	87	84	11	3	1	4.36	0.72	87.2
	2	30	57	50	35	14	3.29	1.17	65.8
	3	5	88	23	9	1	4.11	0.84	82.2
	4	74	84	22	6	0	4.22	0.78	84.4
	5	45	68	54	16	3	3.73	0.98	74.6
	6	98	62	11	14	1	4.30	0.92	86.0
	Mean = 4.0 SD = 0.60 80								
Perceived usefulness	1	98	64	9	12	3	4.30	0.94	86.0

Variable	Item	Responses					Mean	SD	%
	2	106	63	9	5	3	4.42	4.42	88.4
	3	77	75	22	7	5	4.14	0.95	82.8
	4	85	73	13	12	3	4.21	0.94	84.2
	5	69	74	27	12	4	4.03	0.99	80.6
	6	75	79	21	5	6	4.14	0.95	82.8
	Mean = 4.21 SD = 0.83 84.2								
User satisfaction	1	113	63	8	2	0	4.54	0.63	90.8
	2	75	87	17	6	1	4.23	0.79	84.6
	3	85	82	9	8	2	4.29	0.83	85.8
	4	63	76	29	12	6	3.96	1.02	79.2
	Mean = 4.26 SD = 0.70 85.2								
Performance impact	1	69	82	24	9	2	4.11	0.88	82.2
	2	66	87	22	7	4	4.10	0.90	82
	3	48	73	46	15	4	3.78	0.99	75.6
	4	42	86	43	12	3	3.82	0.91	76.4
	5	34	69	53	26	4	3.55	1.01	71
	6	67	90	18	6	5	4.12	0.90	82.4
	7	33	71	61	16	5	3.60	0.97	72
	8	40	71	57	12	6	3.68	0.99	73.6
	9	68	77	27	10	4	4.05	0.96	81
	10	78	73	24	7	4	4.15	0.94	83
	11	64	96	13	9	4	4.11	0.89	82.2
	12	76	73	24	8	5	4.11	0.97	82.2
	Mean = 3.93 SD = 0.79 78.6								

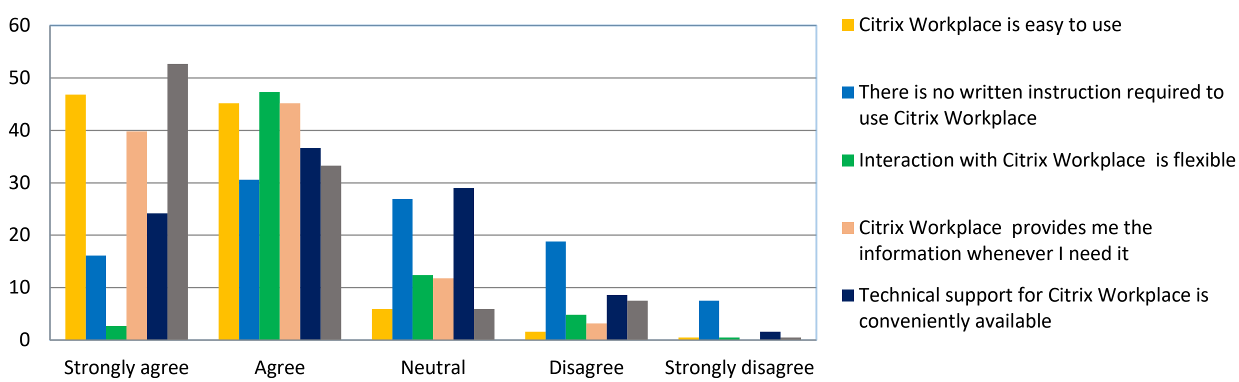


Fig. 1. Participants Responses for the Items of Perceived ease of use Variable.

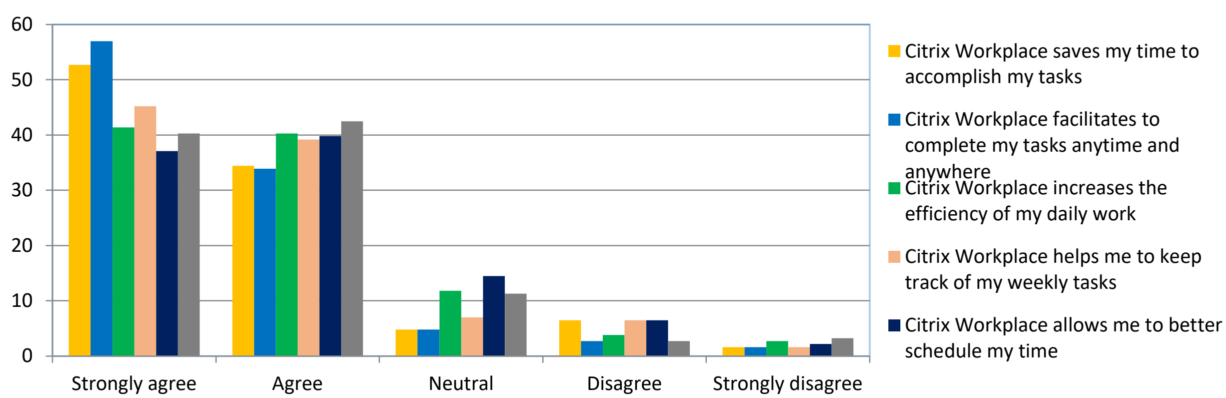


Fig. 2. Participants Responses for the Items of Perceived usefulness Variable.

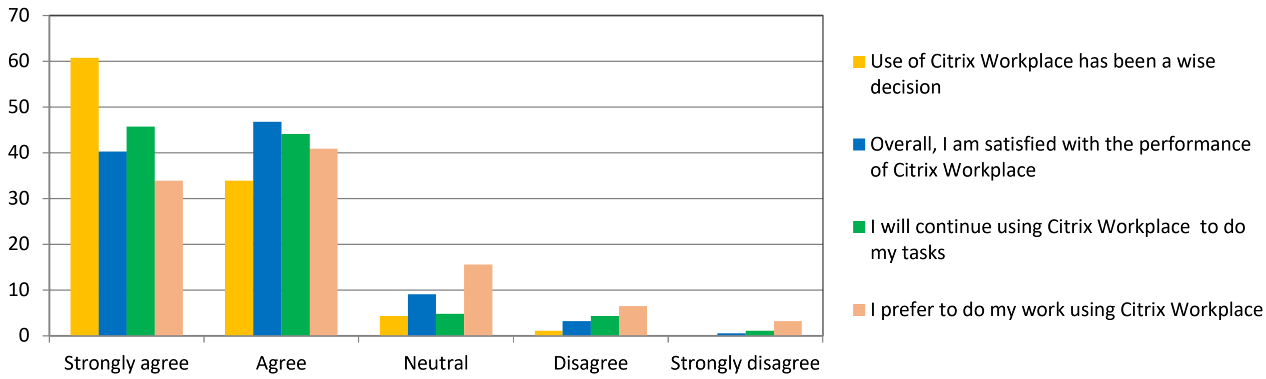
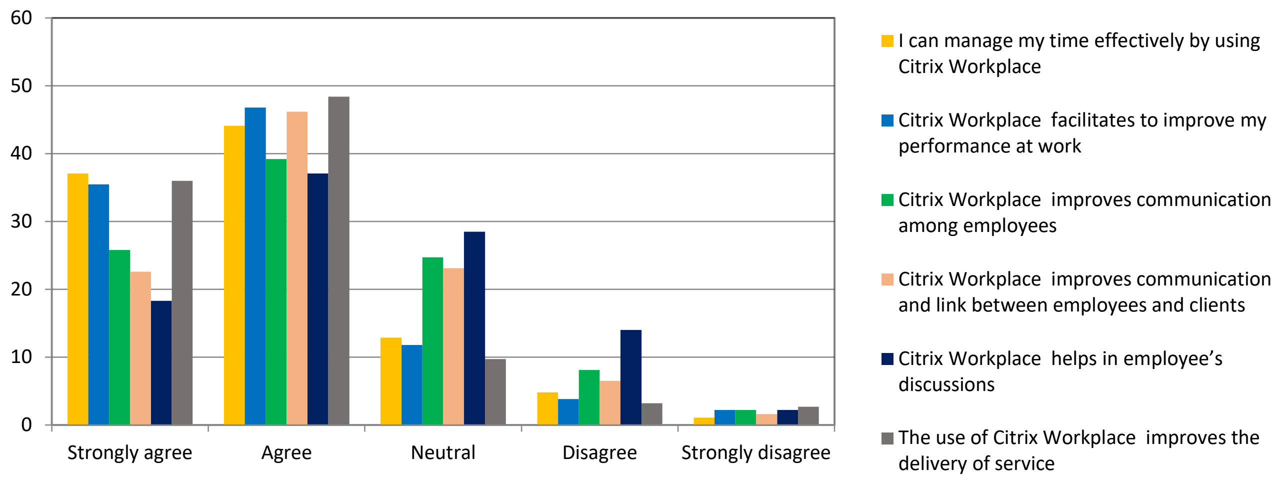
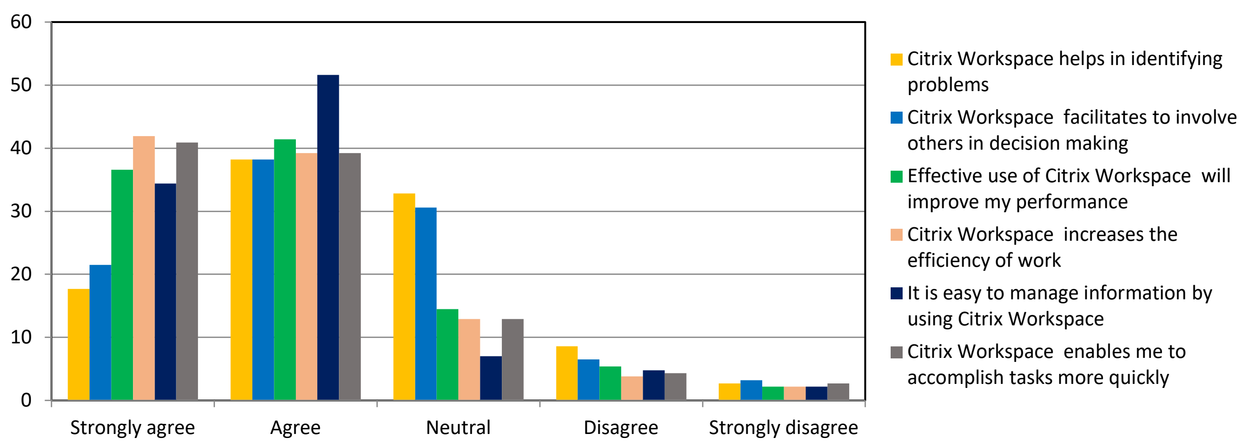


Fig. 3. Participants Responses for the Items of user Satisfaction Variable.



(a)



(b)

Fig. 4. (a) Participants Responses for First 6 Items of Performance Impact Variable. (b) Participants Responses for Rest of Items of Performance Impact Variable.

B. Test of Hypotheses

We postulated four hypotheses for our study in the beginning of this research. Now following are the tests for each of the hypothesis.

H1: Perceived ease of use has a significant positive influence on perceived usefulness.

This study found that there was statistically significant and strong positive relationship between perceived ease of use and perceived usefulness which is compatible with results of the study [24].

The Pearson correlation coefficient between perceived ease of use and perceived usefulness was a strong positive correlation, the value of the correlation coefficient was 0.669.

The correlation was statistically significant at the level of significance (α ⩽0.01).

H2: The Pearson's correlation between Perceived ease of use and actual usage.

The study revealed statistically significant positive relationship between perceived ease of use and actual usage at a statistical significance level. This is in common with other studies which found a significant relationship between ease of use and actual usage [25,26].

The Pearson correlation coefficient between perceived ease of use and actual usage was a week positive correlation, the value of the correlation coefficient was 0.264. The correlation was statistically significant at the level of significance (α ⩽ 0.01.

H3: Actual usage has a significant positive influence on User satisfaction.

This states that actual usage has a significant positive influence on user satisfaction which is consistent with the findings of previous studies [27-28]; another statistically significant relationship was found between actual usage and user satisfaction.

The Pearson correlation coefficient between actual usage and user satisfaction was a moderate positive correlation, the value of the correlation coefficient was 0.394. The correlation was statistically significant at the level of significance (α≤0.01).

H4: User satisfaction has a significant positive influence on Performance.

Higher satisfaction implies higher performance as the same finding was found in parallel studies indicate that the higher the user satisfaction, the higher performance impact [27, 29].

The Pearson correlation coefficient between user satisfaction and performance was a very strong positive correlation, the value of the correlation coefficient was 0.838. The correlation was statistically significant at the level of significance (α≤0.01).

V. Conclusion and Future Work

This study has helped to understand if a relationship existed between an employee's satisfaction and performance with using Citrix Workspace to work remotely. The remote work has become more important with advent of Covid19 and employees have been given access of using technologies remotely [30].

The results revealed that there is a high level of satisfaction among respondents regarding the decision to use the Citrix Workspace. The main motive for most of the participants to use Citrix Workplace was their ability to work remotely. There is statistically significant positive relationship between perceived ease of use and perceived usefulness. Also this study revealed statistically significant positive relationship between perceived ease of use and actual usage. Another statistically significant relationship was found between actual usage and user satisfaction the relation was found to be positive. This study also revealed statistically significant relationship between user satisfaction and positive performance impact.

For future research, findings of this study may lead to similar studies being conducted to measure users' satisfactions and performance with other remote work platforms in other Saudi organisations. Furthermore, they can be used as a road map to apply TAM to conduct investigations and research to assess employees' satisfaction and performance with using other information systems in Saudi Arabia and other places.

References

[1] A. Felstead and G. Henseke, "Assessing the growth of remote working and its consequences for effort, well-being and work-life balance.", *New Technology Work and Employment* , vol. 32, no. 1, pp. 1-18, 2017.

[2] J. Kilgore, "HR technology in the new remote work environment." *Workforce Solutions Review*, pp.11–14, 2020.

[3] M. Hyder, S. Tooba, and Waseemullah, "Performance Evaluation of RSA-based Secure Cloud Storage Protocol using OpenStack", *Engineering, Technology and Applied Science Research*, Vol 11, No. 4, pp. 7321-7325, 2021.

[4] P. Choudhury, "Our Work-from-Anywhere Future.", *Harvard Business Review, vol. 98, no. 6*, pp. 58–67, 2020.

[5] P. Poonam, K. Krishan, S. Bharanidharan, C. Kheng, "A theoretical review of social media usage by cybercriminals, "in Conference on Computer Communication and Informatics, 2017, India.

[6] C. Saran,."Supporting Mobile Workers Everywhere." *Computer Weekly*, pp. 17–22, 2020.

[7] M. Dhall, and Q. Tan, "A Profitable Hybrid Desktop as a Service Solution." in IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), 2019, pp. 55-62.

[8] N. Bloom, J. Liang, J. Roberts, and Z. Ying, "Does Working from Home Work? Evidence from a Chinese Experiment*. *Quarterly Journal of Economics*, vol. 130, no. 1, pp. 165–218, 2015.

[9] B. Wng, Y. Liu, J. Qian, and S. Parker, "Achieving Effective Remote Working During the COVID-19 Pandemic: A Work Design Perspective.", *Applied Psychology*, vol. 70, no. 1, 2021.

[10] S. Shagvaliyeva, and R.Yazdanifard, "Impact of Flexible Working Hours on Work-Life Balance." *American Journal of Industrial and Business Management*, 4(1), pp. 20-23, 2014.

[11] S. Wadhawan, "Impact of Flexible Working Arrangements on Employee Satisfaction in IT Sector." *BVIMSR Journal of Management Research*, vol. 11, no. 1, pp. 59, 2019.

[12] K. Faqih, "An empirical analysis of factors predicting the behavioral intention to adopt Internet shopping technology among non-shoppers in a developing country context: Does gender matter?." *Journal of Retailing and Consumer Services*, vol. 30, pp. 140-164, 2016.

[13] F. Davis, "Perceived usefulness perceived ease of use and user acceptance of information technology.", *MIS. Quarterly*., vol. 13, pp. 319-340, 1989.

[14] N. Elkhani, S. Soltani, and A. Nazir, "The effects of transformational leadership and ERP system self-efficacy on ERP system usage." *Journal of Enterprise Information Management*., vol. 27, pp.759-785, 2014.

[15] V. Bhatiasevi, and C. Yoopetch, "The determinants of intention to use electronic booking among young users in Thailand." *Journal of Management Information Systems*, vol. 23, pp. 1-11, 2015.

[16] O. Isaac, Z. Abdullah, T. Ramayah, A. Mutahar and I. Alrajawy, "Perceived usefulness, perceived ease of use, perceived compatibility and Net benefits: An empirical study of internet usage among employees in Yemen." in 7th International Conference on Postgraduate Education (ICPE7), Sep 8, 2016, Dewan Sri Budiman, Shah Alam Malaysia, pp. 899-919.

[17] A. Bag, and Y. Lin, "Customers' Satisfaction on Technology Adoption: A Study on the Smartphones Usagee." *Intenrational Journal of Scientific and Engineering Research*., vol. 9, pp. 31-39, 2018.

[18] M. Khan, "Investigation of Critical Factors that Perturb Business-IT Alignment in Organizations", *International Journal of Advanced Computer Science and Applications*, vol. 8, pp. 133-137, 2017.

[19] H. Son, Y. Park, C. Kim, and J. Chou, "Toward an understanding of construction professional's acceptance of mobile computing devices in South Korea: An extension of the technology acceptance model." Automation in Construction., vol. 28, pp. 82-90, 2012.

[20] A. Ameen, D. Al-Ali, O. Isaac, and F.Mohammed, "Examining relationship between service quality, user satisfaction and performance impact in the context of smart government in UAE." *International Journal of Electrical & Computer Engineering*, vol. 10, no. 6, pp. 6026–6033, 2020.

[21] G. Goertz, and J. Mahoney, "A tale of two cultures: Qualitative and quantitative research in the social sciences." NJ: Princeton University Press, 2012.

[22] H. Chatti and S. Hadoussa, "Factors Affecting the Adoption of E-Learning Technology by Students during the COVID-19 Quarantine Period: The Application of the UTAUT Model", *Engineering, Technology and Appled Science Research*, vol. 11, no. 2, pp. 6993–7000, 2021.

[23] Z. U. Rehman and F. A. Shaikh, "Critical Factors Influencing the Behavioral Intention of Consumers towards Mobile Banking in Malaysia", *Engineering, Technology and Appled Science Research*, vol. 10, no. 1, pp. 5265–5269, 2020.

[24] H. Widiastuti, and S. Hayati, "User's Perceived Usefulness, Quality and Ease of Use of Internet Financial Reporting (IFR) on Their Intention to Use IFR for Investment Decision Making" in 5th International Conference on Accounting and Finance (ICAF-2019), November, 2019, Indonesia.

[25] N. Samuel, S. Onasanya, C. Olumorin, "Perceived usefulness, ease of use and adequacy of use of mobile technologies by Nigerian university lecturers" *International Journal of Education and Development using Information and Communication Technology*, vol. 14, no. 3, pp. 5-16, 2018.

[26] H. Kim,, H. Chan and S. Gupta, "Value-based adoption of mobile internet: An empirical investigation." *Decision. Support System.*, vol. 43, pp. 111-126, 2018.

[27] M, Uddin, O. Issac, I. Alrajawy, M. Maram, "Do User Satisfaction and Actual Usage of Online Learning Impact Students Performance?" *International Journal of Management and Human Science*, vol. 3, no. 2, pp. 60-67, 2019.

[28] S. Al Kati, and M. Khan, "Challenges and Issues of E-Learning Using Education Cloud – A Review in Context of Covid-19 Pandemic", *International Journal of Computer Applications Technology and Research*, vol. 10, pp. 106-110, 2021.

[29] M. Makokha, and D. Ochieng, "Assessing the success of ICT's from a user perspective: Case study of coffee research foundation," *Kenya. Journal of. Management Strategy*, vol. 5, pp. 46-54, 2014.

[30] K. Ninaus, S. Diehl, R. Terlutter, "Employee perceptions of information and communication technologies in work life, perceived burnout, job satisfaction and the role of work-family balance", *Journal of Business Research*, vol. 136, pp. 652-666, 2021.

Bearing Fault Detection based on Internet of Things using Convolutional Neural Network

Sovon Chakraborty[1], F. M. Javed Mehedi Shamrat[2], Rasel Ahammad[3]
Md. Masum Billah[4], Moumita Kabir[5], Md Rabbani Hosen[6]
Department of Computer Science and Engineering
Ahsanullah University of Science and Technology, Dhaka, Bangladesh[1]
Department of Software Engineering, Daffodil International University, Dhaka, Bangladesh[2, 3, 4]
United International University, Dhaka, Bangladesh[5]
Department of Computer Science and Engineering, European University of Bangladesh, Dhaka, Bangladesh[6]

***Abstract*—In the age of the industrial revolution, industry and machinery are elements of the utmost importance to the development of human civilization. As industries are dependent on their machines, regular maintenance of these machines is required. However, if the machine is too big for humans to look after, we need a system that will observe these giants. This paper proposes a convolutional neural network-based system that detects faults in industrial machines by diagnosing motor sounds using accelerometers sensors. The sensors collect data from the machines and augment the data into 261756 samples to train (70%) and test (30%) the models for better accuracy. The sensor data are sent to the server through the wireless sensor network and decomposed using discrete wavelet transformation (DWT). This big data is processed to detect faults. The study shows that custom CNN architectures surpass the performance of the transfer learning-based MobileNetV2 fault diagnosis model. The system could successfully detect faults with up to 99.64% accuracy and 99.83% precision with the MobileNetV2 pre-trained on the ImageNet Dataset. However, the Convolutional 1D and 2D architectures perform excellently with 100% accuracy and 100 % precision.**

Keywords*—Accuracy; convolution 1D; convolution 2D; data loss; faulty machinery; mobileNetV2; precision

I. Introduction

Industries are getting smarter day by day. Early identification of fault inside industry machinery plays a significant role in this modern era. To increase the productivity of production systems, precise production techniques are vital. Extensive research has been done to detect early faults and classify those machines and defects. Monitoring the faulty machines and taking remedial action makes for a safe environment and reduced failures [1]. Researchers have suggested different classifiers for the early diagnosis and detection of faults. Over time, machines have become complex, and for complex machines, data-driven methods have shown efficiency, whereas non-parametric methods are applied for extracting related information from the data [2, 3]. Machine learning algorithms such as SVM have significant difficulty in selecting features for identifying defects, particularly in induction motors as seen in [4]. Deep learning addresses the problem of feature selection by extracting features directly from raw data. Deep learning models are not precise enough when the amount of data is not sufficient. Generalization is also a significant drawback for deep learning algorithms. Accuracy is dependent on the proper distribution of data. Later, Transfer learning methods start getting famous in the Fault Detection and Diagnosis (FDD) field to prevent industry anomalies [5, 6]. Transfer learning methods learn from a place that is enriched in data, and later these models are applied in a place where the amount of data is an issue. This research has used Convolution 1D, Convolution 2D, and MobileNetV2 architecture to detect faulty machinery.

A sufficient amount of research has been done to find better classifiers for detecting faults in industrial machinery. An Artificial Intelligence-based fault detection system has been proposed by Lei et al. For learning features from raw signals, they present a sparse filtering and neural network [7]. Boukra et al. proposed a hybrid method based on feature reduction uses two parameters. Their proposed method is not manipulated by load conditions [8]. Machine learning algorithms are used for identifying faults in the motor drive where supervised learning plays a significant role [9]. Although supervised learning downs efficiency with unwanted data. Feature extraction from raw signals plays a significant role in detecting induction motors' faults. Wavelet analysis, time-domain analysis, frequency domain analysis, time scale frequency analysis, and time scale frequency analysis all aid in the extraction of characteristics [10-12]. The features extracted here affect accuracy in classification and proper fault recognition. Deep learning reduces this problem. Deep learning algorithms show a significant result in detecting faults for bearings and gearboxes [13]. Deep CNN shows remarkable accuracy in fault diagnosis from raw vibrating signals in the anti-noise domain [14]. Deep learning models are not efficient enough when the amount of training data is not sufficient. Transfer learning models show better accuracy when the amount of data is less. Adversarial transfer learning algorithms can identify erroneous signals by first converting them to RGB pictures and then training the model on those images [15]. Transfer learning-based autoencoders are proposed by Wan et al. [16]. High fault classification results have been shown by VGG 19 in detecting faults in induction motors [17].

In the study, the authors worked with the Dataset obtained from Case Western Reserve University to detect the fault of bearings of the machinery based on a wireless IoT framework. The Dataset was further divided into sub-classes and environments. After label encoding and splitting to training and test set, the processed data were classified using CNN algorithms. The classification methods of MobileNetV2, Convolution 1D, and Convolution 2D are used. The performance of the algorithms is then measured to determine the best method to detect faults in machines in industries.

II. Literature Review

The diagnosis of bearing faults is a hot topic in mechanical condition monitoring. The key phases in bearing defect diagnostics are feature extraction and pattern classification from monitoring data. The bearings will create extra vibrations when a specific bearing element fails. The fault characteristic frequency is a connection between the frequency of the extra vibration and the bearing speed. We can locate the failing bearing by analyzing the original vibration signal's frequency components [18].

Feature extraction techniques used in classical signal processing for bearing defect diagnostics using vibration signals include the Hilbert– Huang transform (HHT), the wavelet transform, empirical mode decomposition, and approaches. HHT was used by V.K. Rai et al. [19] to extract frequency domain features from bearing fault data to identify bearing fault categories. Xinsheng Lou et al. [20] used wavelet transform and neuro-fuzzy classification to develop a novel ball bearing problem diagnostic system. The wavelet transform was employed to retrieve the accelerometer signal's feature vectors. Once the adaptive neural-fuzzy inference system had been trained to categorize the feature vectors, it was used to classify data. The suggested approach worked effectively even with a variable load.

SVMs were effectively used in the area of fault diagnostics by P. Konar et al. [21]. The feature vectors were extracted using a continuous wavelet transform (CWT), and the monitoring data for the three-phase induction motor was classified using a support vector machine (SVM). In order to recognize the early problem of the bearing, Zhuanzhe Zhao et al. [22] presented an intelligent fault detection approach based on a backpropagation (BP) neural network. It was suggested that the intrinsic mode functions (IMFs) be first acquired using a wavelet packet decomposition approach, and then the EMD method was utilized to get them. A BP neural network with three layers was built to recognize the monitoring signal fault pattern.

Using naive Bayes classifier and Bayes net classifier carried out fault diagnostics [23]. Before using the suggested technique, the vibration signals are wavelet-analyzed to extract the discrete wavelet features, then utilized as input into the Bayes net for classification.

An EMD method was utilized by Lei et al. [23], and a kurtosis-based method was offered to identify the sensitive characteristics for defect diagnostics based on bearing vibration signals. When dealing with nonlinear, non-stationary, and composite signals, Lin et al. [24] employed an enhanced EMD approach to extract features. The acoustic emission data from bearing tests were pre-processed using He et al.'s short-time Fourier transform (STFT) approach [25]. Feature selection techniques were commonly used to choose the most representative features from the collected data, including linear discriminant analysis (LDA) and principal component analysis (PCA). A diagnostic defect technique based on decision trees and PCA was reported by Sun et al. [26]. After feature extraction, PCA is used to minimize the number of features.

S. G. et al. [27] developed a continuous wavelet transform and CNN approach to accurately, robustly, and generally diagnose rotating machines faults. This study by S. S. et al. [28] demonstrated that deep convolutional neural networks (DCNNs) could learn from various sensor outputs to identify induction motor faults with consistency and accuracy. By including freshly produced extra features for self-update to incorporate new aberrant samples and fault classes, W. Y. et al. [29] used a wide convolutional neural network to increase diagnostic performance and incremental learning capabilities.

This research proposed a solution based on a wireless Internet of Things architecture for detecting machinery bearing faults using data from Case Western Reserve University. The Dataset was augmented in 261756 samples and further divided into sub-classes and environments. In order to classify the processed data, CNN methods were used after label encoding and partitioning it into training (70%) and test sets (30%). MobilNetV2, Convolution 1D, and Convolution 2D classification algorithms are used. For industrial machine fault detection, compare several algorithms' results side by side to see which one performs better.

III. Proposed Methodology

Fig. 1 depicts a ZigBee-based wireless sensor network (WSN) concept for the proposed fault diagnostic technique. The WSN was created to gather accelerometer sensor data from a variety of industrial bearings. The TMS320F28335 digital signal microcontroller is utilized as a microcontroller unit (MCU) in the WSN model, collecting data from accelerometers and transmitting it through an XBee (Pro Series 3) radio transmitter. In addition, the XBee module is used to configure a coordinator gateway that receives signals from sensor nodes and sends them to the diagnostic server, which hosts the suggested fault diagnosis model. The ZigBee network protocols are used to create a mesh network of XBee devices. Besides, fault diagnosis methods comprise the following steps: a) data collection through accelerometer sensors, b) apply DWT to decomposed signals, c) preprocess data steps including creating a set of classes, set reshaping, and label encoding, d) splitting the dataset into 30% for testing and 70% for training, e) then classify data for fault detection and evaluate the performance.

To identify defects, the suggested approach used convolutional neural network algorithms. The proposed method's process is shown in Fig. 1.

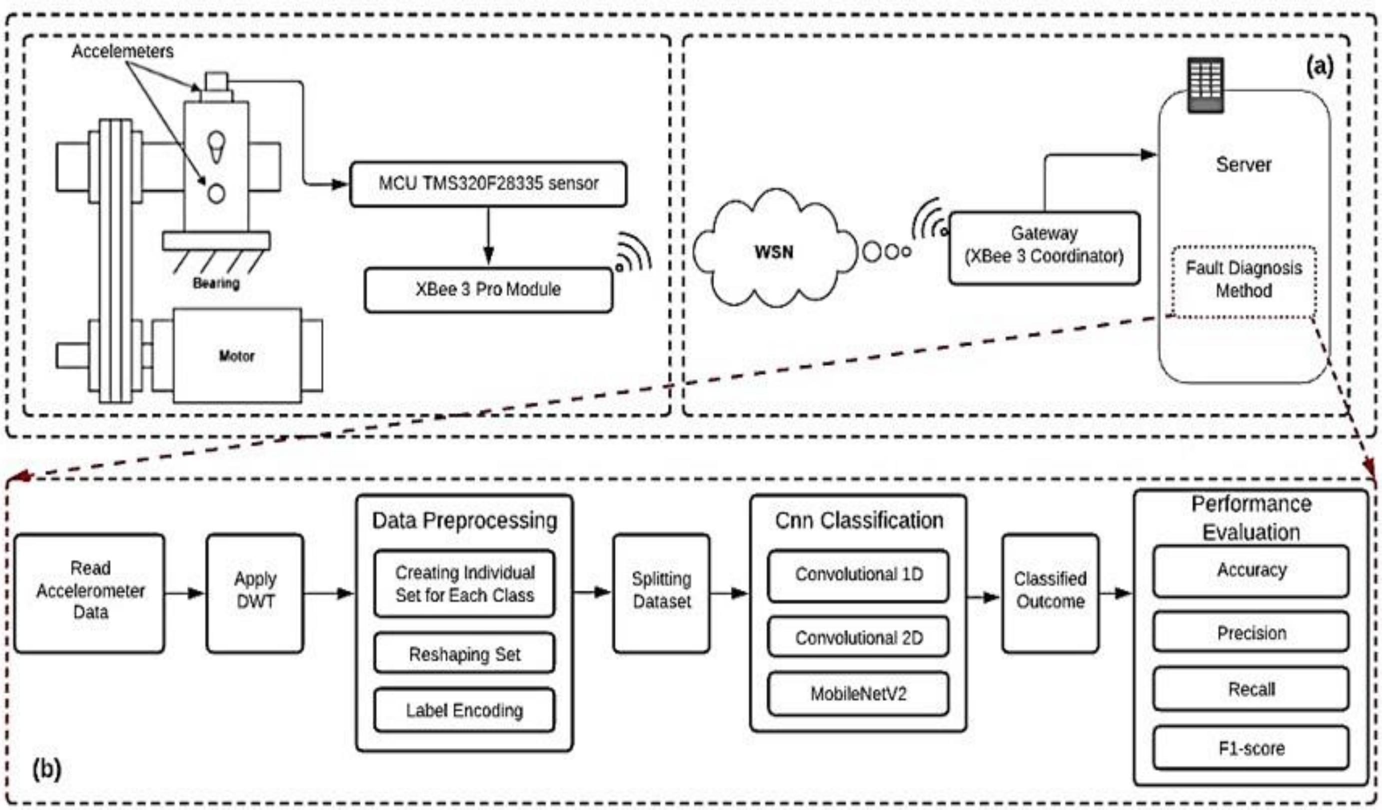


Fig. 1. Wireless Sensor Network ZigBee based and (b) Proposed Fault Diagnosis Approach.

A. Data Collection and Pre-Processing

In this study, the Case Western Reserve University dataset [30] is used as a standard guide to assessing the efficiency of the fault detection algorithm. The CWRU data center examined a 2 hp Reliance electric induction motor for regular bearings, single-point drive end (DE), and fan end (FE) faults under diverse settings. Accelerometers with magnetic bases connected to the housing were used to gather vibration data. At both the driving and fan ends of the motor casing, accelerometers were installed at midnight. An accelerometer was also connected to the motor that supported the base plate in specific tests. A 16-channel DAT recorder was used to capture vibration data, then analyzed in a MATLAB environment. MATLAB (*.mat) format is used for all data files. For drive and bearing problems, digital data was captured at 12000 samples per second and data was gathered at 48000 samples per second, the final dataset volume was around 261756 samples [31]. Data on speed and horsepower was collected using a torque transducer/encoder and manually recorded.

The dataset included three working environments with the following conditions:

1) Data collection at 12000 samples per second.
2) Motor load range 0~2.
3) Fault diameter 0.007 inches.

So, from the big data warehouse of the CWRU data center, we gathered a dataset of 16 signals, of which four were normal baseline. The rest had four inner race faults, ball and outer race faults each. These 16 significant signals were individually sliced into 5949 samples so we can later reshape them as needed for feeding into the neural networks. Hence, the total length of the Dataset is 23796 samples and augmented the final Dataset into 261756 samples. The dataset is homogeneous and balanced. A simulator of data collection from bearing based on Accelerometer's sensors is shown in Fig. 2.

The faults were artificially introduced on the SKF Drive End bearing (6205-2RS JEM) using Electro-Discharge Machining. For environments 1, 2, and 3, the collected data provide a varied motor speed of 1797, 1772, and 1750, respectively. Fig. 3 shows the histogram of all elements in the data collection used in the analysis.

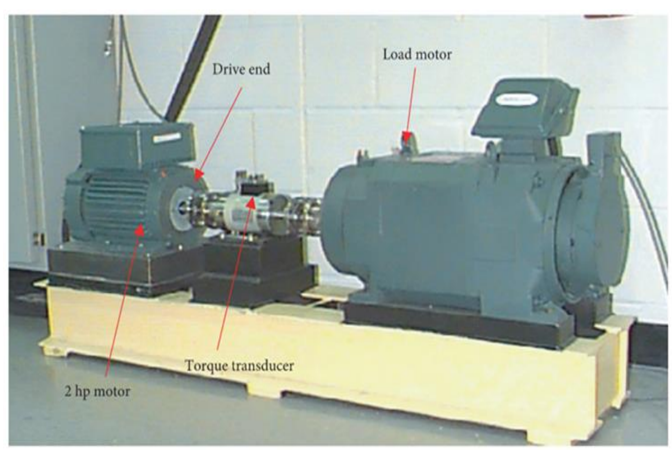


Fig. 2. Accelerometers Sensors placed on bearing Simulator of CWRU.

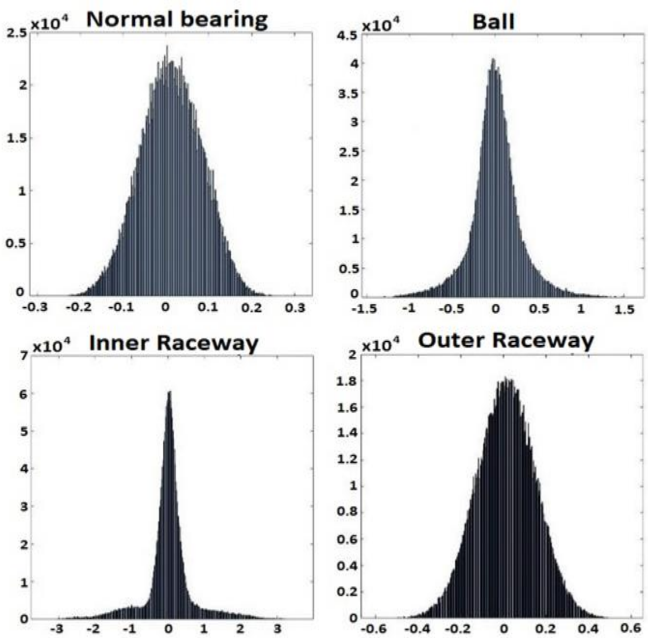


Fig. 3. Histogram for all Elements in Dataset.

We started by loading all of the signal data and creating three sets of data. Label Binarizer was used to label the products and reshaped all three datasets. This work used a random seed value of 0.2 and utilized 30% of our data for testing and 70% of our data for training.

For training our model with Convolution 1D, Convolution 2D, and MobileNetV2 used Tensorflow and Keras library because of its convenient coding environment and ability to train a state-of-an-art algorithm for signal processing and computer vision.

B. Convolution 1D and Convolution 2D

The 2D in Conv2D refers to the fact that each channel in the input and filter is two-dimensional, while the 1D in Conv1D refers to the fact that each channel in the input and filter is one-dimensional. Normalizing the data in our Convolution 1D and Convolution 2D models initially set up the first hidden layer with 100 nodes and implemented the RELU activation feature. MaxPooling1D was used with a pool size of 2 to minimize the dimension feature. The second and third convolution layers have 32 and 10 nodes, respectively. The signals were translated into NumPy arrays to speed up the computation. For backpropagation, the learning rate was set at 0.001 and used Categorical Cross-Entropy and the Adam optimizer equation to calculate the loss function. The categorical Cross-Entropy loss function is used for the multiclass classification of the dataset. After applying all of the optimizers (Adam, Nadam, Adagrad, RMSProp, Adadelta, SGD, Adamax), Adam optimizer is chosen for the highest accuracy on the dataset. Since the batch size for instruction is 32, and the decay is set to 0.1. Tables I and II depict the convolution 2D and convolution 1D models, respectively.

C. MobileNetV2

MobileNetV2 is a CNN architecture that tends to be efficient on mobile devices. MobileNetV2 has 32 filters on its initial fully convolution layer. There exist 19 residual bottleneck layers. It is utilized for image classification, object detection, quantization, and so on [32].

Two types of blocks are introduced in MobileNetV2.

1) Residual block of stride 1.
2) Block for downsizing with 2 strides.

Both the blocks are made up of three layers, as illustrated in Fig. 4. With 1x1 convolution, the ReLU6 activation mechanism is used in the first layer. On the second sheet, a depth-wise is added, and the third layer is also a 1x1 convolution, save for some non-linearity. The activation mechanism of ReLu is often included in the third layer [33]. MobileNetV2 performs well when the mathematical operations and the number of parameters are kept low. The MobileNetV2 architecture is about 35% faster than the previous version, MobileNetV1.

TABLE I. PROPOSED CONVOLUTION 2D MODEL

Layer (type)	Output shape	Parameter
Conv2d_4 (Conv2D)	(None, 30, 30, 32)	320
max_pooling2d_4 (MaxPooling2D)	(None, 15, 15, 32)	0
conv2d_5 (Conv2D)	(None, 13, 13, 64)	18496
max_pooling2d_5 (MaxPooling2D)	(None, 6, 6, 64)	0
dense_4 (Dense)	(None, 6, 6, 10)	650
dropout (Dropout)	(None, 6, 6, 10)	0
max_pooling2d_6 (MaxPooling2D)	(None, 3, 3, 10)	0
flatten_2 (Flatten)	(None, 90)	0
dense_5 (Dense)	(None, 4)	364
Total params: 19,830 Trainable params: 19,830 Non-trainable params: 0		

[a] Shape and parameters are depended on Convolution 2D model layers. Various layers are implemented.

TABLE II. PROPOSED CONVOLUTION 1D MODEL

Layer (type)	Output shape	Parameter
conv1d_2 (Conv1D)	(None, 31, 100)	6500
dropout_2 (Dropout)	(None, 31, 100)	0
max_pooling1d (MaxPooling1D)	(None, 15, 100)	0
conv1d_3 (Conv1D)	(None, 14, 32)	6432
dense_2 (Dense)	(None, 14, 10)	330
max_pooling1d_1 (MaxPooling1D)	(None, 7, 10)	0
flatten_1 (Flatten)	(None, 70)	0
dense_3 (Dense)	(None, 4)	284
Total params: 19,830 Trainable params: 19,830 Non-trainable params: 0		

[a] The shape and parameters are determined by the layers of the Convolutional 1D model. Numerous layers have been introduced.

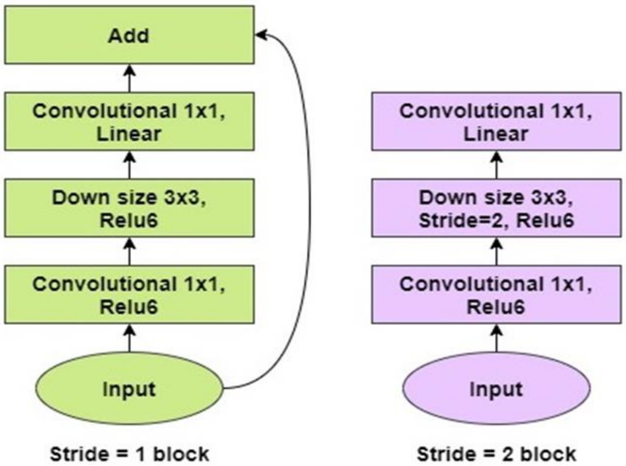


Fig. 4. MobileNetV2 Architecture.

The base layer of the MobileNetV2 platform has been frozen and replaced with the proposed trainable layer. The Relu activation feature is extended to 512 nodes in the proposed layer. In the output sheet, there are 4 nodes and a SoftMax activation mechanism for classifying faults. The loss function is determined using categorical cross-entropy, and the learning factor is set to 0.001.

D. Performance Measures

Precision, recall, f1-score, and accuracy evaluate the models' performance after completing the training and testing phase. Precision is the closeness of the measurements to each other, while accuracy is the proximity of the measurements to a particular value. The equations used to calculate the measures are stated in Eq. (1) and Eq. (2). The ability of a model to identify all of the data points of relevance in a dataset is referred to as recall. F1-score measures the rate of recognizing real threats and not being distracted by false alarms. The measures of Eq. (3) and Eq. (4).

$$Precision = \frac{TP}{TP+FP} \tag{1}$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$F1-Score = 2 * \frac{Recall * Precision}{Recall + Precision} \tag{4}$$

In the Equations, TP represents true positive, TN represents true negative, and FP and FN represent false positive and false negative.

IV. Result Analysis and Discussion

The proposed methods can successfully identify faulty machinery with high accuracy [34]. After processing collected data samples, all three elements (inner raceway, outer raceway, and ball) considered in our Dataset were found to have 0.011 inches fault depth.

A. MobilenetV2 Result Analysis

Table III shows that the best training accuracy was at epoch number 17, where the accuracy was 99.86 percent, and the best validation accuracy was 99.64 percent. The model has a 99.83 percent precision.

Fig. 5 depicts the precision of training and testing for each epoch and the data loss graph for MobileNetV2 in the training and test sets. The graph shows that with increasing epochs, detection accuracy increases, and data loss gradually decreases for both training and test sets.

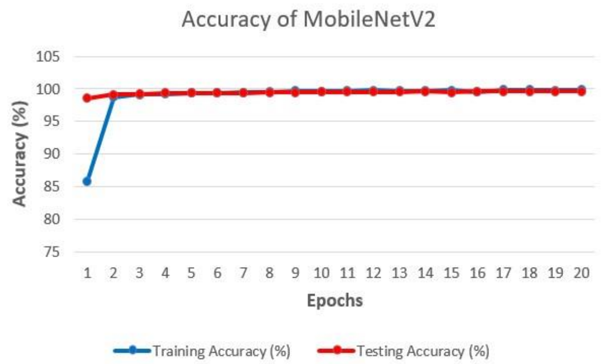
The classification report for each class (normal, ball, inner and outer raceway) of the test dataset is shown in Fig. 6. The model shows high performance for all the classes individually.

From the test data, the heat map is shown below in Fig. 7. The MobileNetV2 predicted 19229 properly in a normal class, where 222 samples are not correctly predicted. In the case of the class ball, 20572 samples are correctly predicted, where 425 data are not correctly predicted. Similarly, in the case of inner class and outer class, 18571 and 17604 samples are correctly predicted, where 749 and 1155 samples are improperly classified.

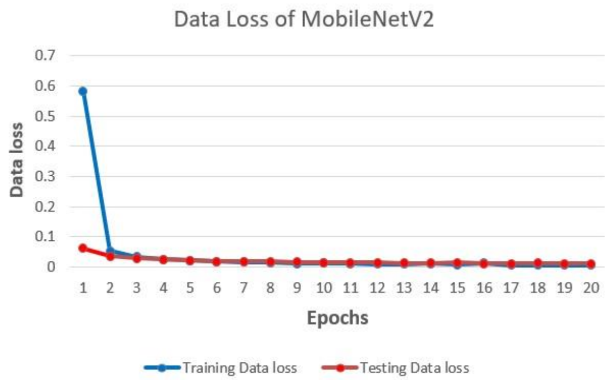
TABLE III. Data Loss and Accuracy Obtained From MobileNetV2

Epoch	**Training Set**		**Testing Set**	
	Data loss	***Accuracy (%)***	***Data loss***	***Accuracy (%)***
1	0.5838	85.82	0.0624	98.60
2	0.0527	98.74	0.0360	99.12
3	0.0338	99.12	0.0291	99.20
4	0.0262	99.23	0.0248	99.34
5	0.0214	99.36	0.0213	99.34
6	0.0186	99.40	0.0192	99.41
7	0.0165	99.47	0.0184	99.41
8	0.0155	99.54	0.0183	99.45
9	0.0117	99.67	0.0168	99.44
10	0.0129	99.61	0.0145	99.52
11	0.0109	99.68	0.0149	99.51
12	0.0098	99.74	0.0148	99.54
13	0.0105	99.73	0.0134	99.52
14	0.0107	99.67	0.0129	99.61
15	0.0071	99.75	0.0142	99.45
16	0.0134	99.54	0.0119	99.62
17	0.0062	99.86	0.0119	99.64
18	0.0070	99.85	0.0123	99.61
19	0.0059	99.80	0.0119	99.64
20	0.0061	99.85	0.0116	99.62

[b] Data sample of the MobileNetV2 model's 20 epochs.



(a)



(b)

Fig. 5. (a) Accuracy for each Epoch, (b) Data Loss Graph, for MobileNetV2 Architecture.

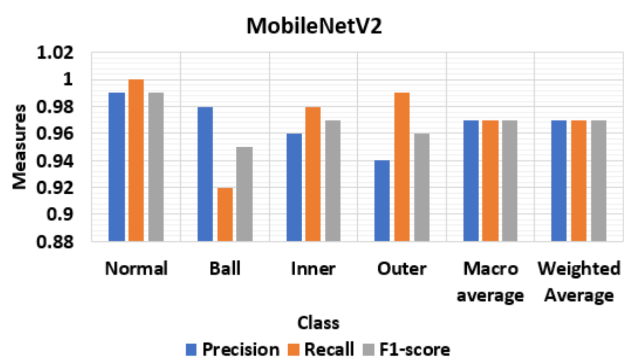


Fig. 6. Classification Report for each Class.

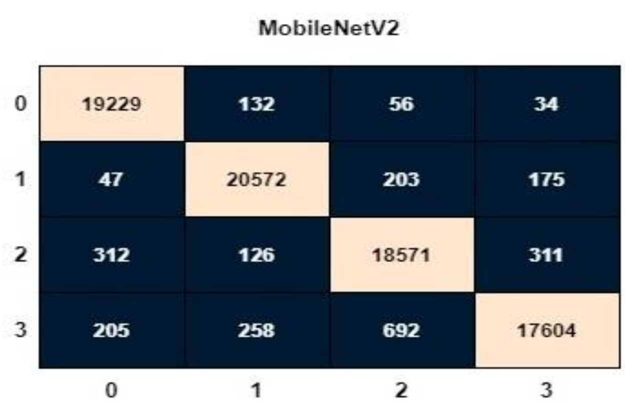


Fig. 7. Heat Map for MobileNetV2.

B. Convolution 1D result Analysis

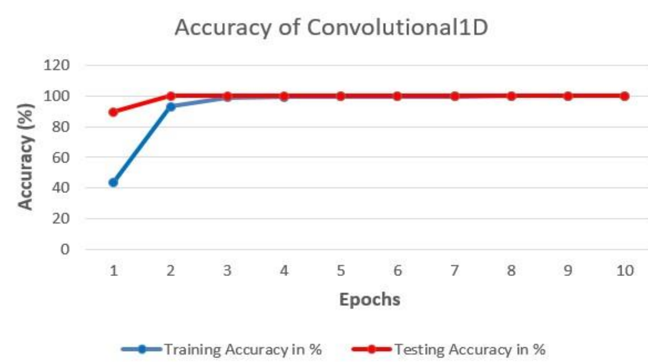
Table IV shows the accuracy of the training and test sets for Convolution 1D. The highest outcome seen in the training set is 99.97%, while the maximum accuracy demonstrated in the test set is 100 percent. Set 3 was used to train the model, and Set 2 and Set 1 were used for testing.

TABLE IV. THE DATA LOSS AND ACCURACY OBTAINED FROM CONVOLUTION 1D.

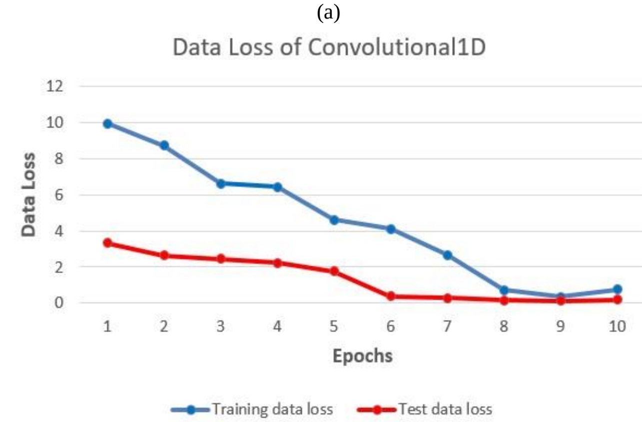
Epochs	Training data loss	Training Accuracy in %	Test data loss	Testing Accuracy in %
1	9.96	43.88	3.34	89.60
2	8.74	92.98	2.65	99.89
3	6.64	99.03	2.45	100.0
4	6.43	99.42	2.25	100.0
5	4.62	99.48	1.76	99.96
6	4.11	99.65	0.38	100.0
7	2.68	99.71	0.29	100.0
8	0.73	99.79	0.17	99.95
9	0.35	99.87	0.13	100.0
10	0.77	99.97	0.19	100.0

[a] Data sample of the Convolution 1D 's 10 epochs.

The data loss and accuracy graph are shown in Fig. 8. For the ten epochs, accuracy increases and data loss decreases from training data and test data.



(a)



(b)

Fig. 8. (a) Data Accuracy, (b) Data Loss Graph, of Convolution 1D.

The scores of precision, recall, accuracy, and f1- score are shown in Fig. 9. The model provides perfect scores for each of the performance tests for all the data classes.

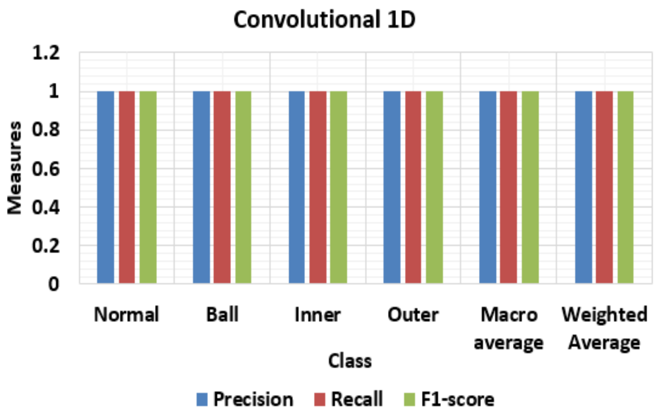


Fig. 9. Classification Report of Convolution 1D.

In Fig. 10, the heat map demonstrates the predicted and non-predicted test data. The Convolution 1D accurately predicted 19451 samples in a normal class; however it incorrectly predicted 0 samples. A total of 20997 samples are accurately predicted when it comes to the class ball, whereas no samples are incorrectly predicted. Additionally, 19320 and 18759 samples from the inner and outer classes are accurately predicted, but no samples are incorrectly classified.

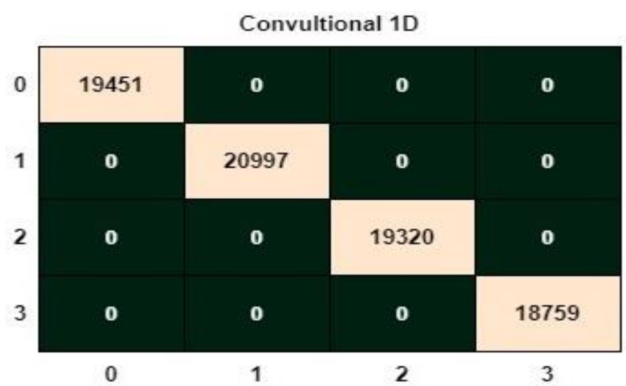


Fig. 10. Heat Map for Convolution 1D.

C. Convolution 2D Result Analysis

In Table V, the result of Convolution 2D is stated. Both the training and validation sets have an accuracy of 100%. The model shows high accuracy on each epoch. Additionally, the data loss for the test set is much lower.

The data accuracy and loss graph are seen in Fig. 11. It is seen that the data loss for the training set gradually decreases with the increased number of epochs. The accuracy rate stays consistently high for the test set, and the data loss stays consistently low on each epoch.

The classification report in Fig. 12 illustrates that the Conventional 2D model shows a perfect performance score throughout all the classes of the test Dataset, the same as the Conventional 1D model.

TABLE V. THE DATA LOSS AND ACCURACY OBTAINED FROM CONVOLUTION 2D

Epochs	Training data loss	Training Accuracy in %	Test data loss	Testing Accuracy in %
1	12.04	84.95	0.07	99.98
2	10.67	99.78	0.06	99.95
3	8.35	99.84	0.03	99.94
4	7.55	99.92	0.03	100.0
5	6.21	99.97	0.02	99.93
6	3.00	99.99	0.03	99.98
7	1.96	99.89	0.03	100.0
8	0.97	99.97	0.02	100.0
9	0.033	100.0	0.02	99.94
10	0.031	99.99	0.02	100.0

b. Data sample of the Convolution 1D 's 20 epochs.

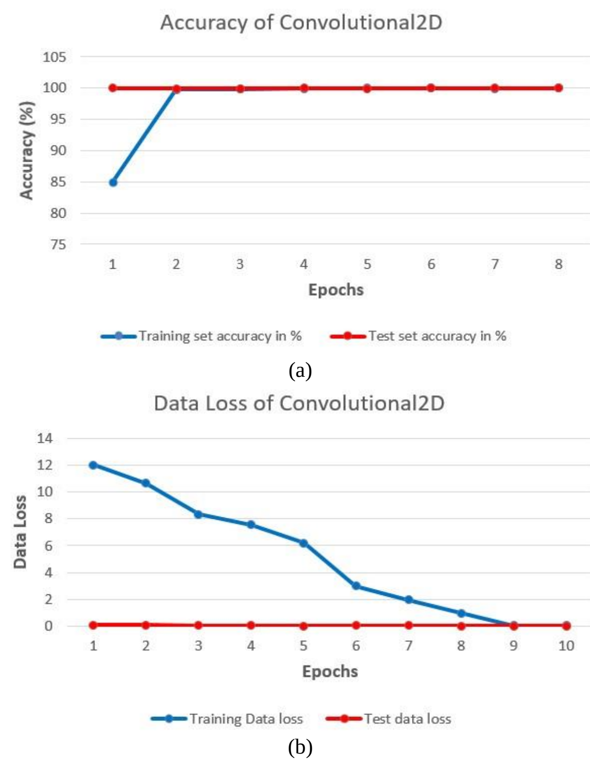


Fig. 11. (a) Data Accuracy, (b) Data Loss Graph using Convolution 2D.

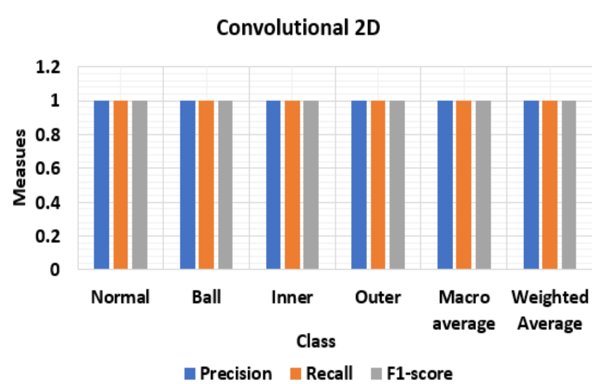


Fig. 12. Classification Report of Convolution2D.

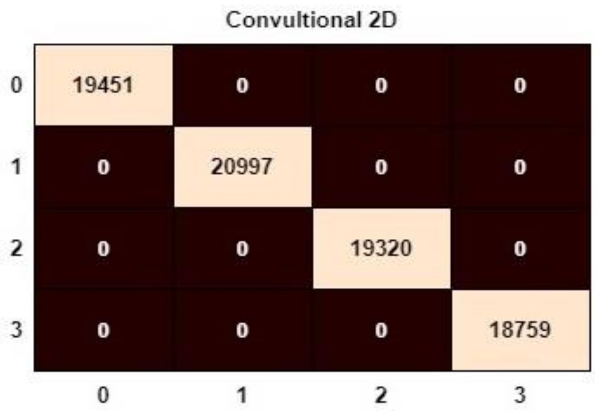


Fig. 13. Heat map for Convolution 2D.

In Fig. 13, the heat map is illustrated the predicted and non-predicted test data. The Convolution 2D successfully predicted 19451 samples in a normal class, yet it wrongly predicted Nil samples. When it comes to the class ball, a total of 20997 samples are properly predicted, while nil samples are wrongly predicted. Furthermore, 19320 and 18759 samples from the inner and outer classes are properly predicted, so the predicted rate is 100%.

D. Result Discussion and Comparison

The CNN algorithms used in the study all show promising results. MobileNetV2 shows the accuracy of 99.64% and lowest data loss of 0.01, whereas both the Convolutional1D and 2D show a perfect accuracy score of 100% at the 10th epoch with CPU time of 82.5secs/epoch and 93secs/epoch, respectively. The comparison among the performance of the algorithms' prediction is illustrated in Table VI.

TABLE VI. COMPARISON OF THE ALGORITHMS

CNN Models	Ep	Tr_ loss	Tr_ Acc in %	T_ loss	T_Acc %	CPU_T secs/Ep
M_NetV2	20	0.006	99.86%	0.01	99.64%	108.65
Cn_1D	10	0.77	99.97%	0.19	100.0%	82.5
Cn_2D	10	0.03	99.99%	0.02	100.0%	93.0

[a.] Convolution 1D and Convolution 2D Perform better with highest accuracy and low CPU time than MobileNetV2. Here, Epochs is abbreviated 'Ep', Training loss is abbreviated 'Tr_loss', Training accuracy is abbreviated 'Tr_Acc', Test loss is abbreviated 'T_loss', Training Accuracy is abbreviated 'T_Acc'. CPU Time is abbreviated 'CPU_T',

All three algorithms of the proposed system achieve very high performance in detecting faults in machinery. However, fault detection has been done in the past using other machine learning algorithms [35-37]. A comparison of the performance of such algorithms with the proposed system is made in Table VII. From the comparison, it is observed that even though CNN and GAN algorithms achieve an accuracy of over 97%, it is much lower than the accuracy achieved by the algorithms proposed in the study.

The proposed system successfully detects a fault in industrial machinery up to 100%. Compared to the existing techniques, the performance is much higher. It is essential to detect faults with the highest accuracy in industrial machinery, as even a minor increase in fault detection accuracy might prevent a tragic accident. Since the machines used in industrial work generally contain a lot of small or big parts, at the same time, machines themselves can be huge, maintaining it can be difficult for humans. The proposed fault detection system can be used to make the process automatic and efficient. Using the system, the sound vibration from the machinery can be processed by MobileNetV2, Convolution 1D, and Convolution 2D to detect faults in the machines to avoid future hassles.

TABLE VII. COMPARISON OF THE PROPOSED SYSTEM WITH EXISTING SYSTEMS

Models	Proposed System	Existing System 1 [38]	Existing System 2 [39]	Existing System 3 [40]	Existing System 4 [27]
MobileNetV2	99.64%	-	-	-	-
Convolutional 1D	100%	-	-	-	-
Convolutional 2D	100%	-	-	-	-
SVM	-	97.3%	-	-	-
CNN	-	98.1%	-	-	-
SVM + PCA	-	74.4%	-	-	-
Stack Denoising Auto Encoder	-	-	91.67%	-	-
GAN	-	-	97.96%	-	-
DNN	-	-	94.4%	-	-
ANN	-	-	-	95%	
ANN + 10-fold Cross Validation	-	-	-	93.54%	-
CNN	-	-	-	-	71.67% 86.67% 98.33%

[a.] A comparison between the proposed study and existing related work. Proposed study offers better findings than existing related studies.

V. CONCLUSION

As scientific and technical knowledge increases, mechanical equipment becomes more sophisticated and automated. Mechanical equipment relies significantly on spinning mechanical components like bearings and lead screws to work properly. Damaged or failed bearing components will cause equipment failure and fatalities. As a result, monitoring the bearing components' performance is essential. This article presented a viable method for bearing defect detection based on accelerometer sensors and the wavelet transformation (DWT) signal processing methodology, with a ZigBee-based wireless sensor network architecture for effectively sending data to a diagnostic server. MobileNetV2 architecture pre-trained model compares the model with two custom CNN models: 1D and 2D deep CNN architectures. A bearing dataset collected by accelerometers sensors is used to validate the models that consist of 4 types of fault signals. Upon the four classes, we have achieved a satisfactory result. With MobileNetV2, the system was able to identify problems with 99.64% accuracy and 99.83% precision. It achieves up to 100% accuracy and precision when utilizing the Convolutional 1D and 2D architectures. In the future, the proposed architecture may be tested with more parameters besides implementing hybrid transfer learning models and the most reliable IoT framework.

REFERENCES

[1] Souza, R. M., Nascimento, E. G., Miranda, U. A., Silva, W. J., & Lepikson, H. A. (2021). Deep learning for diagnosis and classification of faults in industrial rotating machinery. *Computers & Industrial Engineering*, *153*, 107060.

[2] Ahmad, W., Khan, S. A., Islam, M. M., & Kim, J. M. (2019). A reliable technique for remaining useful life estimation of rolling element bearings using dynamic regression models. *Reliability Engineering & System Safety*, *184*, 67-76.

[3] Baptista, M., Sankararaman, S., de Medeiros, I. P., Nascimento Jr, C., Prendinger, H., & Henriques, E. M. (2018). Forecasting fault events for predictive maintenance using data-driven techniques and ARMA modeling. *Computers & Industrial Engineering*, *115*, 41-53.

[4] Guo, X., Chen, L., & Shen, C. (2016). Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. *Measurement*, *93*, 490-502.

[5] Li, W., Gu, S., Zhang, X., & Chen, T. (2020). Transfer learning for process fault diagnosis: Knowledge transfer from simulation to physical processes. *Computers & Chemical Engineering*, *139*, 106904.

[6] Chuang, Y. C., Chen, T., Yao, Y., & Wong, D. S. H. (2018). Transfer learning for efficient meta-modeling of process simulations. *Chemical Engineering Research and Design*, *138*, 546-553.

[7] Lei, Y., Jia, F., Lin, J., Xing, S., & Ding, S. X. (2016). An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. *IEEE Transactions on Industrial Electronics*, *63*(5), 3137-3147.

[8] Boukra, T., Lebaroud, A., & Clerc, G. (2012). Statistical and neural-network approaches for the classification of induction machine faults using the ambiguity plane representation. *IEEE Transactions on Industrial Electronics*, *60*(9), 4034-4042.

[9] Murphey, Y. L., Masrur, M. A., Chen, Z., & Zhang, B. (2006). Model-based fault diagnosis in electric drives using machine learning. *IEEE/ASME Transactions On Mechatronics*, *11*(3), 290-303.

[10] Wang, J., Gao, R. X., & Yan, R. (2014). Multi-scale enveloping order spectrogram for rotating machine health diagnosis. *Mechanical Systems and Signal Processing*, *46*(1), 28-44.

[11] Powers, E. J., Shin, Y. J., Mack Grady, W., Böhme, J. F., Carstens-Behrens, S., Papandreou-Suppappola, A., ... & Onchis, D. M. (2016). Time-frequency diagnosis, condition monitoring, and fault detection.

[12] Yan, R., Gao, R. X., & Chen, X. (2014). Wavelets for fault diagnosis of rotary machines: A review with applications. *Signal processing*, *96*, 1-15.

[13] Chakraborty, S., Shamrat, F. J. M., Afrin, S., Saha, S., Ahmed, I., & Thapa, S. (2021, October). Industrial Fault Detection Using Transfer Learning Models. In *2021 2nd International Conference on Smart Electronics and Communication (ICOSEC)* (pp. 1-6). IEEE.

[14] Zhang, W., Peng, G., Li, C., Chen, Y., & Zhang, Z. (2017). A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors*, *17*(2), 425.

[15] Zhang, Z., Li, X., Wen, L., Gao, L., & Gao, Y. (2019, August). Fault diagnosis using unsupervised transfer learning based on adversarial network. In *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)* (pp. 305-310).

[16] Wan, Z., Yang, R., & Huang, M. (2020). Deep Transfer Learning-Based Fault Diagnosis for Gearbox under Complex Working Conditions. *Shock and Vibration*, *2020*.

[17] Chong, U. P. (2011). Signal model-based fault detection and diagnosis for induction motors using features of vibration signal in two-dimension domain. *Strojniški vestnik*, *57*(9), 655-666.

[18] Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., & Nandi, A. K. (2020). Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical Systems and Signal Processing*, *138*, 106587.

[19] Rai, V. K., & Mohanty, A. R. (2007). Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert–Huang transform. *Mechanical systems and signal processing*, *21*(6), 2607-2615.

[20] Lou, X., & Loparo, K. A. (2004). Bearing fault diagnosis based on wavelet transform and fuzzy inference. *Mechanical systems and signal processing*, *18*(5), 1077-1095.

[21] Konar, P., & Chattopadhyay, P. (2011). Bearing fault detection of induction motor using wavelet and Support Vector Machines (SVMs). *Applied Soft Computing*, *11*(6), 4203-4211.

[22] Bin, G. F., Gao, J. J., Li, X. J., & Dhillon, B. S. (2012). Early fault diagnosis of rotating machinery based on wavelet packets—Empirical mode decomposition feature extraction and neural network. *Mechanical Systems and Signal Processing*, *27*, 696-711.

[23] Lei, Y., He, Z., & Zi, Y. (2011). EEMD method and WNN for fault diagnosis of locomotive roller bearings. *Expert Systems with Applications*, *38*(6), 7334-7341.

[24] Lin, L., & Hongbing, J. (2009). Signal feature extraction based on an improved EMD method. *Measurement*, *42*(5), 796-803.

[25] He, M., & He, D. (2017). Deep learning based approach for bearing fault diagnosis. *IEEE Transactions on Industry Applications*, *53*(3), 3057-3065.

[26] Sun, W., Chen, J., & Li, J. (2007). Decision tree and PCA-based fault diagnosis of rotating machinery. *Mechanical Systems and Signal Processing*, *21*(3), 1300-1317.

[27] Guo, S., Yang, T., Gao, W., & Zhang, C. (2018). A novel fault diagnosis method for rotating machinery based on a convolutional neural network. *Sensors*, *18*(5), 1429.

[28] Shao, S., Yan, R., Lu, Y., Wang, P., & Gao, R. X. (2019). DCNN-based multi-signal induction motor fault diagnosis. *IEEE Transactions on Instrumentation and Measurement*, *69*(6), 2658-2669.

[29] Yu, W., & Zhao, C. (2019). Broad convolutional neural network based industrial process fault diagnosis with incremental learning capability. *IEEE Transactions on Industrial Electronics*, *67*(6), 5081-5091.

[30] Dataset: https://github.com/XiongMeijing/CWRU-1?fbclid=IwAR0RacMKCv2CIB1GD3JOYHSmSraXixoc5S8lyIltvacvbFDlTkgL08AD1_Y. Last Accessed: 18 August, 2021.

[31] Duan, S., Zheng, H., & Liu, J. (2019). A novel classification method for flutter signals based on the CNN and STFT. *International Journal of Aerospace Engineering*, *2019*.

[32] MobileNetV2: https://paperswithcode.com/method/mobilenetv2. Last Accessed: 24 August, 2021.

[33] Akter, S., Shamrat, F. M., Chakraborty, S., Karim, A., & Azam, S. (2021). COVID-19 Detection Using Deep Learning Algorithm on Chest X-ray Images. Biology, 10(11), 1174.

[34] Biswas, A., Chakraborty, S., Rifat, A. N. M. Y., Chowdhury, N. F., & Uddin, J. (2020, August). Comparative Analysis of Dimension Reduction Techniques Over Classification Algorithms for Speech Emotion Recognition. In *International Conference for Emerging Technologies in Computing* (pp. 170-184). Springer, Cham.

[35] Ghosh, P., Azam, S., Karim, A., Hassan, M., Roy, K., & Jonkman, M. (2021). A comparative study of different machine learning tools in detecting diabetes. *Procedia Computer Science*, *192*, 467-477.

[36] Ghosh, Pronab, et al. "A performance based study on deep learning algorithms in the effective prediction of breast cancer." 2021 International Joint Conference on Neural Networks (IJCNN). IEEE, 2021.

[37] Shamrat, F. J. M., Chakraborty, S., Ahammad, R., Shitab, T. M., Kazi, M. A., Hossain, A., & Mahmud, I. (2022). Analysing most efficient deep learning model to detect COVID-19 from computer tomography images. *Indonesian Journal of Electrical Engineering and Computer Science*, *26*(1), 462-471.

[38] Verstraete, D., Ferrada, A., Droguett, E. L., Meruane, V., & Modarres, M. (2017). Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings. *Shock and Vibration*, *2017*.

[39] Neupane, D., & Seok, J. (2020). Bearing fault detection and diagnosis using case western reserve university dataset with deep learning approaches: A review. *IEEE Access*, *8*, 93155-93178.

[40] Vakharia, V., Gupta, V. K., & Kankar, P. K. (2016). Bearing fault diagnosis using feature ranking methods and fault identification algorithms. *Procedia Engineering*, *144*, 343-350.

Three Layer Authentications with a Spiral Block Mapping to Prove Authenticity in Medical Images

Ferda Ernawan[1], Afrig Aminuddin[2], Danakorn Nincarean[3], Mohd Faizal Ab Razak[4], Ahmad Firdaus[5]
Faculty of Computing, Universiti Malaysia Pahang, Pekan 26600, Malaysia[1, 3, 4, 5]
Faculty of Computer Science, Universitas Amikom Yogyakarta, Sleman, 55283, Indonesia[2]

***Abstract*—Digital medical image has a potential to be manipulated by unauthorized persons due to advanced communication technology. Verifying integrity and authenticity have become important issues on the medical image. This paper proposed a self-embedding watermark using a spiral block mapping for tamper detection and restoration. The block-based coding with the size of 3×3 was applied to perform self-embedding watermark with two authentication bits and seven recovery bits. The authentication bits are obtained from a set of condition between sub-block and block image, and the parity bits of each sub-block. The authentication bits and the recovery bits are embedded in the least significant bits using the proposed spiral block mapping. The recovery bits are embedded into different sub-blocks based on a spiral block mapping. The watermarked images were tested under various tampered images such as blurred image, unsharp-masking, copy-move, mosaic, noise, removal, and sharpening. The experimental results show that the scheme achieved a PSNR value of about 51.29 dB and a SSIM value of about 0.994 on the watermarked image. The scheme showed tamper localization with accuracy of 93.8%. In addition, the proposed scheme does not require external information to perform recovery bits. The proposed scheme was able to recover the tampered image with a PSNR value of 40.45 dB and a SSIM value of 0.994.**

Keywords*—*Fragile watermarking; self-embedding; image authentication; self-recovery; medical image; spiral block mapping

I. Introduction

The development of internet technology has grown exponentially in the last decade. Internet technology made digital multimedia data easy to be distributed to the world. With the rapid development of information technology, multimedia data was produced and sent seamlessly without boundaries and limitations. The digital image can be produced by various electronic devices such as digital cameras and X-ray machines for medical images [1], [2]. In addition, there are many available images editing software to modify these digital images, such as Photoshop, Lightroom, and GIMP. However, the advancement of this software also enables unauthorized modification to a digital image that makes it hard to identify the authenticity and integrity of the images. The medical images require the images to be authentic to prevent false decisions under some circumstances. Thus, any alteration or slight modifications cannot be accepted. Digital images could be modified by unauthorized persons for illegal use. The modification can be invisible or visible to the human eye. Furthermore, the illegal modification may lead to illegal action against the law. Image watermarking techniques can be an alternative solution to authenticate digital image content [3].

Digital image watermarking is the process of embedding a watermark into digital images [4]. The watermark itself can contain a logo, serial number, or a security key. The watermark is then embedded into the cover image to be visible or imperceptible. An example of a visible watermark is a transparent logo as seen embedded in the corner of an image to prevent copyright violation. However, this type of watermark was prone to watermark removal by simply cropping or overlaying the new watermark over the original watermark. Hence, the researcher primarily focused on the imperceptible watermark with a security feature to not be easily removed by any attack. Digital watermarking techniques can be used to protect copyright and authenticate digital image content. There are three types of image watermarking: robust, semi-fragile, and fragile watermarking. Robust image watermarking is mainly used for copyright protection so that any modification on the image content, the embedded watermark should be resistant against several attacks. In contrast, semi-fragile and fragile watermarking are primarily utilized for image authentication. Fragile image watermarking is not resistant to any modifications on the watermarked image. Fragile watermarking has been widely used to detect tamper localization and authenticate the image content. Fragile watermarking does not allow modification to the image content, while semi-fragile image watermarking will tolerate minor changes to the image, such as image compression [5], [6].

A fragile image watermarking for authentication provides four important aspects which are watermark generation, insertion technique, tamper localization, and tamper recovery. The watermark generation presented a way to generate a watermark image to be embedded into the cover image. The existing schemes use the part of the cover image as the watermark image, which is called self-embedding fragile image watermarking [7], [8]. The watermark generation also can be obtained from a set of embedding bits that contains the authentication bit and recovery bits [9], [10]. The embedding watermark in fragile watermarking can be performed by modifying the bits of the host image. The embedding of watermark bits into the first Least Significant Bit (LSB) made it invisible to the human eye [11]. The embedding of the watermark into the second LSB significantly contributes to the reconstruction error. Lastly, the tamper recovery algorithm plays a key role in recovering the tampered area of the image.

The recovered image quality depends on the size of the tampered area.

Image authentication techniques can be used for verifying or authenticating the integrity of digital media content. Image authentication can be classified into passive and active authentication. Active authentication requires preliminary data from the original image. This data could be stored on a secured database so that the authentication can be done by comparing the data from the database and the extracted data from the receiving ends [12]. Another scheme directly embedded the data into the cover images, so it does not need any intermediate database to store the authentication data, while it slightly reduces the quality of the image [13], [14]. In addition, some active authentication schemes also support tamper recovery along with authentication and tamper detection [15], [16]. In contrast, passive authentication was performed without requiring any preliminary image information for authentication. Passive authentication was categorized as forgery dependent and independent. On one hand, forgery dependent will only detect some types of forgery, such as copy-move and splicing forgery. This type of forgery may not leave any visual trace, but the inconsistencies of the underlying statistics can detect it. On the other hand, forgery independent will detect the tampered image based on the visible artifacts left during the resampling process and lighting inconsistencies [17]. Thus, passive authentication has limited tamper detection capability depending on its forgery types and the visual artifacts left behind. Furthermore, passive authentication did not support tamper recovery compared to active authentication [18], [19].

The existing authentication schemes still do not achieve high accuracy in tamper detection. The high accuracy of tamper detection is significantly important for achieving the high quality of the self-recovered image. Therefore, this study presents three-layer authentications with a spiral block mapping that can support tamper recovery. The parity bit of each block is extracted to obtain the authentication bit for the first level. If the bit value is the same, then the second authentication bit will be computed and compared with the average pixel of its block image. If the average pixel of the sub-block is greater than the average pixel of its block, then the third authentication is performed to check the output of the second layer authentication in three RGB channels. The proposed scheme embeds the image content itself into the host image based on a spiral block mapping. The embedding process will be performed for each block of 3×3 pixels, the nine least significant bit is modified for authentication and recovery. The proposed embedding scheme utilized a set of conditions bits. The recovery bits are embedded into the different locations of sub-block images. The proposed spiral block mapping can detect any tamper occurred on the watermarked image. The proposed scheme will produce a higher accuracy of the tamper detection and higher quality of the recovered image than other existing schemes after tampering the image.

II. Related Work

Belferdi et al. [15] presented a self-embedding fragile watermarking scheme for color image tamper detection and recovery. The scheme implemented the Bayer pattern to reduce the size of embedding watermarks into the original image. The watermark was generated from the original image by decomposing the image into red, green, and blue channels. The scheme converted the color watermark into a grayscale image by using the Bayer filter to reduce the watermark size. This filter selects one color sample from each color channel. This grayscale watermark was decomposed into four sub-blocks to improve security and enhance robustness. Each sub-image was then permuted three times based on automorphic permutation using three keys to ensure maximum security. Finally, each permuted sub-images is converted into a binary sequence and embedded three times into the least significant bits (LSB) of the three RGB components of the host images. The scheme can improve detection accuracy if one channel was destroyed.

Gul and Ozturk [20] showed a fragile image of watermarking and tamper-detection based on the SHA-256 hash function. The scheme divided the cover image into non-overlapping blocks with the size of 32×32 pixels. Each block was then sub-divided into four sub-blocks of 16×16 pixels. The watermark was generated using the entire SHA-256 hash value of the first three sub-blocks. The resulting 256-bit binary watermark was embedded into the least significant bit (LSB) of the fourth sub-block. The tamper detection works by comparing the hash value of the first three blocks to the extracted watermark obtained from the fourth block.

Hisham et al [2] presented a watermarking scheme for tamper detection and self-recovery. The scheme presented a spiral pattern during embedding the watermark image. While the pattern of embedding a watermark in the LSB does not give effect to the quality of the watermarked image and security. The spiral pattern did not work for the non-square image; some parts of the image can't be embedded using a spiral pattern. The embedding watermark using a spiral pattern doesn't give any contributions to the watermarked image especially for modifying LSB. The embedding scheme was performed on each block of 8×8 pixels. Each block was divided into sub-block with the size of 4×4 pixels. Each sub-block with the size of 4×4 was embedded by nine bits, including two (v and p) authentication bits and seven recovery bits. The scheme considers only 9 bits over 16 bits of LSB on each sub-block. The scheme did not fully utilize the rest of the seven bits for embedding watermark images. The scheme has the potential to consume a large computational time and it did not optimize the embedding space of the host image. The recovery bits in the scheme are generated from the average pixel value on its image sub-block. The scheme is designed to recover the tampered area by using an average value of its sub-block image with the size of 4x4 pixels. If any small tamper occurred in the sub-block of 4×4 pixels, the average pixel value will replace it. The scheme has the potential to produce less quality of the recovery bits due to average pixels, even if the tamper that occurred was a small area.

III. Proposed Scheme

A. Proposed self-embedding Watermark

The experiments utilize seven medical images with different sizes namely "Abdomen", "Brain", "Breast", "Chest", "Eye", "Teeth", and "Womb" images. The block diagram of the proposed scheme is depicted in Fig. 1.

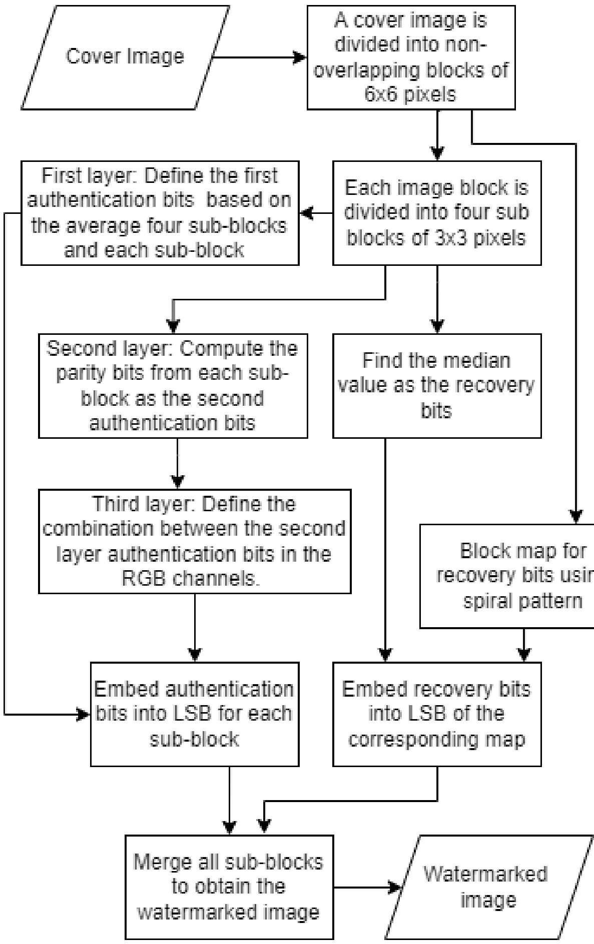


Fig. 1. Block Diagram of the Proposed Embedding Scheme.

A watermark image is obtained from the two bits, v and p for authentication and seven bits of the sub-block recovery. Each pixel of the cover image was embedded with watermark bits on the LSB of the cover image. The selected sub-block of 3×3 pixels in the embedding watermark is to achieve an optimal embedding of two authentication bits and seven recovery bits into the LSB of the cover image. The modifying LSB of the cover image does not give a significant effect on the quality of the medical image. The following steps discuss the proposed self-embedding watermark:

1) A cover image is divided into non-overlapping blocks of 6×6 pixels. Then, each block is divided into four sub-blocks with the size of 3×3 pixels. The visualization of the block and four sub-block images is shown in Fig. 2.

2) The average pixel value of image block AvgB is calculated, and each of its sub-block AvgSB.

3) The first authentication bits were computed by comparing the average image block of 6×6 pixels AvgB and each of its sub-block with the size of 3×3 pixels AvgSB. If the average AvgB is larger than AvgSB, the authentication bit denoted as v is 1, and vice-versa.

4) The second authentication bits were generated from the parity bits of each sub-block. The authentication bit denoted as p is 1 if the parity number is equal to an odd number, and 0 if the parity bit is an even number.

5) The first and second authentication bits are embedded in each LSB of each sub-block. The illustration of each embedding authentication bits, v and p into the sub-block is shown in Fig. 3.

According to Fig. 3, the white region represents the original bit of pixels on each sub-block, *P* represents the pixel value, *v* is the first layer authentication bit and *p* is the second layer authentication bit.

6) The recovery bits of each sub-block are generated by finding the median value of its sub-block. The seven most significant bits of median pixel value are used as recovery bits. The recovery bits are embedded based on a spiral pattern as visualized in Fig. 4. In the proposed scheme, the authentication bits and recovery bits are embedded in different locations. The recovery bits on the first block are embedded into the nineteenth block and so on. The recovery bits for each block are embedded based on the spiral block mapping as shown in Fig. 4. The proposed spiral block mapping can avoid tampered coincidence.

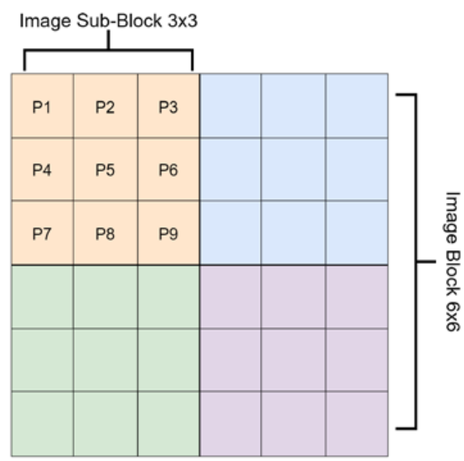


Fig. 2. Illustration of Image Block and each of its Sub-block.

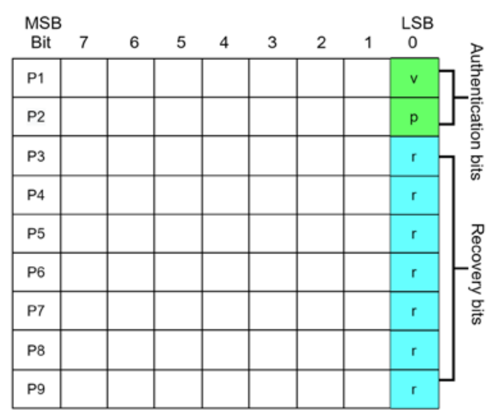


Fig. 3. The Location of the Watermark Bits in the Cover Image.

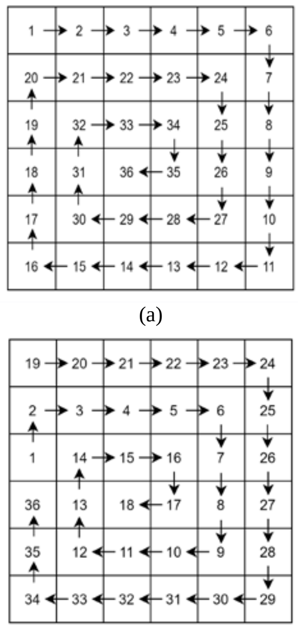


Fig. 4. The Illustration of (a) A Spiral Pattern of its Block and (b) A Spiral Block Mapping for Embedding the Recovery Bits.

7) The self-embedding watermark is repeated for all sub-blocks to obtain the watermarked image.

B. Proposed Tamper Detection and Recovery Scheme

The watermarked images are tested under several attacks such as blur, unsharp, cloning, mosaic, noise, removal, and sharpen images. The proposed scheme provides two-level authentications. The procedures of tamper detection are defined by:

1) First-level authentication: A tampered image is divided into non-overlapping blocks with the size of 6×6 pixels. Then, each block is divided into four sub-blocks with the size of 3×3 pixels. The extract bit v from LSB on the first pixel of its sub-block in the watermarked image. The parity bit of each block is computed, if the parity number is an odd number, the parity bit v' is set to 1, otherwise, set it to 0. The parity bit v' is compared with the extracted v from each LSB of its sub-block. If the bit value of v' and v are not equal, then mark it as a tampered image, otherwise, no tamper is detected. In addition, if the bit value between v' and v is the same, then the second authentication bit p will be checked in the second level authentication.

2) Second-level authentication: In the second authentication bit, the extracted bit *p* from LSB on the second pixel is compared with the *p*' represents algebraic relation between sub-block with the size of 3×3 pixels and block image with the size of 6×6 pixels. The average pixel of each sub-block is computed and compared with the average pixel of its block image. if the average pixel of the sub-block is greater than the average pixel of its block, the *p*' is 1, otherwise, the *p*' is 0. If the bit value of p' and p are not equal, then mark it as a tampered image, otherwise, no tamper is detected. The proposed authentication is defined in Algorithm 1 as follows:

Algorithm 1. The proposed multilayer authentication levels

```
Input: v, v', p, p'
α = 0;
for i=1 to 4
 if (v ~= v')
 # mark as tampered sub-block
 α = 1;
 else
 if (p ~= p')
 # mark as tampered sub-block
 α = 1;
 elseif (α == 1)
 # mark as tampered sub-block
 α = 0;
 else
 # mark as untampered sub-block
 α =0;
 end
end
Output: tampered, untampered sub-block
```

where *v* denotes the extracted bit from LSB on the first pixel of its sub-block in the watermarked image, *v*' represents the parity bit of each sub-block, *p* is the extracted bit from LSB on the second pixel of its sub-block, and *p*' represents algebraic relation between sub-block and block image. The α value is 1 if the sub-block is marked as a tampered sub-block. The α value is used as a reference of its block that has tamper detection. If the previous sub-block has tampered with the attack, the next sub-block has a possibility of tampered image.

3) Third-level authentication: The third layer authentication checks the result obtained from the second layer authentication in three RGB channels. If an image block of the RGB channel is detected as a tamper, then set all blocks of RGB channels on its block locations to be tampered. This third layer authentication further reduces the false-negative detection. The diagram of the proposed tamper detection and the recovery bits is shown in Fig. 5.

C. Evaluation of the Proposed Scheme

The proposed tamper detection algorithm is evaluated by using a confusion matrix such as True Positive Rate (TPR), False Negative Rate (FNR), and False Positive Rate (FPR). TPR represents the ratio between the detected area against the real tampered. The highest TPR means that the tamper detections are correctly detected in the tampered regions. In contrast, FNR means the ratio between the undetected area compared to the real tampered area. The high FNR means inaccurate tamper detection in the tampered area of the images. Next, the FPR represents the ratio between the false detected area against the untampered area. The range of FPR values is between 0 to 1, the higher FPR value represents the higher detection of the untampered area as tampered area or false detection. The proposed scheme is also measured in terms of precision and accuracy for evaluating tamper detection. The TPR and FNR are defined by [11]:

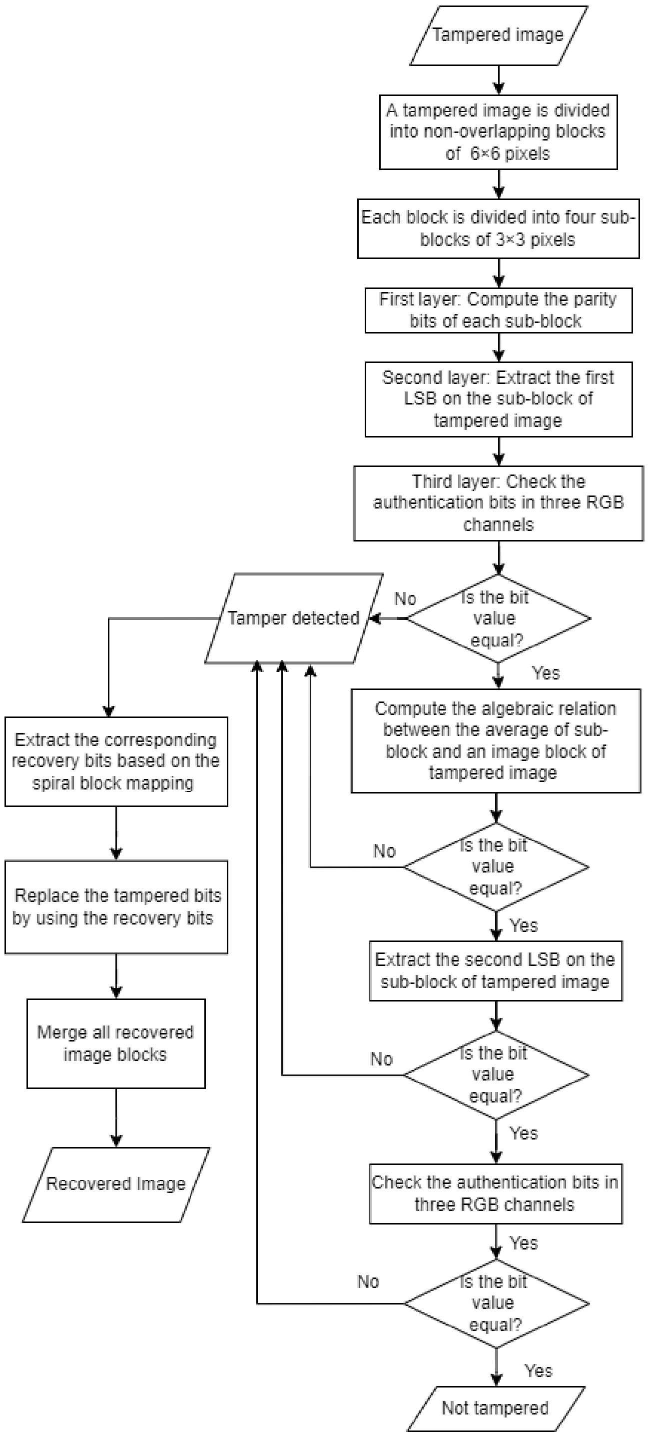


Fig. 5. Block Diagram of the Tamper Detection and Recovery Schemes.

$$TPR = \frac{TP}{TP + FN} = \frac{TP}{P} = 1 - FNR \quad (1)$$

$$FNR = \frac{FN}{TP + FN} = \frac{FN}{P} = 1 - TPR \quad (2)$$

where TP represents the number of true-positive tampered pixels, FN denotes the number of false-negative tampered pixels, P represents the number of real tampered pixels. The FPR is defined by:

$$FPR = \frac{FP}{FP + TN} = \frac{FP}{N} \quad (3)$$

where FP represents the number of false-positive tampered pixels, TN denotes the number of true-negative tampered pixels, N represents the number of untampered pixels. The precision and accuracy of the tampered detection are defined by [11]:

$$Precision = \frac{TP}{TP + FP} = \frac{TPR}{TPR + FPR} \quad (4)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{TP + TN}{P + N} \quad (5)$$

The precision represents the precise tamper detection on the image. The higher precision value means that the proposed scheme was able to produce high true positive values and low false-positive values. The accuracy of tampering detection is evaluated in order to measure the effectiveness of the proposed scheme. High accuracy on tamper detection can improve the recovery bits of the tampered image.

The quality of the watermarked image and recovered image is evaluated in terms of the imperceptibility by using PSNR, MSE and SSIM. The PSNR is defined as follows [21]-[24]:

$$PSNR = 10 \log \frac{(S)^2}{MSE} \quad (6)$$

$$MSE = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \left(f(i,j) - g(i,j) \right)^2 \quad (7)$$

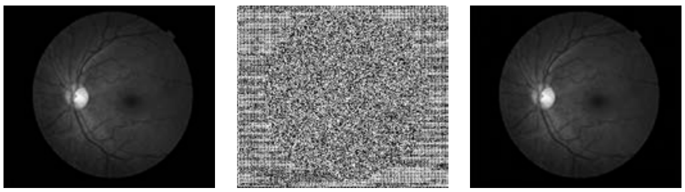
where $f(i,j)$ is the original medical image, $g(i,j)$ is the watermarked image, M, N denotes as the row and column sizes of the medical image. MSE represents the difference value between original and watermarked image, S denotes a maximum pixel value of the medical image. The SSIM is defined as follows [25]-[32]:

$$SSIM(x, y) = [l(x, y)]^f \cdot [c(x, y)]^g \cdot [s(x, y)]^h \quad (8)$$

where $\alpha>0$, $\beta>0$, $\gamma>0$, are parameters to define the relative importance of the three components l, c and s. The SSIM index is defined as three similarity terms of luminance, contrast, and structural between two local windows. The SSIM considers image degradation as the perceived change in structural information, separating the measure of similarity into luminance, contrast, and structure [33]. The range SSIM value in between 0 to 1, the higher SSIM value indicated that the watermarked image is structurally similar to the original cover image.

IV. Experimental Results

The proposed scheme embeds self-image content as a piece of watermark information on the least significant bit (LSB) of the cover image. Each pixel of the cover image has 8-bits of information which corresponds to the range of pixel values 0 to 255. Embedding watermark information into LSB has some advantages include less distortion on the watermarked image and fast embedding watermark image. Thus the embedding LSB algorithm can preserve the quality of the cover image. The embedding watermark by using LSB can provide high sensitivity against an altered image. A small altered image into the pixel value can be identified. Therefore, the LSB method has been widely used on fragile watermarking for image authentication. This study proposed a self-embedding scheme for tamper detection and restoration in the medical image by using multilayer authentications. The visual original covers a medical image, a watermark image obtained from its image content and the watermarked image is shown in Fig. 6.



(a) (b) (c)

Fig. 6. (a) Original Eye Image, (b) Watermark Image, (c) Watermarked Image.

According to Fig. 6, the experimental results showed that the proposed scheme achieves high quality of the watermarked image with the PSNR value of 51 dB. The experimental results of the proposed scheme have been compared with the scheme by Hisham et al. [2]. The imperceptibility of the watermarked image in terms of MSE, PSNR, and SSIM is shown in Table I. Referring to Table I, it can be noticed that a scheme by Hisham et al. [2] achieved a lower MSE value compared with the proposed method. In addition, the scheme also produced higher PSNR and SSIM values than the proposed scheme.

A scheme by Hisham et al. [2] embedded nine bits out of sixteen bits as watermark information on each sub-block of 4x4 pixels. A scheme by Hisham et al. [2] used a sub-block size of 4x4 pixels which contains 16 pixels on each sub-block. All pixels are not embedded by the watermark bits, only nine bits 56.25% of pixels are modified on the LSB of the sub-block pixels. The remaining seven bits 43.75% are maintained as original bits or unmodified. The proposed scheme embedded the watermark bits on the sub-block of 3x3 pixels which contains 9 pixels. All the pixels are modified with watermark bits on the LSB of each sub-block. The proposed scheme embeds two bits as authentication bits and seven bits as recovery blocks as the watermark bits. Modifying all pixel values of the sub-block image can contribute to the reconstruction error. Therefore, the scheme by Hisham et al. [2] produced higher PSNR and SSIM values than the proposed scheme. The comparison of the existing methods using the block-based scheme for image authentication is shown in Table II.

Referring to Table II, it can be noticed that modifying 1 LSB of the cover image produced a higher quality of the watermarked image with a PSNR value of 50 dB above. Schemes by Huang et al. [7], Sing et al. [10], and Tohidi et al. [19] did not inform the detection accuracy. In addition, the existing watermarking schemes based on block coding required an original cover image or external information in order to recover the image. The schemes by Belferdi et al. [15] and Tohidi et al. [19] embedded watermark information into the first and second LSB of the cover image. The schemes produce PSNR values of about 44 dB. At the same time, the schemes produced high distortion on the recovery image. The schemes by Belferdi et al. [15] and Tohidi et al. [19] produced the quality of the recovered image with PSNR values of 40dB and 32 dB respectively. The schemes by Huang et al. [7] and Sing et al. [10] embedded watermark data into the first, second, and third LSB of the image. The scheme produced a PSNR value of about 39 dB due to modifying 3 LSB on the cover image. The recovered image still produced a PSNR value of 41 dB and 39 dB, respectively. Even though the scheme presented a self-embedding watermark, it still required reference bits for authentication. A scheme by Hisham et al. [2] was able to produce high imperceptibility of the watermarked image with a PSNR value of 53 dB. The scheme embedded nine bits out of sixteen bits for authentication and recovery bits. The remaining seven bits are maintained as the original bits. Therefore, the scheme can achieve a high PSNR value of the watermarked image. While the scheme produced low accuracy on the tamper detection due to limited embedding watermark for each sub-block. A scheme by Hisham et al. [2] also required reference bits for recovering the medical images. The scheme produced a quality of the recovered image with a PSNR value of 40 dB. The proposed scheme achieved a slightly low quality of the watermarked image compared with a scheme by Hisham et al. [2] with a PSNR value of 51 dB. In addition, the proposed scheme provides higher accuracy of tamper detection than a scheme by Hisham et al. [2]. The proposed scheme produced a superior quality of the recovered image than the existing benchmark of the watermarking schemes. The proposed scheme is tested under various tampering rates, and then the results are evaluated by using True Positive Rate (TPR), False Negative Rate (FNR), and False Positive Rate (FPR). The comparison of the TPR, FNR and FPR under various tampering conditions is listed in Table III.

TABLE I. THE COMPARISON OF THE IMPERCEPTIBILITY PERFORMANCE BETWEEN THE PROPOSED SCHEME AND A SCHEME BY HISHAM ET AL. [2]

Medical Image	**PSNR (dB)**		**SSIM**	
	Hisham et al. [2]	**Proposed scheme**	**Hisham et al. [2]**	**Proposed scheme**
Abdomen	53.6267	51.1424	0.9995	0.9992
Brain	53.8694	51.3455	0.9978	0.9938
Breast	54.1428	51.7481	0.9951	0.9876
Chest	53.6200	51.1389	0.9996	0.9993
Eye	53.1564	50.9746	0.9953	0.9923
Teeth	53.6523	51.1339	0.9978	0.9961
Womb	54.1882	51.6033	0.9970	0.9910
Average	53.7508	51.2981	0.9974	0.9942

TABLE II. COMPARISON OF THE QUALITY, TAMPER LOCALIZATION, RECOVERY OF THE EXISTING WATERMARKING SCHEME

Method	PSNR Embedding	Embedding Location	Detection	PSNR Recovery	Recovery	Authentication type
Hisham et al. [2]	53.7508 dB	1 LSB	0.7668	40.7861 dB	Require original cover image	Active
Belferdi et al. [15]	44.2495 dB	2 LSB	1.0000	40.7300 dB	Using Bayer Pattern	Active
Huang et al. [7]	39.0900 dB	3 LSB	-	41.3200 dB	Using ROI & RONI	Active
Sing et al. [10]	39.8600 dB	3 LSB	-	39.1400 dB	Using Block Truncation Coding	Active
Tohidi et al. [19]	44.0000 dB	2 LSB	-	32.0000 dB	Using Compression Strategy	Active
Proposed scheme	51.2981 dB	1 LSB	0.8774	44.7922 dB	Not require the original cover image	Active

The watermarked medical images obtained from the scheme by Hisham et al. [2] and the proposed scheme are tested under various attacks and each tamper attack was subjected to the same size of the attack. According to Table III, the seven medical images have been tested under various tampering rates. The experimental results show that the proposed scheme achieves a higher average TPR than a scheme by Hisham et al. [2]. The scheme by Hisham et al. [2] presented multilayer authentication based on a set of conditions on the average value of sub-block and parity bits of sub-block with the size of 4x4 pixels. The scheme has the potential for detecting 75% certainty of the tampered sub-block. First, the parity bit on each sub-block was compared with the first extracted bit of its sub-block. The first layer authentication has a probability of being 50% undetected. The second layer authentication performed a comparison of the average sub-block and block image. This stage also provided a 50% probability of the previous undetected tampered area. The proposed scheme presented multilayer authentication bits by considering the previous tampered block. The first layer authentication compared the average value of the sub-block with the size of 3x3 pixels and block with the size of 6x6 pixels. If the bits are not equal, the sub-block was marked as tampered sub-block and the α was set to 1. Otherwise, the bits are the same, then the second layer authentication is computed to check the authenticity of bits. The second layer of authentication compared the parity of bits on each sub-block with the extracted bits obtained from LSB on its sub-block. If the bit values are not the same, then the sub-block was assigned as a tampered sub-block. In addition, if the α value is 1, the sub-block image was marked as a tampered block. If the previous sub-block has a tampered with, its block image has a high probability tamper. The proposed tamper detection algorithm successfully increased the detection rate compared with the scheme by Hisham et al. [2]. A scheme by Hisham et al. [2] produced an average TPR rate of 76.68%. The proposed scheme successfully detects the tampered image with a high TPR rate of 87.74%. The accuracy of the tamper detection scheme under various tamper attacks is shown in Fig. 7.

According to Fig. 7, the proposed scheme also provided a slightly higher FPR rate than the scheme by Hisham et al. [2], it does not significantly affect the reliability of the proposed scheme. It has been proven by achieving a higher accuracy rate compared to the scheme by Hisham et al. [2] as shown in Table III. According to Table III, our scheme produced slightly lower precision than the scheme by Hisham et al. [2]. The proposed scheme achieved higher accuracy than other schemes with an average accuracy of about 0.997. Both schemes provide multilayer authentication for detecting tamper images. Our scheme proposed tamper localization by using multilayer authentication with considering the previous sub-block α. Referring to Fig. 7, it has been proven that the proposed authentication algorithm can improve the accuracy of tamper detection. The proposed authentication algorithm can detect higher accuracy of the tamper localization compared with the scheme by Hisham et al. [2]. The proposed scheme is also able to recover bits against various tampered attacks. The experimental results of the recovered image are listed in Table IV.

TABLE III. COMPARISON OF THE TPR, FNR AND FPR UNDER VARIOUS TAMPERING RATES

Medical Image	Tamper Attacks	Tampering Rate (%)	TPR		FNR		FPR		Accuracy	
			Hisham et al. [2]	Proposed	Hisham et al. [2]	Proposed	Hisham et al. [2]	Proposed	Hisham et al. [2]	Proposed
Abdomen	Blurring	0.95	0.7261	**0.8566**	0.2739	**0.1434**	**0.0010**	0.0012	0.8626	**0.9277**
Brain	Unsharp Mask	0.70	0.7532	**0.8642**	0.2468	**0.1358**	**0.0007**	0.0010	0.8763	**0.9316**
Breast	Copy-Move	0.60	0.7493	**0.9523**	0.2507	**0.0477**	**0.0004**	0.0006	0.8745	**0.9759**
Chest	Mosaic	0.80	0.8343	**0.8523**	0.1657	**0.1477**	**0.0011**	0.0014	0.9166	**0.9255**
Eye	Noise	2.0	0.7639	**0.8790**	0.2361	**0.1210**	**0.0017**	0.0023	0.8811	**0.9384**
Teeth	Removal	1.35	0.7691	**0.8409**	0.2309	**0.1591**	0.0015	**0.0013**	0.8838	**0.9198**
Womb	Sharpening	0.91	0.7716	**0.8962**	0.2284	**0.1038**	**0.0008**	0.0011	0.8854	**0.9476**
Average			0.7668	**0.8774**	0.2332	**0.1226**	**0.0010**	0.0013	0.8829	**0.9380**

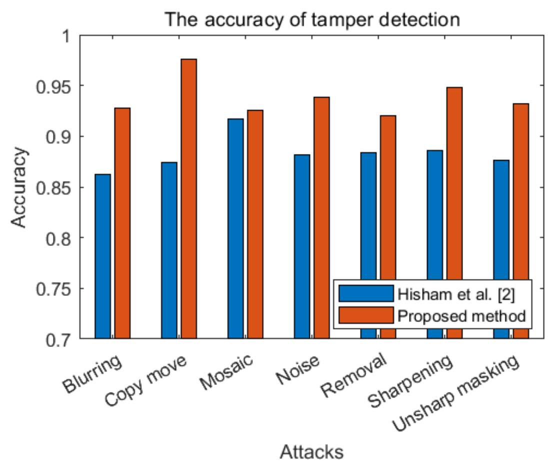


Fig. 7. The Comparison on the Accuracy of Tamper Detection between the Proposed Scheme and a Scheme by Hisham et al. [2].

TABLE IV. COMPARISON OF THE QUALITY OF THE RECOVERED BITS UNDER VARIOUS TAMPERING RATES

Medical Image	PSNR (dB)		SSIM	
	Hisham et al. [2]	**Proposed scheme**	**Hisham et al. [2]**	**Proposed scheme**
Abdomen	39.4512	**49.4710**	0.9960	**0.9988**
Brain	43.5453	**45.4482**	0.9948	**0.9955**
Breast	44.3019	**46.8068**	0.9968	**0.9985**
Chest	42.6248	**48.8064**	0.9972	**0.9984**
Eye	36.9968	**39.6542**	0.9773	**0.9812**
Teeth	40.8027	**42.9007**	0.9920	**0.9935**
Womb	37.7798	**40.4584**	0.9966	**0.9975**
Average	40.7861	**44.7922**	0.9930	**0.9948**

The watermarked images obtained from the proposed scheme were tested under various tamper attacks. Each watermarked image was tampered with using the common attack in the medical images such as blurring, unsharp masking, copy-move, mosaic, noise, removal, and sharpening. Each attack is carried out in a round shape with a diameter of 60 pixels to simulate a real-world attack. The visualization of the comparison between the proposed scheme and other existing schemes for the recovered image is depicted in Fig. 8.

The scheme by Hisham et al. [2] used the spiral block mapping for embedding watermark and retrieved the recovery embedded bits from another block by using pseudorandom equations. The secret key used the total number of blocks in the image. This block mapping introduced a tamper coincidence problem during the recovery process. The recovery bits may be altered due to randomized blocks. Therefore, the scheme was not able to retrieve the embedded recovery bits. In addition, the scheme only can be performed with the square image size. The block numbering starts from the center of the image and spirals outward to the edge of the image. If the image input is not a rectangle, then the block mapping algorithm will not be able to handle outside the square area in the center of the image and the outside area will be unprotected. This study proposed a spiral inward block mapping started from the top-left of the image and ended in the center of the image. To solve this tamper coincidence problem, the proposed method maps the first half of spiral inward mapping to the second half of the spiral block mapping. This technique greatly decreases the tamper coincidence problem during the recovery bits. This approach also can overcome the square-image problem presented by Hisham et al. [2]. The proposed method successfully protects all the image pixels as well as the rectangle images. The proposed method successfully recovers the image from the tampered image. The proposed scheme achieved higher quality recovered images than the scheme by Hisham et al. [2] with an average PSNR value of 44.7922 dB and an average SSIM value of 0.9948. The proposed scheme presented a smaller sub-block code with the size of 3x3 pixels for embedding watermark. Our scheme proposed a spiral block map to prevent the tamper coincidence problem. The proposed scheme is also tested under different tampering rates. The quality of the recovered images is shown in Fig. 9.

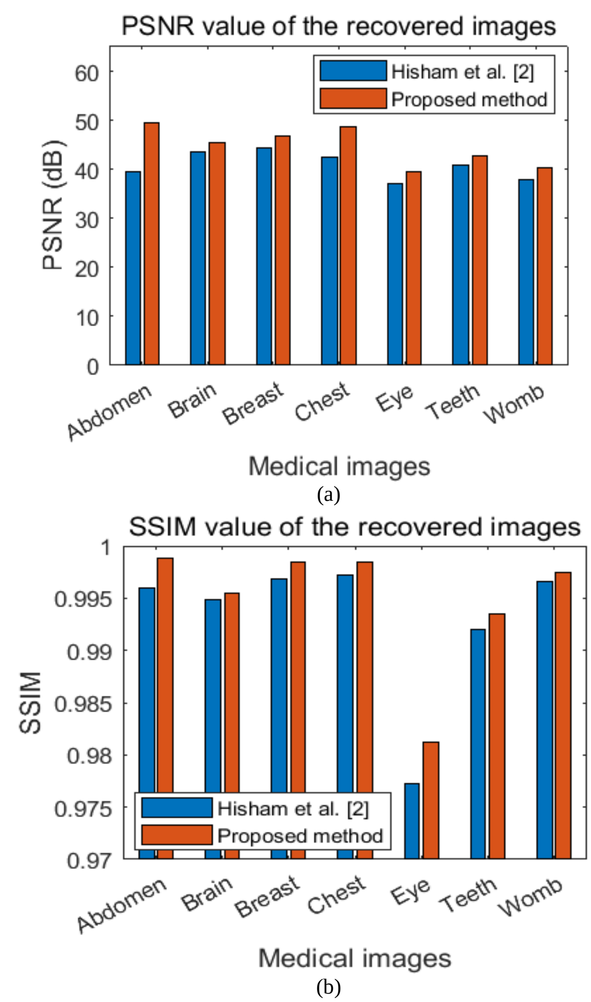
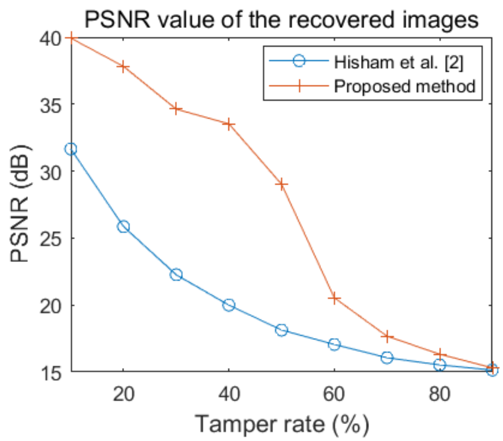
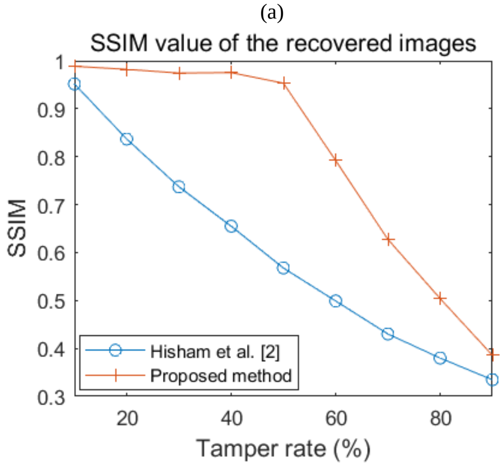


Fig. 8. The Quality of the Recovered Image from the Proposed Scheme and a Scheme by Hisham et al. [2] (a) PSNR (b) SSIM.



(a)

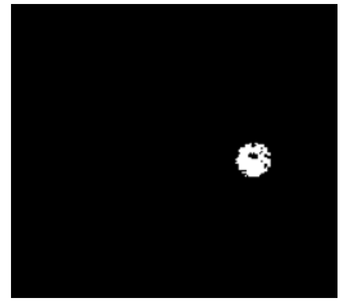


(b)

Fig. 9. The Quality of the Recovered Image: (a) PSNR Value, (b) SSIM Value under different Types of Tampering Rate.

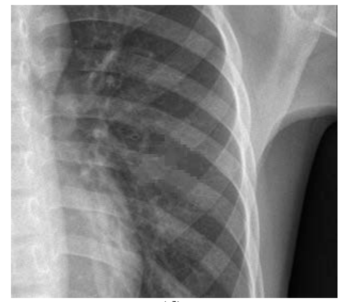
In order to evaluate the effectiveness of the proposed scheme, the different amounts of tampering rates are applied to the watermarked image. The experimental results have shown that the proposed scheme achieved superior quality of the recovered image under different tampering conditions. According to Fig. 9, it can be seen that the proposed scheme can achieve high image quality of the recovered image under tampering rates of 10%, 20%, 30%, 40%, and 50%. The quality of the recovered image decreased when the watermarked image tampered with the tampering rates of 60%, 70%, 80%, and 90%. The proposed scheme can achieve PSNR values of 40db under 10% tampering rate, 35dB under 30% tampering rate, and significantly decrease to 20dB under 60% tampering rate. The proposed scheme produced a high-quality image in terms of SSIM value of 0.99 under tampering rates of 10%, 20%, 30%, 40%, and 50%, while it suddenly decreased when the watermarked image was tampered with at a 60% tampering rate. The visual tamper detection and recovered image from the tamper attack are shown in Fig. 10.

(a)



(b)

(c)



(d)

Fig. 10. (a) Tampered Image, (b) Tamper Detection, (c) Image Recovery, (d) Image Recovery with Zoom in 200%.

According to Fig. 10, the proposed scheme was able to recover the tampered image using the mosaic attacks. Fig. 10(a) shows the tampered "Chest" image by a mosaic attack where the tampered area is shown on the right side. Fig. 10(b) presented that the proposed scheme can show the tamper localization as shown in the white color. The proposed scheme can achieve a detection rate value of 85.23% for the tampered "Chest" image. Next, Fig. 10(c) shows the recovered "Chest" image, and the recovered image with zoom in 200% is shown in Fig 10(d). According to Fig. 10(d), our scheme can achieve high quality of the recovered image, it is closer to the original medical image. In addition, the proposed scheme can achieve high accuracy of tamper detection. The visual comparison of the recovered image between Hisham et al. [2] and the proposed scheme is shown in Table V. The detail visual tamper detection and recovered image under various attacks is shown in Tables VI and VII.

V. Conclusion

This research has presented self-embedding a watermark for medical image authentication and recovery. The watermark bits have been generated from its image content; it consists of two authentication bits and seven recovery bits. The first authentication bit is obtained from the comparison between an average pixel of sub-block and block image. The second authentication bit has been generated from the parity bits of each sub-block. The recovery bits are determined by finding the median value of each sub-block. The recovery bits are embedded in the different locations based on spiral block mapping. The proposed scheme maps the first half of spiral inward mapping to the second half of the spiral block mapping.

TABLE V. VISUAL COMPARISON OF THE QUALITY OF THE RECOVERED BITS UNDER VARIOUS TAMPERING RATES

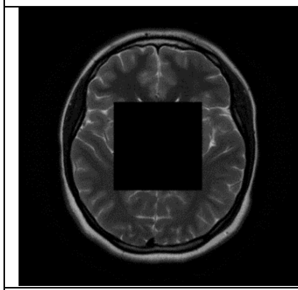
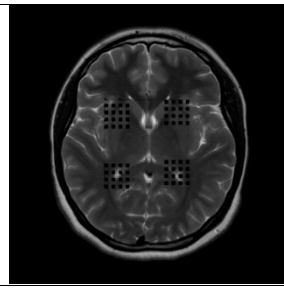
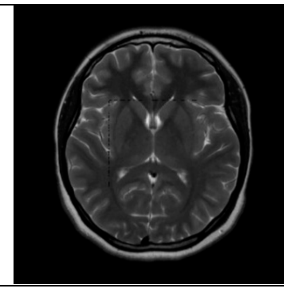
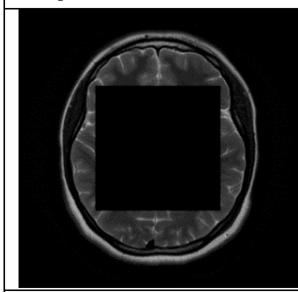
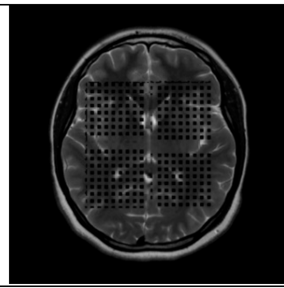
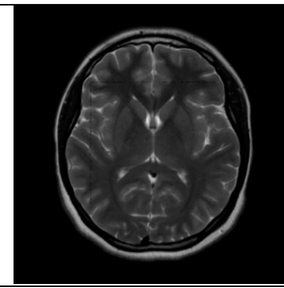
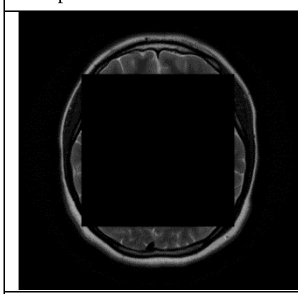
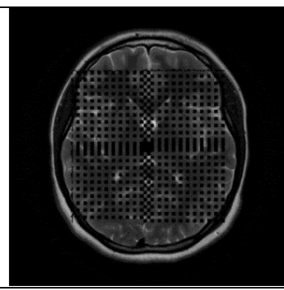
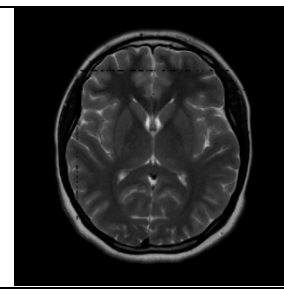
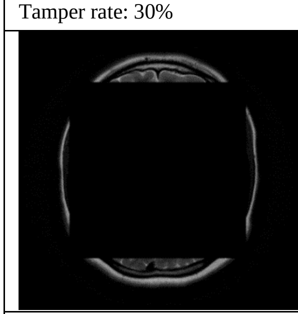
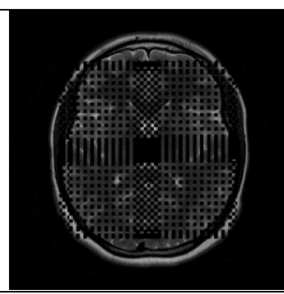
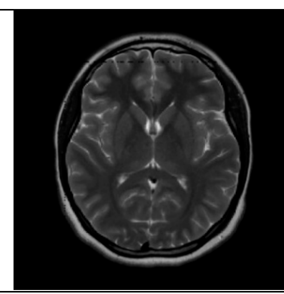
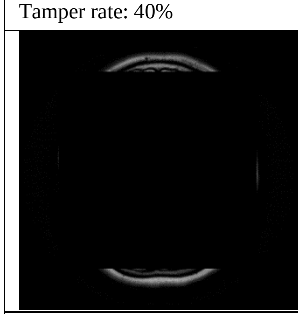
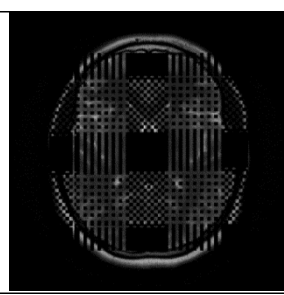
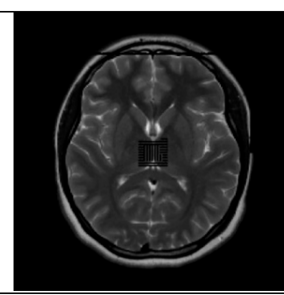
Tampered Image (Tampering rate)	**Recovered Image**	
	Hisham et al. [2] (PSNR / SSIM)	**Proposed Method (PSNR / SSIM)**
		
Tamper rate: 10 %	31.6484 dB / 0.9516	39.9283 dB / 0.9885
		
Tamper rate: 20%	25.8563 dB / 0.8368	37.8172 dB / 0.9822
		
Tamper rate: 30%	22.2600 dB / 0.7372	34.6273 dB / 0.9744
		
Tamper rate: 40%	19.9915 dB / 0.6550	33.5292 dB / 0.9754
		
Tamper rate: 50%	18.1313 dB / 0.5677	29.0157 dB / 0.9533

TABLE VI. VISUAL TAMPER DETECTION AND RECOVERED IMAGE FROM THE PROPOSED SCHEME UNDER VARIOUS TAMPER ATTACKS

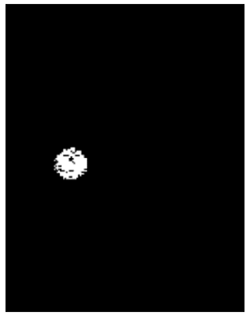
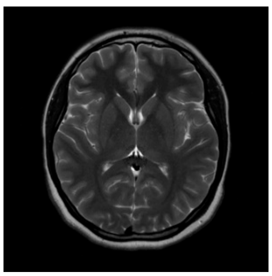
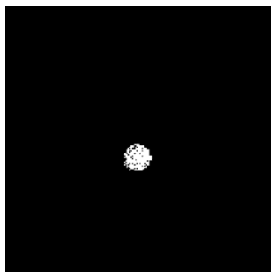
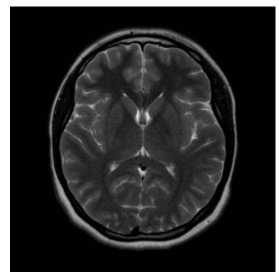
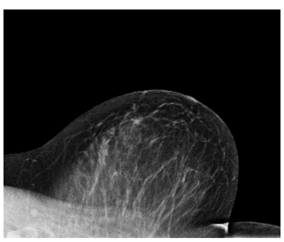
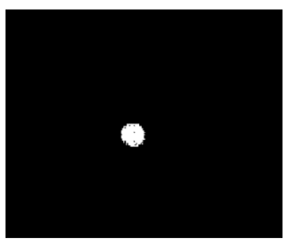
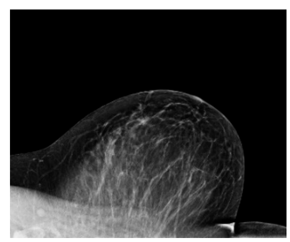
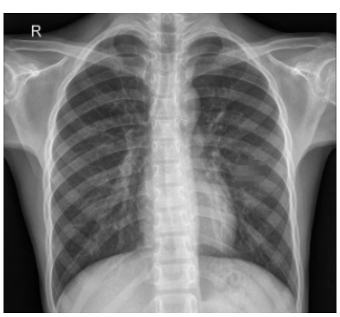
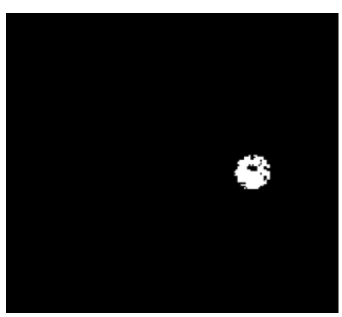
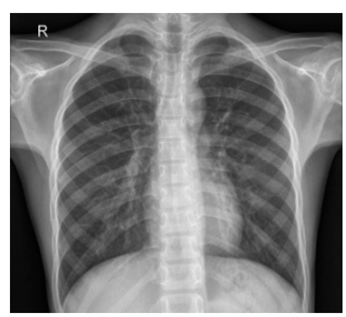
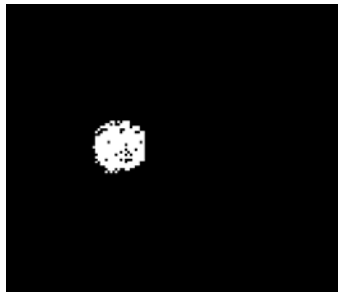
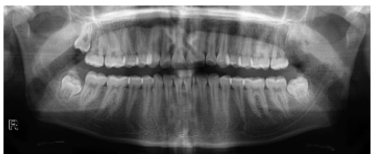
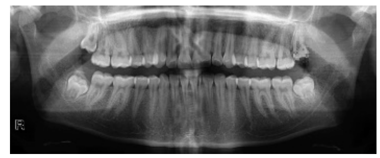
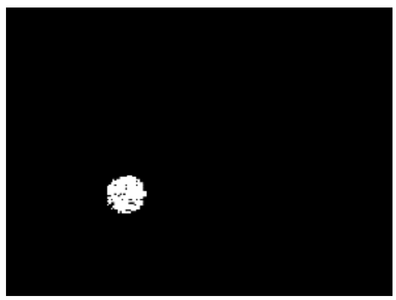
Tampered Image (Tampering rate)	**Tamper Detection (TPR / FNR / FPR)**	**Recovered Image (PSNR / SSIM)**
		
Blurring 0.94 %	0.8566 / 0.1434 / 0.0012	49.4710 dB / 0.9988
		
Unsharp masking 0.70 %	0.8642 / 0.1358 / 0.0010	45.4482 dB / 0.9955
		
Cloning 0.64 %	0.9523 / 0.0477 / 0.0006	46.8068 / 0.9985

TABLE VII. VISUAL TAMPER DETECTION AND RECOVERED IMAGE FROM THE PROPOSED SCHEME UNDER VARIOUS TAMPER ATTACKS

Tampered Image (Tampering rate)	**Tamper Detection (TPR / FNR / FPR)**	**Recovered Image (PSNR / SSIM)**
		
Mosaic 0.8 %	0.8523 / 0.1477 / 0.0014	48.8064 dB / 0.9984

		
Noise 2.01 %	0.8790 / 0.1210 / 0.0023	39.6542 dB / 0.9812
		
Removal 1.35 %	0.8409 / 0.1591 / 0.0013	42.9007 dB / 0.9935
		
Sharpening 0.91 %	0.8962 / 0.1038 / 0.0011	40.4584 / 0.9975

The watermarked images have been tampered with using different types of attacks. The proposed scheme was also evaluated by using tampering ratios of 10%, 20%, 30%, 40%, and 50% tampering rates. The experimental results show that it can improve the accuracy of tamper detection. The scheme achieved an average TPR value of 0.87 and an accuracy of 93%. In addition, the proposed scheme performed superior quality of the recovered image than the existing benchmark. The proposed scheme achieved high accuracy of the recovered image under various tamper attacks with a PSNR value of 44.79 dB and an SSIM value of 0.994.

ACKNOWLEDGMENT

This work was supported by Universiti Malaysia Pahang through the Research Grant Scheme (RDU190370).

CONFLICT OF INTEREST

On behalf of all authors, the corresponding author states that there is no conflict of interest.

REFERENCES

[1] T. Tuncer and M. Kaya, “A novel image watermarking method based on center symmetric local binary pattern with minimum distortion,” Optik (Stuttg)., vol. 185, pp. 972–984, May 2019, doi: 10.1016/j.ijleo.2019.04.038.

[2] S. I. Hisham, A. N. Muhammad, G. Badshah, N. H. Johari, and J. Mohamad Zain, “Numbering with spiral pattern to prove authenticity and integrity in medical images,” Pattern Anal. Appl. 2016 204, vol. 20, no. 4, pp. 1129–1144, May 2016, doi: 10.1007/S10044-016-0552-0.

[3] B. Bolourian Haghighi, A. H. Taherinia, and R. Monsefi, “An Effective Semi-fragile Watermarking Method for Image Authentication Based on Lifting Wavelet Transform and Feed-Forward Neural Network,” Cognit. Comput., vol. 12, no. 4, pp. 863–890, Jul. 2020, doi: 10.1007/s12559-019-09700-9.

[4] F. Ernawan, “Robust Image Watermarking Based on Psychovisual Threshold,” Journal of ICT Research and Applications, vol. 10, no. 3, pp. 228-242, 2016, doi: 10.5614/itbj.ict.res.appl.2016.10.3.3.

[5] B. Feng, X. Li, Y. Jie, C. Guo, and H. Fu, “A Novel Semi-fragile Digital Watermarking Scheme for Scrambled Image Authentication and Restoration,” Mob. Networks Appl. 2019 251, vol. 25, no. 1, pp. 82–94, Jan. 2019, doi: 10.1007/S11036-018-1186-9.

[6] H. Rhayma, A. Makhloufi, H. Hamam, and A. Ben Hamida, “Semi-fragile self-recovery watermarking scheme based on data representation through combination,” Multimed. Tools Appl., vol. 78, no. 10, pp. 14067–14089, May 2019, doi: 10.1007/s11042-019-7244-x.

[7] R. Huang, H. Liu, X. Liao, and S. Sun, “A divide-and-conquer fragile self-embedding watermarking with adaptive payload,” Multimed. Tools Appl. 2019 7818, vol. 78, no. 18, pp. 26701–26727, Sep. 2019, doi: 10.1007/s11042-019-07802-y.

[8] A. Aminuddin and F. Ernawan, “AuSR1: Authentication and self-recovery using a new image inpainting technique with LSB shifting in fragile image watermarking,” J. King Saud Univ. - Comput. Inf. Sci., Feb. 2022, doi: 10.1016/J.JKSUCI.2022.02.009.

[9] X. Yuan, X. Li, and T. Liu, “Gauss–Jordan elimination-based image tampering detection and self-recovery,” Signal Process. Image Commun., vol. 90, p. 116038, Jan. 2021, doi: 10.1016/j.image.2020.116038.

[10] D. Singh and S. K. Singh, “Block Truncation Coding based effective watermarking scheme for image authentication with recovery capability,” Multimed. Tools Appl. 2017 784, vol. 78, no. 4, pp. 4197–4215, Feb. 2019, doi: 10.1007/S11042-017-5454-7.

[11] J. Molina-Garcia, B. P. Garcia-Salgado, V. Ponomaryov, R. Reyes-Reyes, S. Sadovnychiy, and C. Cruz-Ramos, “An effective fragile

watermarking scheme for color image tampering detection and self-recovery," Signal Process. Image Commun., vol. 81, p. 115725, Feb. 2020, doi: 10.1016/j.image.2019.115725.

[12] A. Singh and M. K. Dutta, "A robust zero-watermarking scheme for tele-ophthalmological applications," J. King Saud Univ. - Comput. Inf. Sci., vol. 32, no. 8, pp. 895–908, Oct. 2020, doi: 10.1016/j.jksuci.2017.12.008.

[13] J. Molina, V. Ponomaryov, R. Reyes, S. Sadovnychiy, and C. Cruz, "Watermarking Framework for Authentication and Self-recovery of Tampered Colour Images," IEEE Lat. Am. Trans., vol. 18, no. 3, pp. 631–638, Mar. 2020, doi: 10.1109/TLA.2020.9082736.

[14] S. Prasad and A. K. Pal, "A Secure Fragile Watermarking Scheme for Protecting Integrity of Digital Images," Iran. J. Sci. Technol. - Trans. Electr. Eng., vol. 44, no. 2, pp. 703–727, Jun. 2020, doi: 10.1007/s40998-019-00275-7.

[15] W. Belferdi, A. Behloul, and L. Noui, "A Bayer pattern-based fragile watermarking scheme for color image tamper detection and restoration," Multidimens. Syst. Signal Process., vol. 30, no. 3, pp. 1093–1112, Jul. 2019, doi: 10.1007/s11045-018-0597-x.

[16] O. Hemida, Y. Huo, H. He, and F. Chen, "A restorable fragile watermarking scheme with superior localization for both natural and text images," Multimed. Tools Appl. 2018 789, vol. 78, no. 9, pp. 12373–12403, May 2019, doi: 10.1007/S11042-018-6664-3.

[17] S. Mushtaq and A. H. Mir, "Digital Image Forgeries and Passive Image Authentication Techniques: A Survey," Int. J. Adv. Sci. Technol., vol. 73, pp. 15–32, 2014, doi: 10.14257/ijast.2014.73.02.

[18] O. Hemida and H. He, "A self-recovery watermarking scheme based on block truncation coding and quantum chaos map," Multimed. Tools Appl. 2020 7925, vol. 79, no. 25–26, pp. 18695–18725, Jul. 2020, doi: 10.1007/s11042-020-08727-7.

[19] F. Tohidi, M. Paul, and M. R. Hooshmandasl, "Detection and recovery of higher tampered images using novel feature and compression strategy," IEEE Access, vol. 9, pp. 1–1, Apr. 2021, doi: 10.1109/access.2021.3072314.

[20] E. Gul and S. Ozturk, "A novel pixel-wise authentication-based self-embedding fragile watermarking method," Multimed. Syst., vol. 1, p. 3, Feb. 2021, doi: 10.1007/s00530-021-00751-3.

[21] M. Ferroukhi, A. Ouahabi, M. Attari, Y. Habchi, and A. Taleb-Ahmed, "Medical Video Coding Based on 2nd-Generation Wavelets: Performance Evaluation," Electron., vol. 8, no. 1, p. 88, Jan. 2019, doi: 10.3390/ELECTRONICS8010088.

[22] F. Ernawan, D. Ariatmanto, "Image watermarking based on integer wavelet transform-singular value decomposition with variance pixels," International Journal of Electrical and Computer Engineering, vol. 9, no.3, pp. 2185-2195, 2019, doi: 10.11591/ijece.v9i3.pp2185-2195.

[23] F. Ernawan, "Tchebichef image watermarking along the edge using YCoCg-R color space for copyright protection," International Journal of Electrical and Computer Engineering, vol. 9, no. 3, pp. 1850-1860, 2019, doi: 10.11591/ijece.v9i3.pp1850-1860.

[24] F. Ernawan, D. Ariatmanto, A Firdaus, "An Improved Image Watermarking by Modifying Selected DWT-DCT Coefficients," IEEE Access vol. 9, pp. 45474-45485, 2021, doi: 10.1109/ACCESS.2021. 3067245.

[25] A. Ray and S. Roy, "Recent trends in image watermarking techniques for copyright protection: a survey," Int. J. Multimed. Inf. Retr. vol. 9, no. 4, pp. 249–270, Oct. 2020, doi: 10.1007/S13735-020-00197-9.

[26] O. Hemida, Y. Huo, H. He, and F. Chen, "A restorable fragile watermarking scheme with superior localization for both natural and text images," Multimed. Tools Appl., vol. 78, no. 9, pp. 12373–12403, May 2019, doi: 10.1007/S11042-018-6664-3.

[27] W. Belferdi, A. Behloul, and L. Noui, "A bayer pattern-based fragile watermarking scheme for color image tamper detection and restoration," Multidimens. Syst. Signal Process., vol. 30, no. 3, pp. 1093–1112, Jul. 2019, doi: 10.1007/s11045-018-0597-x.

[28] W. Hong, J. Chen, P. S. Chang, J. Wu, T. S. Chen, and J. Lin, "A color image authentication scheme with grayscale invariance," IEEE Access, vol. 9, pp. 6522–6535, 2021, doi: 10.1109/ACCESS.2020.3047270.

[29] E. Gul and S. Ozturk, "A novel triple recovery information embedding approach for self-embedded digital image watermarking," Multimed. Tools Appl., vol. 79, no. 41–42, pp. 31239–31264, Nov. 2020, doi: 10.1007/s11042-020-09548-4.

[30] F. Ernawan, N.A. Abu and N. Suryana, "Adaptive tchebichef moment transform image compression using psychovisual model," Journal of Computer Science, vol. 9, no. 6, pp. 716-725, 2013, doi: 10.3844/jcssp.2013.716.725.

[31] F. Ernawan, N.A. Abu and N. Suryana, "An adaptive jpeg image compression using psychovisual model," Advanced Science Letters, vol. 20, no. 1, pp. 26-31, Jan. 2014, doi: 10.1166/asl.2014.5255.

[32] D Ariatmanto, F Ernawan, "An improved robust image watermarking by using different embedding strengths," Multimedia Tools and Applications, vol. 79, no. 17-18, pp. 12041 - 12067, May 2020, doi: 10.1007/s11042-019-08338-x.

[33] S. Prasad and A. K. Pal, "A secure fragile watermarking scheme for protecting integrity of digital images," Iran. J. Sci. Technol. - Trans. Electr. Eng., vol. 44, no. 2, pp. 703–727, Jun. 2020, doi: 10.1007/s40998-019-00275-7.

Assessing Node Trustworthiness through Adaptive Trust Threshold for Secure Routing in Mobile Ad Hoc Networks

M Venkata Krishna Reddy[1]
Research Scholar, Dept. of CSE
Jawaharlal Nehru Technological University
Hyderabad
Asst. Professor
Dept. of CSE
Chaitanya Bharathi Institute of Technology (A)
Hyderabad, Telangana, India

Dr. P.V.S. Srinivas[2]
Professor, Dept. of CSE, Vignana Bharathi Institute of Technology (A), Hyderabad, Telangana, India

Dr. M.Chandra Mohan[3]
Professor, Dept. of CSE, Jawaharlal Nehru Technological University, Hyderabad, Telangana, India

***Abstract*—In the field of communication, Mobile Ad-hoc networks (MANET) have become popular and widely used. However, there are many security challenges in communication through these networks due to the presence of malicious nodes. The aim of this article is to present a novel adaptive threshold trust based approach for isolating malicious nodes to establish secure routing between source and destination. Many existing cryptography methods are complex and do not properly address the elimination of malicious nodes. Several trust-dependent mechanisms have been proposed that supplement old traditional cryptography- related security schemes. But it is observed that most of these trust based approaches are using direct trust and comparing with static trust threshold. This article proposes a novel method, secured trust with adaptive threshold (STAT) that uses the Adaptive threshold technique (APTT) combined along with secure trust based approach (STBA) to evaluate the node trustworthiness for efficient routing. Secure trust for a node is calculated based upon three tier observations that includes direct, neighbor, self-historical to enrich the trust factor and adaptive trust threshold is determined based upon network parameters dynamically. Node's secure trust is compared with adaptive trust threshold computed to isolate the malicious nodes from routing. The proposed method is compared with two cases where routing is performed without any trust calculation and routing with trust calculation and compared with static trust threshold approach. Results show significant performance of the proposed work in terms of metrics like packet delivery ratio, delay, throughput, false positive detection ratio and packet drop ratio. The proposed method STAT effectively isolates the malicious nodes and establishes secure routing.**

Keywords*—Node trustworthiness; misbehaving nodes; secure trust; static threshold; adaptive threshold; secure routing

I. Introduction

MANET's are considered to be connected on an infrastructure that provides better linkage between the nodes and its environment [1]. These networks are considered as a part of many applications today [2]. However, though its wide application in many fields, MANET's are vulnerable to many attacks and especially due to its dynamic network topology. These attacks can be overcome using many schemes which are related to the identifying the malicious nodes. These schemes work on the principle that the trust values of the nodes are to be calculated. Later these trust value calculations are compared to the static threshold values known as trust threshold in order to make appropriate routing decision by isolating the malicious nodes. This threshold defines the tolerability of a node in a network [3]. The security challenges of MANET's are identified in the case of their scalability, resource utilization, dynamic topology, and even power consumption and usage. Other challenges are related to the secure environments of the networks [4].

Many trust based schemes proposed have actually made use of static thresholds to identify trust of the node [5]. This type of methods are prone to drawbacks like high error rates. These error rates will influence the timeframe of dropping malicious nodes from routing. Nodes due to environment glitches may drop the packets in some cases. They may also be categorized as malicious nodes due to static threshold strategy which is taken without any consideration of network behavior. Network behavior plays important role in MANET's due to its infrastructure less hierarchy. Existing trust based mechanisms are based upon the two tier observations either direct or combination of direct and indirect trust computations. All these trust based approaches are comparing the evolved trust with static threshold for identifying malicious nodes. As MANET's are dynamic in nature, there is always a need to compute adaptive trust threshold based on network parameters that change dynamically time to time for every node. Every node should have its trust threshold factor computed dynamically. Node's trust value should be compared with adaptive trust threshold to decide its trustworthiness. It is observed form the limitations of the existing methods, there is need of computing node's trust factor with more sophisticated approach and calculation of node's trust threshold using network parameters in adaptive mode.

The proposed work deals with the isolation of malicious nodes. Secure trust computation scheme is used to compute nodes trust value and it is combined with adaptive trust threshold technique (STAT). The work emphasis on adaptive trust threshold technique (APTT) instead of using static trust threshold, where first one is employed here along with sound research on the background on MANET and its challenges in the real-time applications. The study is proposed with a model which aims to design the threshold (adaptive in nature) of each and every node in the network to match the topology of the network. The research is also furnished with enough and appropriate mathematical model and formulas to introduce the Adaptive trust and proposed scheme. Satisfactory results are obtained such that it can be tested by implementing any routing protocol of choice.

Towards the end, the goal of this work is to address the challenge of adaptive trust threshold computation and combining it with three tier observations for generating secure trust in order to establish secure routing.

This article is organized with Introduction to the security challenges in MANET's in Section I. Background research on MANETs and related work is presented in Section II, after that the proposed model is implemented in Section III along with simulation results in Section IV. The conclusions and future directions are given in Section V.

II. Related Work

It is found from many researches usually in their proposed trust schemes fixing a threshold value which ranges between 1 and 0 to decide the fact that the node must be given access to process towards the routing phase or not [6].

Authors in [7] proposed trust computation based on user and self evidences. They evaluated the trust between the range 1 and 0. In [8], direct observations based on probability assignment between two nodes are proposed which also uses the scale of 0 and 1 for trust evaluation. Probability centric model is used for the evaluation of trust in [9] that considers the trust within the values 0-1 and uses static trust threshold concept. A scheme is presented to append the nodes trust values of nodes, according to their behaviors in [10]. A method based on reputation using the concept of polling is proposed in [11]. A local voting trust establishment strategy based on discrete scale for mobile adhoc networks is proposed in [12]. All these approaches are considering the static trust threshold commonly for all the nodes to isolate malicious nodes from routing.

This work finds its grounds on the fact that MANET's are known for their Dynamic Topology where a fixed calculation and pre-defined trust value doesn't make sense [13]. Mobility is also dependent on the behavior of the network. The evaluation of the mobility is seen in many researches which show that it can be considered as a part of the proposed model [14]. Node failures associated with the link also imposes as a threat in real-time scenarios [15]. Hence, all the time considering the static trust seems ignoring the problems that arise due to MANET behavior. This forces that an adaptive nature is to be employed.

Nodes in MANET's will move randomly time to time that leads to the raise of Node degree, a network factor. Every node in the network is linked to the fact that it is going to change for every second that can be interpreted as Rate of Link change. Average trustworthiness of the nodes in 1-hop distance is considered to compute the Adaptive threshold [16]. These are the metrics usually considered for the Topology. Hence, they do play an important role in defining the network topology as well as in secure transmission.

In the case of static methods, the threshold was based on the link as it changes in a linear fashion [17]. But in this case, every node is having its own environment which will obviously affect the link change. It is also stated by the researchers that each node might individually experience the change which is to be evaluated. These can be determined using metrics like node mobility and other parameters.

Hence, the proposed method for adaptive trust threshold computation aims towards estimating the link change at every node, node degree and average nodes trustworthiness which helps for better performance in the real-time scenarios.

III. Proposed Methdology

The proposed work computes the nodes trust based on three tier observations which are quantified into a single value to represent the secure trust factor of the node under consideration in first stage (STBA). In second phase, adaptive trust threshold is evaluated based on the network factors (APTT). In third phase, the evaluated secure trust factor is compared with adaptive trust threshold to classify the nodes category (STAT). It isolates the malicious nodes and performs secure routing only with those nodes evolved as trustworthy. The proposed approach suits to the real time environment which represents the minimum gathering in any real time scenario like a small conference. Work flow is shown the Fig. 1.

A. *Secure Trust based Approach (STBA)*

Secure Trust based approach evaluates nodes resultant secure trust value as a combination of direct, neighbor observations and nodes self-appraisal/historical trust as given in equation 1.

Resultant Secure Trust = Direct Trust + Neighbour Trust + Historical Trust/Self-appraisal of Node (1)

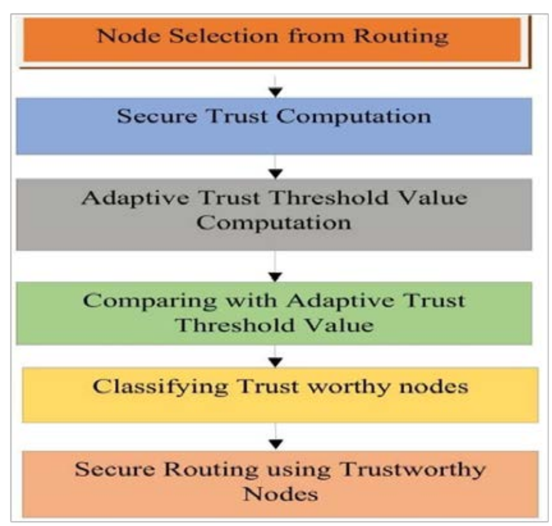


Fig. 1. Work Flow of the Objectives.

Direct Trust (D) : Data Packet Ratio(DF) and Control Packet Ratio(CF) is generated based on the data packets and control packets forwarding nature of the node. Direct Trust is calculated using equation 2.

Direct Trust, $D = t_1 * DF + t_2 * CF$ (2)

Where $t_1 + t_2 = 1$, t_1, t_2 are corresponding weight given to Data Packet Ratio and Control Packet Ratio.

Neighbour Trust (N): All the neighbouring nodes in 1 hop distance will quantify their trust observations on the node specified for which trust is calculated using the equation 3.

Neighbour Trust,

N = (k1*T1 + k2*T2 + k3*T3+ k4*T4…. +knTN) /(Total No of Nodes within 1 Hop distance) (3)

Where T1, T2, T3, T4…. Tn are trust observations by the neighbour nodes. And k1, k2, k3, k4… are corresponding weights given to the neighbour nodes based on their distance in the network.

Nodes Self-Appraisal (H): Nodes can rate themselves with a trust value based upon their performance in terms of packet forwading. Self-Appraisal is quantified based on equation 4.

Self-Appraisal, H = No of Packets forwarded properly/Total no of packets received (4)

Then,

Secure Trust is evaluated based on equation 5.

Secure Trust (STBA), T = m1D + m2N + m3H (5)

Where m1, m2, m3 are corresponding weights assigned to the trust observations.

B. Adaptive Trust Threshold (APTT)

Adaptive trust threshold computation has to represent many different network factors. Each node in the network may encounter various conditions such as node degree variations, rate of link changes and average neighborhood.

Network Parameters to be considered are:

1) Node Degree - σ

It is defined as No of nodes in 1 hop neighbourhood. The statics are presented as follows.

Node n at time t, n(t) = 0, i.e.

Min Node Degree = 0, then it is considered as no neighbours for that node.

Max Node Degree = all are directly connected to Node µ. Node Degree has a direct impact on Trust Threshold; the higher number of nodes in its 1-hop neighbourhood. The higher is the threshold value and vice-versa. Optimal Threshold Value for the Node Degree is calculated based on the equation 6.

Optimal Threshold value

$$\Sigma\,\sigma = \sigma n \,/\, |\,T\,| \quad (6)$$

Where, T = total number of nodes

σn = Node Degree of node $'n'$

and 2 Hop connectivity may be considered.

2) Rate of Link Changes – η

Neighborhood changes occur in MANETs frequently due to Network Mobility. A Node can determine its Neighbor Mobility by computing the Neighborhoods rate of link changes. Higher Mobility leads to higher rate of link changes in Nodes Neighborhood.

Rate of Link changes at Node µ is given by equation 7.

$$\eta\,\mu = \lambda\mu + \mu\mu \quad (7)$$

Where,

λ_μ =Number of new nodes coming in, means Neighbours of Node

µ = Total Link arrival rate at Node µ

μ_μ = Total number of nodes moving out of Node's µ transmission range for time interval.

Minimum Link Rate changes, $\eta_{\mu\ min}$ = 0 = No new nodes arrival, No Link Breakages, implies Temporary Static then considered as no mobility.

Maximum Link Rate changes, $\eta_{\mu\ max}$ = When all the direct neighbors are out of the transmission zone, considered as High mobility.

If the rate of change in Neighborhood is high, set Low Threshold (to avoid false positives).

If the rate of change in Neighborhood is low, set High Threshold (Network is static).

Optimal Threshold value for the Rate of link change factor is given by equation 8.

Optimal Threshold

$$\mathcal{E}\eta = 1 - \eta\,\mu\,/2\sigma\mu \text{ [22]} \quad (8)$$

Where $\eta\,\mu$ = Rate of Link Changes of node 'µ '

$\sigma\mu$ = Node Degree of node 'µ '

3) Average Neighborhood Trustworthiness – $\boldsymbol{\tau_{avg}}$

The formula for τ_{avg} is given by equation 9.

$$\tau_{\mu\ avg} = 1/n \sum_{j=1}^{n} Tj \quad (9)$$

Where, Tj = Trust of all the neighbour nodes of Node µ on it, where Self-appraisal/Historical Trust of nodes is taken into consideration.

$\tau_{\mu\ avg}$ = 1: Good Nodes are available, High Trust Worthy , set High Threshold.

$\tau_{\mu\ avg}$ = 0: More Misbehaving Nodes are available, Low Trust Worthiness, set Low Threshold value.

Optimal Threshold value at node μ for malicious node isolation given be equation 10.

$$\xi_T = \tau_{\mu\, avg} \tag{10}$$

Combining all the network parameters for estimating the proposed adaptive trust threshold using equation 11.

Final adaptive trust threshold, ξ_μ is

$$\xi_\mu = (a\,\xi_\sigma + b\xi_\eta + c\,\xi_T)\,/(a+b+c) \tag{11}$$

Where ξ_σ = Optimal threshold value of Node Degree of the

Node

ξ_η = Optimal threshold value of Rate of Link.

Changes of the Node.

ξ_T = Optimal threshold value of Average.

Neighbourhood Trustworthiness of the Node.

And a,b,c are constants and a+b+c is considered for higher throughput, and the Node Degree and 2 Hop connectivity are given importance. So subsequently α should have more weight.

Then Decision of isolating a node depends upon.

T, Trust Evaluated of the node >= Adaptive Trust Threshold - ξ_{μ}, Node is decided as Trusted Node.

T, Trust Evaluated < Adaptive Trust Threshold - ξ_{μ}, Node is decided as Un Trusted Node, then, isolate the node from routing.

C. Proposed Algorithm for Adaptive Trust Threshold

Considering the 2-hop connectivity an algorithm is proposed for the adaptive trust threshold as follows:

Algorithm Optimal adaptive trust threshold computation

procedure Secure Trust(T, D, N, H)
{
// T = Secure trust of the Node
// D= Direct Trust of the Node
// N=Neighbor Trust of the Node
// H=Historical Trust/Self Appraisal of the Node
Step:1 Node trustworthiness is initiated, for each and every node
Step:2 Data Packet Forward ratio is
$DF = w_1 *(D_{forw}/D_{td}) + w_2 *(D_{drop}/D_{td}) + w_3 *(D_{mr}/D_{td}) + w_4 *(D_{fi}/D_{td})$
Step:3 Control Packet Forward Ratio is
$CF = w_1 *(R_{req}/R_{treq}) + w_2 *(R_{rep}/R_{trep}) + w_3 *(R_{err}/R_{terr}) + w_4 *(R_{tack}/R_{ack})$
Step:4 Driect Trust, $D = t_1 * DF + t_2 * CF$
Step:5 Neighbour Trust, N = (k1*T1 + k2*T2 + k3*T3+ k4*T4…. +knTN) /(Total No of Nodes within 1 Hop distance)
Step:6 Self-Appraisal, H = No of Packets forwarded properly/Total no of packets received
Step:7 Secure Trust(STBA), T = m1D + m2N + m3H
}
end procedure

procedure Network Parameters (T_{avg}, $\Sigma\sigma$, $\xi\eta$)
{
//T_{avg} = Average Neighborhood Trustworthiness
// $\Sigma\sigma$ = Optimal Threshold value of Node Degree
//$\xi\eta$= Optimal Threshold value of Rate of Change in Linkage
Step:1 if new node appears then $\tau_{avg} = \tau_{avg}$+ Threshold of new node.
Step:2 Calculating the optimal threshold through node degree using $\Sigma\sigma = \sigma n / |T|$
Step:3 Based on the mobility of the network calculate the total Change of rate of linkage using $\xi\eta = 1 - \eta\mu / 2\sigma\mu$
Step:4 if $\tau_{\mu\, avg}$=0 then consider it as Malicious node.
Step:5 The threshold for malicious node is calculated $\xi_T = \tau_{\mu\, avg}$.
}
end procedure

procedure Adaptive Threshold (ξ_μ, ξ_σ, ξ_η, ξ_T)
{
// ξ_μ = Adaptive Trust Threshold
//ξ_σ = Optimal Trust Threshold of Node Degree
// ξ_η = Optimal Trust Threshold of Rate of Link Changes
//ξ_T = Optimal Trust Threshold of Average Node Trustworthiness
Step:1 Overall Adaptive Trust Threshold is calculated using $\xi_\mu = (a\,\xi_\sigma + b\xi_\eta + c\,\xi_T)\,/(a+b+c)$
}
end procedure

procedure Routing Decision (T_μ, ξ_μ)
{
// T_μ = Secure trust of the Node μ
// ξ_μ = Adaptive Trust Threshold of the Node μ
if ($T_\mu < \xi_\mu$) then malicious node, isolate the node
else
Trustworthy node, Involve in routing process
end if
}
end procedure

IV. Experiment and Result Analysis

Network Simulator 2 (NS2) is used for simulation of desired network. The network traffic is maintained with a size of 512Bytes with a packet rate of 200 and 100 packets per second. The malicious Nodes are defined in the physical layers. Hence, considering the parameters for configuration, the trust and other metrics are calculated. To analyze the results, the network configuration parameters and simulation parameters are given in Table I.

TABLE I. SIMULATION PARAMETERS

Simulation tool	NS2
Number of Nodes	100
Malicious Nodes	18
Propagation Model	Two ray ground
Malicious Nodes Declaration	0t
Topography	700*500(M)
Simulation Time	500s
Mobility(r)	5m/s

The simulations are carried out for three design goals in which the last scenario is the proposed method where calculations are obtained from the Adaptive trust threshold.

The parameters used to evaluate the results are Packet Delivery Ratio (PDR): It is defined to be the ratio denoting the number of packets received at the destination and the number of packets sent from the source [18].

Packet Drop Rate (PPR): It is defined as a ratio of the number of lost packets to the total number of sent packets [19].

False Positive Detection (FPR): It denotes the ratio the count of good nodes wrongly identified as malicious to the total available count of nodes. It is also used to calculate the FPR (False positive rate) [20, 21].

Malicious Node Detection Ratio (MDR): it gives the ratio of malicious/misbehaving nodes from the total nodes from the network [20, 21].

Throughput (T): It usually defines the amount of data transferability of a network through a period of time [20,21].

Delay (D): This shows the time frame from delivering the packets through source and destination. [22].

These parameters are considered to evaluate the performance of three occurrences, where the first occurrence routing without any prior trust calculation.

The second occurrence considers the routing with nodes trust computation using the methodology Secure trust based approach (STBA) but compared with static trust threshold factor for node isolation.

Third occurrence is the proposed work, routing with Secure trust combined with Adaptive Trust threshold (STAT) which is the combination of Computation of Secure trust (STBA) and Computation of Adaptive trust threshold (APTT). More emphasis is on computation and comparison of Adaptive trust threshold. Results proved the proposed scheme with adaptive trust threshold comparison performs well over other two occurrences mentioned.

A. *Result and Analysis*

1) Secure Trust Based Approach Computation (STBA): Secure Trust is computed based upon the three tier observations using the above mentioned equations. Results obtained for secure trust are tabulated in Table II.

2) Adaptive Trust Threshold Computation (APTT): Adaptive Trust Threshold is computed based upon network parameters Node Degreeσ, Rate of Link changes ŋ, Average Node Behavior τ_{avg} using above mentioned equations. Results generated are tabulated in Table III. Identification of malicious nodes and their isolation using proposed STAT method with Node's Secure Trust computation and comparison with Adaptive Trust Threshold is shown in Table IV.

TABLE II. SECURE TRUST THRESHOLD COMPUTATION

Node	**Direct Trust**	**Neighbor calculation**	**Historical Trust Calculation**	**Node Secure Trust**
0	0.93	0.3417	0.94	0.75451
1	0.75	0.2788	1	0.63364
2	0	0.3792	0.78	0.11376
3	0	0.5238	0.87	0.15714
4	0	0.5992	0.78	0.17976
5	0	0.2955	0.67	0.08865
6	0	0.03075	0.76	0.009225
7	0.03	0.624	0.87	0.2052
8	0.69	0.4672	0.91	0.55416
9	0.83	0.04975	0.995	0.612425
10	0.72	0.414	1	0.6562
11	0.49	0.4096	0.79	0.81688
12	0.47	0.4636	0.86	0.82108
13	0.67	0.2895	0.83	0.88885
14	0.31	0.179	0.85	0.4397

TABLE III. ADAPTIVE TRUST THRESHOLD ξ_{μ} COMPUTATION

Node	#1hop Neighbor -	Node Degree Value σ	Rate of Link Changes	2sigma	d/e	Link change η	Average Neighborhood Trustworthiness - τ_{avg}	Adaptive trust Threshold - ξ_{μ}
0	12	0.5714286	2	24	0.083333	0.916667	0.4792	0.665777
1	11	0.5238095	5	22	0.227273	0.772727	0.4822	0.594324
2	13	0.6190476	4	26	0.153846	0.846154	0.4215	0.667425
3	11	0.5238095	4	22	0.181818	0.818182	0.5268	0.61242
4	13	0.6190476	3	26	0.115385	0.884615	0.5914	0.695953
5	11	0.5238095	5	22	0.227273	0.772727	0.4866	0.594764
6	14	0.6666667	8	28	0.285714	0.714286	0.5489	0.669176
7	15	0.7142857	3	30	0.1	0.9	0.5578	0.754351
8	10	0.4761905	6	20	0.3	0.7	0.5422	0.549934
9	10	0.4761905	3	20	0.15	0.85	0.4911	0.589824
10	11	0.5238095	5	22	0.227273	0.772727	0.4474	0.590844
11	15	0.7142857	3	30	0.1	0.9	0.5781	0.756381
12	15	0.7142857	2	30	0.066667	0.933333	0.5512	0.763691
13	14	0.6666667	2	28	0.071429	0.928571	0.4719	0.725761
14	3	0.1428571	1	6	0.166667	0.833333	0.5144	0.387154

TABLE IV. MALICIOUS NODE ISOLATION

Node	Node Secure Trust	Adaptive trust Threshold - ξ_{μ}	Decision
0	0.75451	0.665777	Trustworthy
1	0.63364	0.594324	Trustworthy
2	0.11376	0.667425	Malicious
3	0.15714	0.61242	Malicious
4	0.17976	0.695953	Malicious
5	0.08865	0.594764	Trustworthy
6	0.009225	0.669176	Malicious
7	0.2052	0.754351	Malicious
8	0.55416	0.549934	Trustworthy
9	0.612425	0.589824	Trustworthy
10	0.6562	0.590844	Trustworthy
11	0.81688	0.756381	Trustworthy
12	0.82108	0.763691	Trustworthy
13	0.88885	0.725761	Trustworthy
14	0.4397	0.387154	Trustworthy

B. Performance Metrics

1) Packet delivery ratio: It was observed that for 100pkts/s, out of 50000 packets sent, 47550 packets received, Packet Delivery Ratio is 95.1% for the proposed method, 91.1 % for the second case where trust is compared with static trust threshold, 53.2 % in case of third design goal where routing involved without trust calculation and for 200pkts/s**,** out of 100000 packets sent, 79235 packets received, Packet Delivery Ratio is 79.2%. In case of proposed method, 75.3%, 30.1%. In case of second and third case, respectively whereas in Fig. 2, shows the Packet Delivery ratio of all the three design goals.

C. Packet Drop Ratio

From the simulation, it was observed that for 100pkts/s, out of 50000 packets sent, 2451 packets lost, Packet Drop Ratio is 4.9% for the proposed method, 5.8 % for the second case where trust is compared with static trust threshold, 43.3 % in case of third design goal where routing involved without trust calculation and for 200pkts/s**,** out of 100000 packets sent, 20785 packets lost, Packet Drop Ratio is 20.765%. In case of proposed method, whereas it is 23.4%, 69.2% for second and third design goal. Fig. 3 shows the Packet drop ratio for the three cases compared.

D. Throughput

From the simulation, it was observed that for 100pkts/s, Throughput is 380.4kbps for the proposed method, 358.2kbps for the second case where trust is compared with static trust threshold, 211.6kbps in case of third design goal where routing involved without trust calculation and for 200pkts/s**,** Throughput is 633.4kbps in case of proposed method, whereas it is 603.5kbps, 229.2kbps for second and third design goal. In Fig. 4, throughput efficiency of the proposed method STAT is shown.

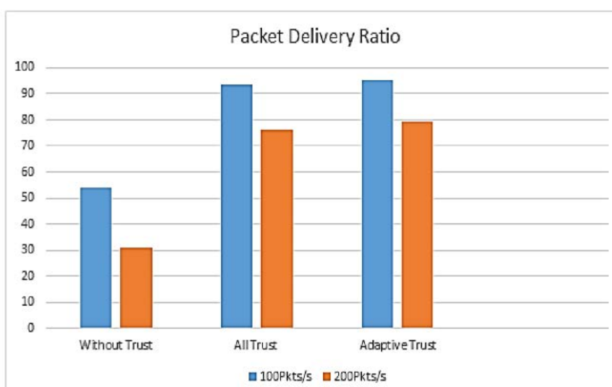


Fig. 2. Packet Delivery Ratio Analysis.

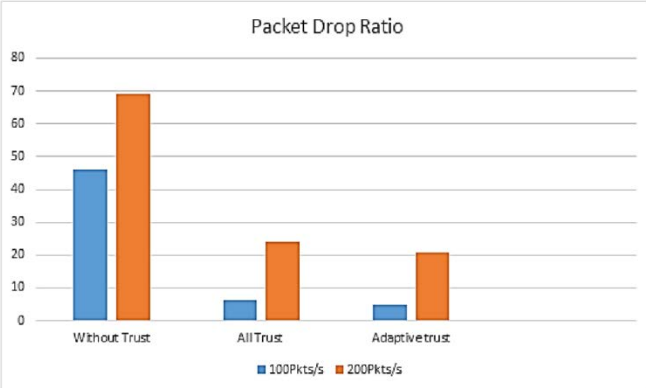


Fig. 3. Packet Drop Ratio Analysis.

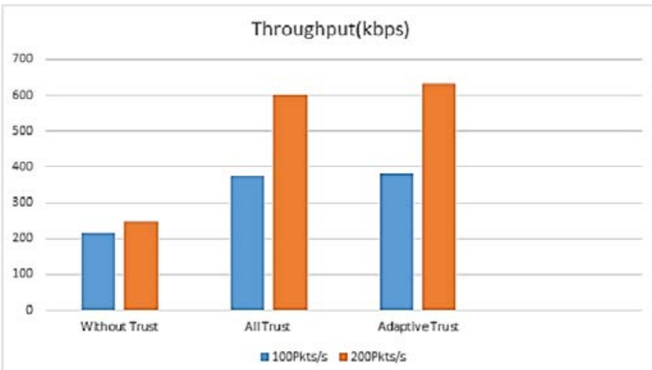


Fig. 4. Analysis of Throughput.

E. Delay

Delay is observed as 187ms for 100pkts/s in case of proposed method, 198ms for the second case where trust is compared with static trust threshold, 232ms in case of third design goal where routing involved without trust calculation and for 200pkts/s**,** Delay is 270ms in case of proposed method, whereas it is 287ms, 302ms for second and third design goal. Fig. 5 depicts the delay parameter in case of three scenarios mentioned and proves the efficacy of the proposed method. Delay in the network in case of packet delivery is illustrated as follows.

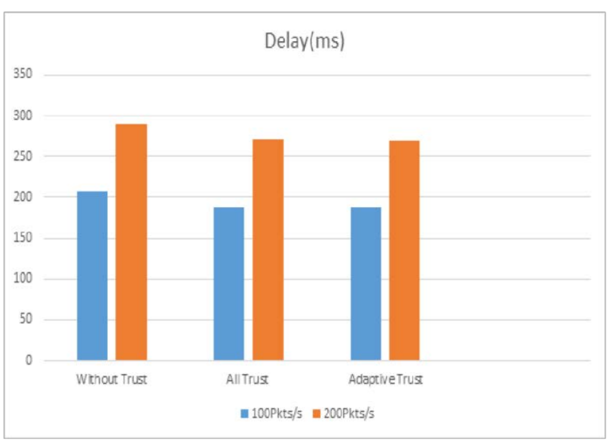


Fig. 5. Delay in the Network.

F. False Positive Detection Rate

False Positive Detection Rate is found as 52% in case of proposed method, 44% for the second case where trust is compared with static trust threshold. In Fig. 6, False Positive Detection Rate is shown for the proposed method.

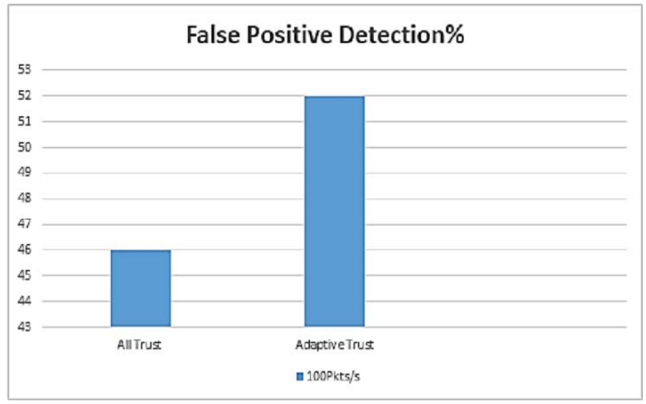


Fig. 6. False Positive Detection Ratio Analysis.

G. Malicious node Detection Rate

Malicious node Detection Rate is found as 23% in case of proposed method, 21% for the second case where trust is compared with static trust threshold. In Fig. 7, proposed method is performing better compared with the secure trust (static threshold) in terms of malicious node detection rate.

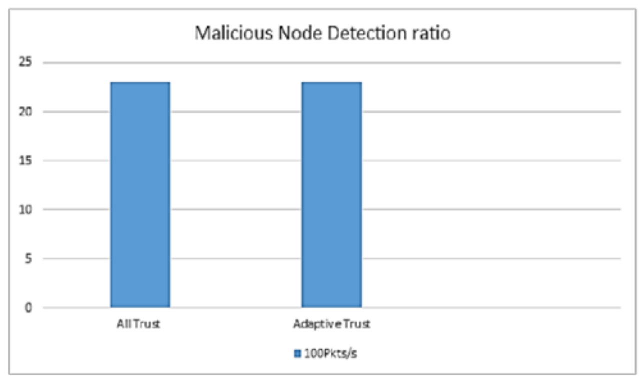


Fig. 7. Detection of Malicious Nodes.

From results, it is interpreted that computation of adaptive trust threshold based on the network parameters when combined with secure trust mechanism, efficiently identifies the malicious nodes which are dropping packets and good natured nodes which are delivering packets and can be used for mobilization.

V. CONCLUSION

The work presented here shows a strategy which can be used to detect the nodes which are misbehaving in a network by considering the network parameters which plays an important role in network. The results evaluated and shown in Fig. 2 to 7 prove the efficacy of the proposed work STAT. The Packet delivery ratio for the method proposed (Adaptive trust Threshold) shows significant growth along with less packet loss ratio as well which makes it easy to consider the technique proposed in the real-time traffic networks. The proposed method is evaluated and seemed better performing than other methods in particular with methods that uses static trust threshold. Results proves the efficiency of the proposed method when compared with other approaches like routing without trust calculation and routing with trust computation and static threshold approaches. The further scope of the work will be extended by considering the Power consumption scenarios in the networks in case of trustworthy nodes.

Acknowledgment

The authors acknowledge the support and cooperation rendered by all the members directly and indirectly.

References

[1] B.V.S Uma Prathyusha, K.Ramesh Babu, "A Node Monitoring Agent based Handover Mechanism for Effective Communication in Cloud-Assisted MANETs in 5G", International Journal of Advanced Computer Science and Applications(2022), Vol. 13, No. 1, 2022, 128-136.

[2] Ahmed, Malik N., et al. "F3TM: Flooding Factor based Trust Management Framework for secure data transmission in MANETs." *Journal of King Saud University-Computer and Information Sciences* 29.3 (2017): 269-280.

[3] AlKhatieb, Anas, Emad Felemban, and Atif Naseer. "Performance evaluation of ad-hoc routing protocols in (FANETs)." *2020 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*. IEEE, 2020.

[4] H C Ramaprasad , S.C. Lingareddy , "A Novel Integrated Scheme for Detection and Mitigation of Route Diversion Attack in MANET", International Journal of Advanced Computer Science and Applications(2022), Vol. 12, No. 11, 2021, 374-381.

[5] Chakraborty, Arpita, Jyoti Sekhar Banerjee, and Abir Chattopadhyay. "Malicious node restricted quantized data fusion scheme for trustworthy spectrum sensing in cognitive radio networks." *Journal of mechanics of continua and mathematical sciences* 15.1 (2020): 39-56.

[6] Jain, Ashish Kumar, Vrinda Tokekar, and Shailendra Shrivastava. "Security enhancement in MANETs using fuzzy-based trust computation against black hole attacks." *Information and Communication Technology*. Springer, Singapore, 2018. 39-47.

[7] Saidi, Ahmed. "Trust evaluation method for Wireless Sensor Networks based on behavioral similarity and similarity coefficient." *2021 International Conference on Networking and Advanced Systems (ICNAS)*. IEEE, 2021.

[8] Khan, Burhan Ul Islam, et al. "A survey on MANETs: architecture, evolution, applications, security issues and solutions." *Indonesian Journal of Electrical Engineering and Computer Science* 12.2 (2018): 832-842.

[9] Zhang, De-gan, et al. "Novel approach of distributed & adaptive trust metrics for MANET." *Wireless Networks* 25.6 (2019): 3587-3603.

[10] Daly, Elizabeth M., and Mads Haahr. "The challenges of disconnected delay-tolerant MANETs." *Ad Hoc Networks* 8.2 (2010): 241-250.

[11] Abdel-Fattah, Farhan, et al. "Security challenges and attacks in dynamic mobile ad hoc networks MANETs." *2019 IEEE jordan international joint conference on electrical engineering and information technology (JEEIT)*. IEEE, 2019.

[12] G. A. S. T. ME, and M. Manikandan. "Trust threshold-based neighborhood-Trustworthy with node certified public key Management scheme in MANET." pp. 28-33. International Journal of Science and Technology, 2019.

[13] Divyashree, H. B., C. Puttamadappa, and KS Nandini Prasad. "Performance Analysis and Enhancement of QoS Parameters for Real-Time Applications in MANETs-Comparative Study." *2020 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)*. IEEE, 2020.

[14] Jhaveri, Rutvij H., et al. "A composite trust model for secure routing in mobile ad-hoc networks." *Adhoc Networks* 2 (2017): 19-45.

[15] Koul, Ajay, and Harinder Kaur. "Quality of Service Oriented Secure Routing Model for Mobile Ad hoc Networks." *Proceedings of the 2017 International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence*. 2017.

[16] M. Ebrahimi, N. , N. Ho, T. Nguyen, and J. Stolmeier. "Evaluation of parameters affecting the performance of routing protocols in mobile ad hoc networks (MANETs) with a focus on energy efficiency." In Future of information and communication conference, pp. 1210-1219. Springer, Cham, 2019.

[17] Saudi, Nur Amirah Mohd, et al. "Mobile ad-hoc network (MANET) routing protocols: A performance assessment." *Proceedings of the third international conference on computing, mathematics and statistics (iCMS2017)*. Springer, Singapore, 2019.

[18] Tamilselvi, P., and C. Ganesh Babu. "An efficient approach to circumvent black hole nodes in manets." *Cluster Computing* 22.5 (2019): 11401-11409.

[19] Anwar, Raja Waseem, et al. "BTEM: Belief based trust evaluation mechanism for wireless sensor networks." *Future generation computer systems* 96 (2019): 605-616.

[20] Sethuraman, Priya, and N. Kannan. "Refined trust energy-ad hoc on demand distance vector (ReTE-AODV) routing algorithm for secured routing in MANET." *Wireless Networks* 23.7 (2017): 2227-2237.

[21] Oubabas, Sarah, et al. "Secure and stable vehicular ad hoc network clustering algorithm based on hybrid mobility similarities and trust management scheme." *Vehicular Communications* 13 (2018): 128-138.

[22] Patel, Surabhi, and Heman Pathak. "A mathematical framework for link failure time estimation in MANETs." *Engineering Science and Technology, an International Journal* 25 (2022): 100984.

An Ontology Model for Medical Tourism Supply Chain Knowledge Representation

Worawit Janchai[1], Abdelaziz Bouras[2], Veeraporn Siddoo[3]*
College of Art, Media and Technology, Chiang Mai University, Chiang Mai, Thailand[1]
Department of Computer Science and Engineering, College of Engineering, Qatar University, Qatar[2]
College of Computing, Prince of Songkla University, Phuket, Thailand[3]

***Abstract*—This study developed an application ontology related to the medical tourism supply chain domain (MTSC). The motivation for developing an ontology is that current MTSC studies use a descriptive approach to provide knowledge, which is difficult to comprehend and apply. The formalization of MTSC knowledge can provide medical tourism stakeholders with a shared understanding of the medical tourism business. As a result, the MTSC domain requires efficient semantic knowledge representation. Ontology is a popular approach for integrating knowledge and comprehension because it presents schema and knowledge base in an accurate and relevant feature. This paper employed the ontology engineering methodology, which included specification, conceptualization, and implementation steps. The conceptual model and facets of the MTSC are proposed. The MTSC objective and scope are tested with semantic competency questions against SPARQL Query formulations. The ontology metrics evaluation was used to verify the ontology quality including the external validation done by the domain experts. The results showed that the MTSC ontology has an appropriate schema design, terminologies, and query results.**

Keywords—Medical tourism; application ontology; supply chain; ontology development; knowledge representation

I. Introduction

The medical tourism (MT) industry is considered a high-value business that can generate a high income for the service providing country [1]. MT is a part of health tourism that focuses on medical intervention to improve or restore an individual's health [2]. MT is a specialized service and differs from general tourism in that medical tourists often have a significant problem and need of medical service, along with the desire to travel as well [3]. The medical tourism supply chain (MTSC) is distinct from both the general and health tourism supply chains. MTSC emphasizes medical and travel intervention, as well as relationships with connected parties who assist the major service providers in both the medical and tourism sectors.

At present, the MT business knowledge tends to relate to the key player side [4]. MT business activities are fragmented between stakeholders [5, 6]. Activities are short-lived interactions between a medical tourist and the providers. As a result, some MT actors are deficient in knowledge of the supply chain (SC). Since the MT business is being challenged by many factors such as the global situation, consumer demand, and business competition [7, 8], a new paradigm for enabling MT success is to integrate intra- and inter-organizational processes [9]. Understanding the MTSC is therefore important.

One method for developing concrete knowledge is to use ontology [10, 11]. Ontology supports aligning the understanding of knowledge by presenting it in an explicit specification and in a meaningful manner [11]. Ontology links the information relationships to a semantic structure. For over a decade now, organizations have used ontologies to build mutual comprehension in the SC [12, 13]. The ontology ideas emphasize making the practitioners understand business logistics, roles, and activities in the SC. We can see studies where ontologies are used for describing SC models, such as [14, 15, 16]. However, an ontology for encouraging MT practitioners to understand the MT business has not been reported. Previous MT studies have emphasized descriptive explanations [4, 6, 17, 18, 19]. At the same time, there is no mention of MT in the SC ontology. Numerous studies have avoided establishing semantic knowledge as a support to MTSC model. Practitioners in MT still lack explicit knowledge that would help them better understand the industry.

Current studies lack a shared understanding of MTSC model between practitioners. Therefore, strengthening business comprehension is of great importance to the business participation and the SC management. This study contributes to the visual linkage of MTSC model.

The paper's key contribution is the MTSC ontology, which was built for the MT domain and enhanced by domain experts. This study developed semantic knowledge utilized in the MT context, to represent MTSC knowledge by using ontology-based tools. The MTSC conceptual model and facets are proposed based on literature extraction and domain expert knowledge. The ontology engineering methodology is presented, adopting ontology engineering methodology of [10, 20]. The instances used for implementation are from Thailand. The evaluation of ontology is performed using the semantic competency questions and the well-known metric-based ontology quality analysis [21, 22]. Domain specialists in the MT industry assisted in the enumeration of terminology and the validation of ontology quality during the development of the ontology.

II. Literature Review

The paper includes MTSC and ontology engineering methodology terms. The literature review discusses the MTSC characteristics and processes followed by a discussion of

*Corresponding Author.

ontology engineering methodology and ontology evaluation. Then the previous ontology works are examined for MTSC ontology development.

A. Medical Tourism Supply Chain

SC is a term used for over a decade. The SC was originally associated with the manufacturing industry, from raw material procurement to manufacture, delivery, and distribution to clients or customers [23]. The system is interconnected and requires performance management [24]. The expected outcome of SC management is improved quality of operations and products that promote profits for the organization [25]. The actors involved in the SC cycle are grouped to suppliers, producers, transporters, distributors, and customers [14]. At each stage of the SC system, products are produced and transmitted depending on business objectives.

Subsequently, the SC system was introduced in the service business. MT industry has been examined using this concept [26]. MTSC is a network of individuals and organizations who handle information, product or service, and financial flows [27]. The objective of MTSC is to manage the healthcare and tourism processes for improving product and service value for all actors and for creating profits to the organization [26]. The two primary industries that MT is associated with are health and tourism [28]. The MTSC involves a process flow from beginning to end that encompasses the transfer of products or services from providers to customers (internal and external) [6, 19].

The MTSC processes commence with the marketing process to attract medical tourists followed by medical tourists' decision-making regarding service, then making the necessary travel arrangements, followed by diagnostic activity and travel to the destination for medical treatment [6]. The subsequent recovery care is the main activity, involving the rechecking of the treatment result and assessing health risks. During the process of medical services in the destination country, medical tourists can participate in tourism activities [6, 18]. At the end of the process, the relevant activities are tourist departure, tracking potential symptoms, and follow-up activities at the country of origin [27, 29]. The actors in MT business are those who participate in the MTSC processes [4, 18]. Actors come from diverse industries and have various business objectives and roles including activities to be carried out inside and outside the organization [12, 15, 28]. The goods or services are produced by actors and delivered to others.

B. Ontology Engineering Methodology

Ontology engineering methodology is a method for aligning understanding of knowledge by presenting it in an explicit specification and in a meaningful manner [11]. Ontology links the information relationships to a semantic structure. Organizations use ontology to build mutual comprehension in the general business and in context-specific domains.

The goal of ontology development is to express knowledge in such a way that both people and computers can comprehend it [22]. Developing an ontology usually begins with recognizing the need for a certain purpose [20]. Then an ontology is created. Many methodologies can be used to develop an ontology. For example, the method of [22] designing and evaluating an ontology combined six steps: scenario motivation, informal competency questions, first-order logic: terminology, formal competency questions, first-order logic: axioms, and completeness theorems; methodology of [20] relating to the knowledge process consisted of five steps: feasibility study, ontology kickoff, refinement, evaluation, and maintenance; the method of [10] consists of seven steps: determining the ontology domain and scope, considering and reusing existing ontologies, enumerating ontology terms, defining the classes and hierarchy, defining class properties, defining the facets, and creating instances; the method of [30] combines development and support processes that can run in parallel with all of the development phases.

We can see that [22] focused on ontology development, whereas [20] detailed all steps in the process of maintaining. Identically [30] enumerated the entire process, but the steps were complex. On the other hand, [10] proposed the clear steps of ontology creation than other studies considered although they did not have much detail relating to the ontology testing process.

Taking into account the ontology development and quality testing objectives, this study employs a hybrid waterfall methodology by combining waterfall methodologies from [10, 20]. The creation of an ontology involves three major steps.

1) Specification The purpose is to gather informal knowledge about the domain. The specification step develops the conceptual taxonomy of MTSC ontology.

2) Conceptualization The purpose is to organize and structure the information into semantic framework. The conceptualization step defines the classes, objects and instances of ontology.

3) Implementation The purpose is to implement, verify and validate the ontology. The implementation step creates a concrete ontology with an ontology implementation tool and validates the ontology quality.

C. Ontology Evaluation

The ontology evaluation indicates the quality of the ontology. This study defines the quality as ontology should be validated by robust approaches in technical aspects and domain experts' involvement. The methodology used in this paper adheres to the semantic competency questions method of [22] and the ontology metrics proposed by [21].

The semantic competency questions technique [22] generates answers to competency questions. The ontology concept is tested against SPARQL Query formulations. Domain experts will involve in the evaluation process by verifying competency questions and answers.

The ontology metrics [21] combine three evaluation aspects.

1) Ontology schema metrics assess ontology design and its capabilities for representing rich knowledge. The proposed metrics are as following:

Relationship richness (RR) relates to an ontology having many interrelated relationships. RR is a percentage representing the number of connections between classes with multiple relationships. If the percentage closes to one, it demonstrates ontology has various connections other than class-subclass. On the other hand, if RR value closes to zero then ontology does not have many relationships.

Attribute richness (AR) relates to an ontology having a sufficient amount of knowledge. AR is a number representing the amount of knowledge that ontology conveyed. The high AR means ontology has a high average attribute or knowledge per class. On the other hand, the lower AR means ontology has less knowledge.

Inheritance richness of schema (IRs) relates to an ontology having a wide knowledge distribution. IRs is a number representing the ontology's ability to arrange knowledge into multiple categories and subcategories. The low IRs means ontology is a very detailed knowledge, while the high IRs reflects a broad spectrum of generic knowledge.

2) Knowledgebase metrics assess the amount of knowledge representation and the design efficacy of the ontology. The proposed metrics include:

Class richness (CR) relates to an ontology having a class with knowledge. CR is a percentage representing the use of class with instances. The high CR indicates ontology has rich knowledge in class. By contrast the low CR reflects low knowledge.

Average population (P) relates to an ontology having described the schema's whole knowledge adequately. P is a number representing the average knowledge per class. The high P shows sufficient knowledge of schema, while the low P means insufficient knowledge representation.

Cohesion (Coh) relates to an ontology having relationships among nodes and edges. Coh is a number representing the isolated components. A high Coh indicates that the ontology knowledgebase has lost its connection. If the Coh value is small, it means that knowledge is connected.

3) Class metrics investigate the knowledge distribution of all ontology classes. The proposed metrics include:

Importance (Imp) concerns the instance distribution in each class. Imp value shows which classes have enough instances and which classes have fewer instances.

Connectivity (Cn) concerns the relationship of the instance in one class over another class. Cn number refers to how well instances are linked together.

Readability (Rd) concerns whether the ontology contains human-readable descriptions. Rd is calculated from comments, labels, or captions of ontology.

D. Medical Tourism Supply Chain Ontologies

MTSC ontology is categorized as application ontology because the terminology depends on both specific tasks that are MT processes and the domain that encompasses health and tourism [31]. Ontology has been proven to help practitioners clarify SC models including fragmented processes [32]. However, during this research, no ontology for building MTSC comprehension was discovered. Therefore, the common SC ontologies based on SC definition given by [23] and ontologies related to health tourism domain were reviewed.

Üreten and Đlter [14] present SCM ontology derived from the common concept of SC. There are four main classes: Agent, Product, Operations, and Flow. The Agent subclasses include supplier, producer, transporter, distributor, and customer. Agent has data properties that are production activity type, utility created, size, geographic location, service content, and make-buy decisions. The Product is identified to two subclasses: Goods, Services. Goods have data properties: process type, weight, size, resistance, durability, characteristics, and product type; while Service has service content, property, location, and degree of customer contact. The Operation subclasses are Plan, Source, Make, Deliver, and Return. There are sub-activities in a given operation. The Flow has three subclasses: Material, Money, and Information, with flow direction and flow between properties. There are relationships between the common classes. SCM ontology shows that Agent and Product have a link with produces object property. Also, Agent links with Operation by operation performed property and by contrast Operation has an inverse property performed by to Agent.

Ye, Yang, Jiang, & Tong [15] introduce Onto-SCM ontology for sharing the SCM conceptualization. The idea of Onto-SCM is evolved from general SCM. The ontology combines five main classes: SC_Management, SC_Structure, SC_Activity, SC_Resource, and SC_Item. The SC_management implies the management of SC operation in which activities and the important elements are linked from the remaining four main classes. SC_Structure classifies agent classes into five subclasses:Dyadic_Structure,Serial_ Structure,Divergent_Structure,Convergent_Structure,and Network_Structure. The SC_Structure has Strategy that is derived from Goal class and has SC_Process that implies the implementation of SC to achieve the goal. SC_Activity has Activity class. Activity has sub-activity property that is connected under the interval time. SC_Resource presents resources used in SCM. There are four resource classes: Production_Resource, Storage_Resource, Transportation_ Resource, and Human_Resource. Each resource has a role. SC_Item refers to products produced from SC. There are three item classes: Offer, Business_Order, and Plan. Offer has meant the products that are separated to Product or Service. Business_Order means the flow of object that transfers from one activity to other. Onto-SCM can explain the relationship between the activity and goal, the agent who performs activity, and the resource used in SCM at a given activity.

Soares, Azevedo, and Sousa [16] propose an ontology to support the process flow management of semiconductor manufacturing. The ontology helps agents to communicate regarding the requirement of the business's particular planning and control system. There are three semantic processes: Organizations, Plans Management, and Orders Management. Organizations have VE Unit class that represents the agents involved in the SC process. Plan management has three main classes: Plan presents the semantics of activities; Planning presents the semantics of action to implement Plan; Resource presents the necessary entity that serves a given activity. Resource class has a capacity as a data property, which means the allocation of resources used in each activity. Order management includes the classes of Order and Product. Order has been described by Order Item and its data properties such as quantity, due date, delivery date. Product is produced by supplier that is included in VE Unit.

Vegetti et al. [33] propose ontology for the SC domain. The goals are to demonstrate the SC structure, process, resource, and application evaluation. The structure identifies SC infrastructure, including roles, and combines the following classes: Supply Chain, Organizational Unit, Organizational Unit Role, and Functional Area. The process dimension defines SC's processes and subprocesses, including the execution of a process within a specific timeframe. Process, Process Occurrence, Temporal Relationship, and Utilization are the classes that make up the process. The resource dimension depicts resource utilization and its role in the SC process. The resource combines the classes Resource and Resource Role. Each class has data properties as well as relationships to other classes.

Spoladore et al. [34] propose health tourism ontology. The goal is to connect natural resources with visitor demands. There are six main classes: Destination, Descriptor, Variable, KPI_Families, Optima, and Target_Groups. The Destination explains the tourism inflows, whereas Descriptor focuses on regional services and characteristics. The Variable, KPI_Families, and Optima depict the value that health tourists will receive. Target_Groups has six subclasses that explain different types of health tourists. Chantrapornchai and Choksuchat [35] provide an ontology that collects information on health tourism. Many classes make up the ontology, including Product, Package, Facility, Service, Treatment, and so on. Almost every class has subclasses; for example, the Service class has two subclasses: BeautyService and Massage. Ontology has re-used classes and inferred concepts.

Previous studies [14,15,16,33] created ontologies using the core SC concept. This study may benefit from some structure design. However, there is a difference between the SC and MTSC concepts. For example MTSC agents are grouped based on service provided to medical tourists such as medical providers, medical tourists, tourism providers whereas SC agents are derived from general industry such as suppliers, customers, transporters. The MTSC involves the process of utilizing medical tourists' services until the follow-up process, while SC includes the process of producing items until selling products to customers. The flow of data, money, and product differs between MTSC and general SC. Medical tourists' service requests and consumption of services are the objects exchanged between or within the activity flows of MTSC. On the other hand, SC's flows are concerned with raw materials, physical goods trading, and purchase orders.

Taking into account the health tourism ontologies [34, 35], the earlier designs focused on gathering and formalizing information for tourist decision-making. The MTSC domain cannot be represented by previous ontology structures. Some health tourism terms, on the other hand, can be reused. The considerations for the reuse of ontologies are discussed in the following section. Table I compares previous studies that were unable to provide semantic knowledge about MTSC.

TABLE I. COMPARISON OF PREVIOUS ONTOLOGIES WITH THE PRESENT WORK

Aspects	Domain	Target user	Source of knowledge	MTSC related
[14]	General industry	Practitioners	International standard	No
[15]	General industry	Practitioners	Previous work, case study	No
[16]	Semiconductor industry	Practitioners	Previous work, case study	No
[33]	Food industry	Practitioners	Previous work, case study	No
[34]	Health tourism	Tourists, government	Case study	No
[35]	Health tourism	Tourists	Case study	No
This Work	Medical Tourism	Practitioners	Previous work, case study	Yes

III. BUILDING A MEDICAL TOURISM SUPPLY CHAIN ONTOLOGY

The goal of ontology development is to express knowledge in such a way that both people and computers can comprehend it [22]. Developing an ontology usually begins with recognizing the need for a certain purpose [20]. This study used a hybrid waterfall methodology with three key steps in the formation of the MTSC ontology adapted from [10, 20].

A. Specification

1) Determination of ontology domain and scope: The MTSC ontology focuses on the MT domain. This ontology represents the domain knowledge model that captures information of the MT attributes based on MTSC concept. The competency questions were created and measured the degree to which the ontology nailed its purpose and scope [22]. As a basis, the ontology should answer general questions, for example:

- How many actors are featured in the MTSC?
- What is the actor's primary role in the MTSC?
- What is the main process chain of MTSC?
- What are the activities in a given MTSC process?
- What is the core product produced in MTSC and by whom?
- How many flows do appear in the MTSC?

For complex knowledge representation, the ontology should answer further questions, for example:

- Which actors are involved in a given MTSC process?
- What is the actor's role when involving a given MTSC activity?

- What are the flows that are activated in a given MTSC activity?
- What are the inputs and outputs that relate to a given MTSC activity?
- What is the direction of the MTSC flow in each activity and between whom?

2) Consideration of the existing ontology: The literature demonstrates the use of ontology to explain the SC semantically. Although the existing ontologies did not pertain to the MTSC domain, there is some idea that can be reused. Previous ontology for sharing SC information at the top-level [14] has been considered as a starting point for the MTSC ontology model and the idea of the class (actor, action, product, and flow) is re-used. Some terminology from [35] is taken into account in the MTSC ontology. However, there is no ontology in the MTSC domain that represents enough terminologies. As a result, the next step involves the enumeration of terminology.

3) Enumeration of important terms in the ontology: Because terms constitute the foundation of an ontology, it's critical to be precise and unambiguous about the main terms and their associated attributes. The motivation of the ontology representation is derived from [14]. Fig. 1 shows an ontology model. The model is composed of seven main concepts: MT actors, MT processes, MT activities, MT roles, MT products, MT flows, and MT flow between.

MT actors represent the semantic knowledge of who is involved in the MTSC. MT actors can be organization units, companies who participate in MTSC system, or even individuals. MT actors have a role in MTSC. The MT actor and MT role conceptual model are derived from the stakeholder investigations [36, 4, 37, 18].

MT processes and MT activities represent the semantic knowledge of operations performed in the MTSC. The process description visualizes the logistics and flow of information and includes many activities. MT processes have been activated by MT actors. The actions of MT actors cause the forwarding of MT products and MT flows. The MT process and MT activity conceptual model is suggested by [6, 29, 27, 18].

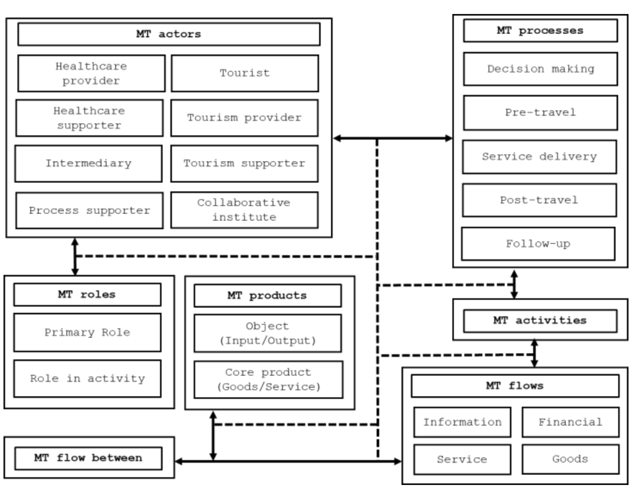


Fig. 1. MTSC Ontology Representation Adopted from [14].

MT products represent the semantic knowledge of objects produced from MT actors including objects required to perform activity. MT product conceptual model is captured from [14, 18, 35].

MT flows represent the semantic knowledge of object movement from a process starting point to the end of MTSC system. Flow is a thing with distinct and independent existence and involves the transfer of an object between respective activities and actors. The MT flow and MT flow between conceptual model is extracted from [14, 6, 29, 18].

This study focuses on the trustworthiness and practical benefits of ontology. Interviews with MT experts were used to build and verify the ontology structure and terminologies. Sixteen domain experts who have experience in MT business from Thailand, took part in the face-to-face interviews. The interviewing techniques begin by presenting the ontology design and components. Second, interviewers were requested to corroborate the ontology conceptualization and share MT knowledge. The final stage is to summarize the understanding and double-check the accuracy of the interview results with the interviewers right away. Table II lists the brief profiles of the MT experts who assisted with the interview parts.

TABLE II. NUMBER OF MT EXPERTS IN THE INTERVIEW SECTIONS

MT Domain Group	**Number of MT experts**	**Approximate experience in MT**
Health provider (Hospital, Dental clinic)	4	10-20 years
Tourism provider (Hotel, Tour operator)	3	15-20 years
Collaborative institute (University)	2	12-25 years
Health and tourism supporter (Interpreter, Insurance, Shop, Restaurant)	7	5-21 years

As a result of sharing knowledge with domain experts, terms such as important MT actor terms, process and activity names, and linkages among terms were discussed according to actual usage. The initiated terms can be used to advance the construction of ontologies. The obvious benefit of bringing in a diverse group of experts is that they can share their knowledge and illustrate relationship design approaches.

B. Conceptualization

1) Definition of the classes and the class hierarchy: The class hierarchy is designed using the top-down approach [10]. Classes and subclasses are categorized based on MTSC ontology model designed with the terms mentioned earlier. The ontology classes are as follows:

MT_Actor: refers to the stakeholders involved in the business and has eight subclasses.

MT_Product: refers to the objects produced by actors including input/output used in activities and has two subclasses.

MT_Process: refers to the processes in MTSC.

MT_Activity: refers to the activities that are part of the process.

MT_Flow: refers to an entity that transfers objects within an activity or from one activity to another activity and has four subclasses.

MT_FlowBetween: refers to the interaction of the actors in any flow.

MT_Role: refers to the responsibility/task of actors in MTSC and has two subclasses.

2) Definition of classes-slot properties: The data properties were defined as the qualification for each particular class. Table III identifies data property, representation, and domain usage of ontology creation in this study.

TABLE III. DATA PROPERTIES OF MTSC ONTOLOGY CLASSES

Data property(Domain usage)	**Representation**
hasActorName(MT_Actor)	Name of actor
hasActorDesc(MT_Actor)	Description of actor
hasPrimaryRoleDesc(PrimaryRole)	Description of primary role
hasRoleInActivityDesc(RoleInActivity)	Description of role in activity
hasProductDesc(MT_Product)	Description of product
hasProductType(MT_Product)	Type of product: *goods, services, input, output*
hasProcessDesc(MT_Process)	Description of process
hasActivityDesc(MT_Activity)	Description of activity
hasFlowDesc(MT_Flow)	Description of flow
hasFlowDirection(MT_Flow)	Direction of flow: *one-way, two-way*

Object properties are defined to link the relationship between two instances. Table IV shows the object properties with the boundaries of their inferences between domains and ranges.

3) Creation of instances: This step provides values to the defined classes and properties. For the instances of MTSC ontology, they are derived mainly from academic literature and a case study for credibility and practicality. The case study is from the document of experienced private hospital agency in Thailand that provides medical tourism services to foreign patients. The knowledge obtained from the experiences correlates with academic education. The knowledge agents presented in MTSC are reliable and can be used to explain MTSC.

For example, Hospital/Clinic is an instance of the MT_Actor class and the HealthProvider subclass. Its data property is hasActorName and the attributions of the property are "Hospital, Clinic, Health Institution, Health Center, Medical Center". This implies that the *hospital/clinic* has multiple names used in the MT business. We present an excerpt of the instances of the MTSC ontology in the implementation section.

TABLE IV. OBJECT PROPERTIES OF ONTOLOGY CLASSES FOR ROLE-BASED RECOMMENDATION

Object property(Domain > Range)	**Inference**	**Cardinality**
hasPrimaryRole(MT_Actor > PrimaryRole)	An actor has primary role	1,1
hasRoleInActivity(MT_Actor > RoleInActivity)	An actor has role in activity	1,N
makeProduct(MT_Actor > MT_CoreProduct)	An actor makes core product	1,1
performActivity(MT_Actor > MT_Activity)	An actor performs activity	1,N
consistOfActivity(MT_Process > MT_Activity)	A process consists of activity	1,N
activateFlow(MT_Activity > MT_Flow)	An activity activates flow	1,N
isPerformedBy(MT_Activity > MT_Actor)	An activity is performed by actor	1,N
relateToProcess(MT_Activity > MT_Process)	An activity relates to process	1,1
getInput(MT_Flow > MT_Object)	A flow gets input to perform activity	1,N
returnOutput(MT_Flow > MT_Object)	A flow returns output after performing activity	1,N
isFlowBetween(MT_Flow > MT_FlowBetween)	A flow is action between actors	1,N

C. Implementation

1) MTSC ontology with protégé 5.0: This study implemented an ontology by using Protégé 5.0 software with OWL language. The tool can present knowledge and collaborate with other tools and is available as software implementation [38]. The OWL is a markup language focusing on semantic web representation and mostly manipulates XML language expressions. The MTSC ontology taxonomy is presented using OntoGraf (Protégé-OWL plugin). Fig. 2 shows the MTSC hierarchy and relationships in the top-level design.

Fig. 3 and 4 present an excerpt of the instances of the MTSC ontology.

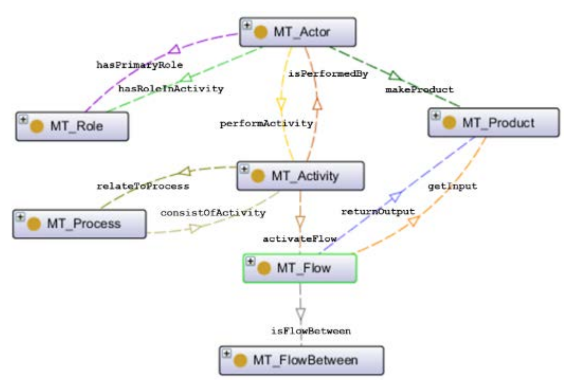


Fig. 2. A Top-level Hierarchy of MTSC Ontology.

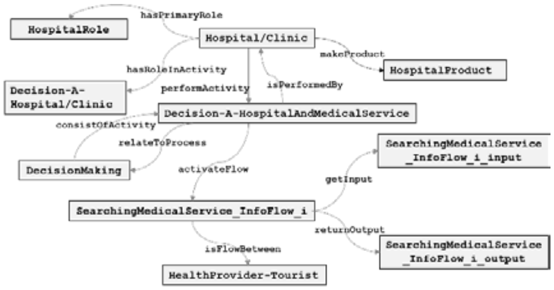


Fig. 3. An Excerpt of Health Provider Instance of MTSC Ontology.

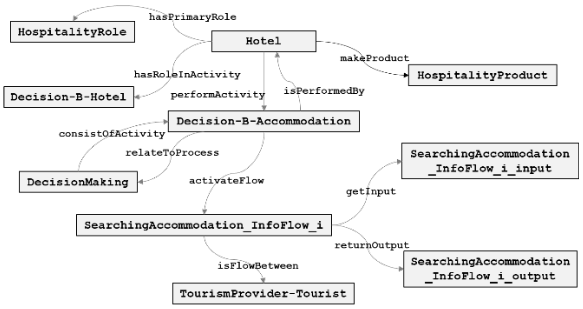


Fig. 4. An Excerpt of Tourism Provider Instance of MTSC Ontology.

2) Ontology evaluation: The MTSC ontology is assessed by competency questions and ontology metrics. The MTSC objective and scope are tested against SPARQL Query formulations. The query results were supposed to address competency questions. Some sample questions as SPARQL queries are as follows:

CQ: How many actors are featured in the MTSC?

SELECT ?Actor ?ActorName ?ActorDesc

WHERE {?Actor p:hasActorName ?ActorName.

?Actor p:hasActorDesc ?ActorDesc.}

The sample answer:

Actor	ActorName	ActorDesc
Bank/ForeignExchang	"Bank, Financial compa	"A financial institution that is legally responsible for deposits and withdrawals, currency exchange , and other financial tra
FoodSupplier	"Food supplier, Food m	"A producer who produces food products and distributes them to individuals or organizations that have purchased the p
GovernmentAgency	"Accreditation body, Ed	"A government agency that has established and controlled policies, functions, and measures for stakeholders manageme
HealthCareMarketer	"Healthcare marketer, "	"A marketing public relation operator who promotes medical services to be known"^^<http://www.w3.org/2001/XMLSc
Hospital/Clinic	"Hospital, Clinic, Health	"A health care provider who offers medical services to medical tourists"^^<http://www.w3.org/2001/XMLSchema#string
Hotel	"Hotel, Accommodatio	"A hospitality provider who offers accommodation and facilitation for tourists based on their intended purpose"^^<http
InsuranceCompany	"Insurance company, In	"An insurance company that is legally licensed to operate the insurance business, has a duty to pay compensation to the
InternationalTransport	"Airplane, Bus, Boat"^^	"An international transporter who offers international travel in various forms such as planes, boats, cars, trains."^^<http:
InterpreterProvider	"Interpreter, Translator"	"A communication mediator who helps bridge understanding between the speaker and the audience"^^<http://www.w
LocalTransporter	"Local transporter, Bus,	"A local transport provider who services at destination country such as train, subway, taxi, minibus, domestic airplane."^
MedicalSupplier	"Medical supplier, Medi	"A medical device manufacturer, modern medicine provider, including traditional medicine provider who produces and
MedicalTourist	"Medical Tourist, Touris	"A foreign tourist who intends to use medical tourism services"^^<http://www.w3.org/2001/XMLSchema#string>

CQ: What is the actor's primary role in the MTSC?

SELECT ?Actor ?PrimaryRole ?PrimaryRoleDesc

WHERE {?Actor p:hasPrimaryRole ?PrimaryRole.

?PrimaryRole p:hasPrimaryRoleDesc ?PrimaryRoleDesc.}

The sample answer:

Actor	PrimaryRole	PrimaryRoleDesc
Bank/ForeignExchangeProvider	FinancialRole	"Provides deposit and withdrawal services including currency exchange and other financial transactions"^^<http:
FoodSupplier	FoodSupplierRole	"Produces food products and sells them to individuals or organizations that have purchased the products"^^<http
GovernmentAgency	GovernmentAgencyRole	"Issues policy, functions, and measures for industry direction and coordination between stakeholders"^^<http://w
HealthCareMarketer	HealthCareMarketerRole	"Promotes medical services and medical tourism industry to be known"^^<http://www.w3.org/2001/XMLSchema
Hospital/Clinic	HospitalRole	"Offers and serves medical services to medical tourist"^^<http://www.w3.org/2001/XMLSchema#string>
Hotel	HospitalityRole	"Offers and serves accommodation and facilitation for tourists"^^<http://www.w3.org/2001/XMLSchema#string>
InsuranceCompany	InsurerRole	"operates the insurance business and has a duty to pay compensation to the insured from the events that occur ur
InternationalTransporter	InterTransporterRole	"Provides international travel in various forms such as planes, boats, cars, trains to medical tourists"^^<http://www
InterpreterProvider	InterpreterRole	"Communicates between the doctor and the patient using the patient's preferred language"^^<http://www.w3.or
LocalTransporter	LocalTransporterRole	"Services local travel such as trains, subways, taxis, minibusses, domestic airplanes"^^<http://www.w3.org/2001/
MedicalSupplier	MedicalSupplierRole	"Manufacturers medical devices, modern medicines, traditional medicine, and medical-related goods that used m
MedicalTourist	MedicalTouristRole	"Obtains medical tourism services"^^<http://www.w3.org/2001/XMLSchema#string>
[illegible]	[illegible]	[illegible]

CQ: Which actors are involved in a given process of MTSC?

SELECT ?Actor ?Process ?Activity

WHERE {?Actor p:performActivity ?Activity.

?Process p:consistOfActivity ?Activity.}

The sample answer:

Actor	Process	
HealthCareMarketer	DecisionMaking	Decision-A-HospitalAndMedicalService
HealthCareMarketer	DecisionMaking	Decision-D-Intermediary
Hospital/Clinic	DecisionMaking	Decision-A-HospitalAndMedicalService
Hospital/Clinic	FollowUp	Follow-A-TrackSymptom
Hospital/Clinic	PostTravel	Post-B-ShareExperience
Hospital/Clinic	PreTravel	PreTravel-A-ReserveHospitalAndService
Hospital/Clinic	ServiceDelivery	Service-C-ReceiveMedicalService
Hospital/Clinic	ServiceDelivery	Service-D-DischargePayment
Hospital/Clinic	ServiceDelivery	Service-F-PostOperativeCare
Hotel	DecisionMaking	Decision-B-Accommodation
Hotel	PostTravel	Post-B-ShareExperience
Hotel	PreTravel	PreTravel-B-BookAccommodation

Table V summarizes the result of ontology metrics.

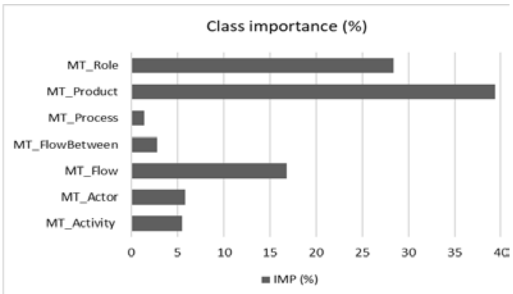
TABLE V. THE ONTOLOGY SCHEMA AND KNOWLEDGEBASE METRICS ASSESSMENT RESULT

Metrics	Ontology schema metrics			Knowledgebase metrics		
	RR	AR	IRs	CR	P	Coh
MTSC ontology	50%	4.70	0.70	100%	15.78	0

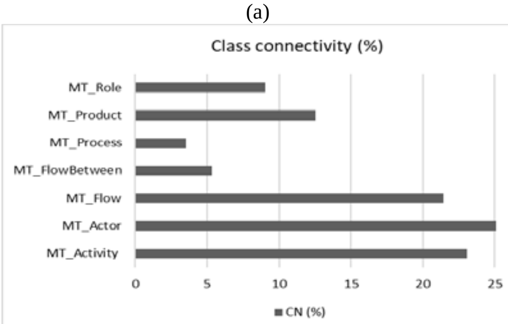
The ontology schema metrics reflect the quality of ontology structure for representing knowledge. The RR result (50%) indicates that MTSC ontology has many connections other than class-subclass, such as object property relationships. The AR result (4.70) indicates that the ontology has a good average of attributes per class. There are many types of knowledge presented. The IRs result (0.7) indicates that the ontology has a general distribution of subclasses and is not considered detailed knowledge. Overall, the MTSC schema metrics indicate good quality of structure design.

The results of knowledgebase metrics show that all ontology classes are implemented to deliver knowledge by the CR result of 100%. The P result (15.78) indicates that this ontology has a not too small knowledge in schema. The Coh result (0) indicates that all components in the ontology are used to describe MTSC. Overall, the knowledgebase metrics have appropriate values. The MTSC ontology has enough knowledge representation.

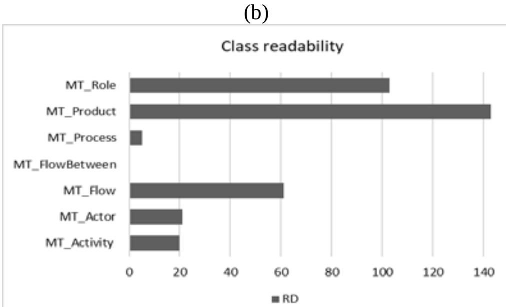
Fig. 5 presents the class metric results. The class metrics results indicate that the most important classes are MT_Product, MT_Role, and MT_Flow. The Imp result makes logical because these three classes store detailed knowledge. The central connectivity is MT_Actor followed by MT_Activity and MT_Flow. The Cn result reflects that the MTSC can explain knowledge using actors and activities as the center linked with other knowledge classes. Rd reflects that almost all of the classes provide descriptions for the users to understand. Overall, the class metrics have appropriate values.



(a)



(b)



(c)

Fig. 5. Class Metric Results in Values of (a) Imp, (b) Cn, and (c) Rd.

The MTSC ontology has a good knowledge distribution. This study was carried out with the collaboration of domain experts. The answers to SPARQL Queries and the ontology metrics values were verified by experts. The findings are adequate to clarify the study's goal and scope. The same group of 16 experts approved the MTSC ontology. Ontology structure and terminologies have quality.

IV. Discussion

This study presented the development of an application ontology for representing MTSC knowledge. The ontology engineering methodology consisted of specification, conceptualization, and implementation steps. The steps are enriched with knowledge engineering techniques and present a practical conceptual design. The key views of MT were verified. The MTSC ontology combines seven main classes namely 'MT_Actor', 'MT_Role','MT_Process', 'MT_Activity', 'MT_Flow', 'MT_FlowBetween', and 'MT_Product'; and has sixteen subclasses. The class conceptualizations can describe MTSC structurally. The concepts are in line with general SC presented by [14].

This MTSC ontology provides instances of 21 actors, 21 primary roles, 5 processes, 20 activities, 21 core products, 122 objects, 82 roles in activity, and 61 flows. The MTSC logistics has two main routes: medical services or tourism services. The MTSC process is divided into five stages. The actor involved in all processes is a medical tourist. The players on the service side who are involved in all processes of the medical side and tourism side are the health provider and tourism provider, respectively, including many supporters. MTSC ontology offers four types of knowledge flows as in [14]. Flow provides insights into how activities contribute to the flow and who is involved in that activity and flow in MT context. MTSC logistics are different from general SC logistics [23]. The MTSC activities are different from general SC which involves stakeholders. The short flows do not pass data to the next step like a production line, such as flow of providing an interpreter during treatment, or flow of selling souvenirs. However, the knowledge flow representation is consistent with MTSC process of [6] and [18]. Core products in MTSC ontology are mostly classified as services that are different from general SC where the products are usually tangible [16].

The ontology evaluation employed competency questions and ontology metrics. Domain experts had verified the quality of structure, design, and usefulness. According to SPARQL Query results, the MTSC ontology can answer semantic questions given in the Specification section. It denotes that the ontology is trustworthy in terms of objective and scope. The ontology metric results reflect the suitability of ontology structure and design. The body of knowledge generated is sufficient to describe the MTSC. Knowledge is distributed in an acceptable manner.

V. Conclusion

MTSC ontology generates a clear business understanding for stakeholders in the MT domain. The ontology construction demonstrated a high-quality development approach that can be replicated. The ontology development incorporated academic material and case studies into the acquisition process. The collaboration of domain experts in many activities makes ontology more dependable and efficient. The proposed ontology comprises concepts relevant to the MT domain. The ontology can describe the supply chain of the medical tourism domain: it can tell who is involved in which processes, what roles they play, and what flows are taking place. Based on the evaluation results, the ontology design and knowledge representation are of a quality.

We plan to apply the ontology in our next work a role-based recommendation system. Stakeholders can use the system to learn about MTSC and the ontological relationship that exists in the MT domain. The expertise used in this paper, however, came from a case study. The limitation of this study is that some aspects of the overall challenges may have been overlooked as a result of the different perspectives. Those who will use this work context must examine how well it fits within the framework of the business area.

ACKNOWLEDGMENT

This work was supported by the National Research and Innovation Information System (Project-code 2563NRCT321664, 2020).

REFERENCES

[1] D. S. Sandberg, "Medical tourism: an emerging global healthcare industry," International Journal of Healthcare Management, vol. 10, no. 4, pp. 281-288, 2017.

[2] P. M. Carrera and J. F. Bridges, "Globalization and healthcare: understanding health and medical tourism," Expert Review of Pharmacoeconomics & Outcomes Research, vol. 6, no. 4, pp. 447-454, 2006.

[3] J. Connell, "Contemporary medical tourism: conceptualisation, culture and commodification," Tourism management, vol. 34, pp. 1-13, 2013.

[4] A. Kamassi, N. H. A. Manaf and A. Omar, "The identity and role of stakeholders in the medical tourism industry: state of the art," Tourism Review, vol. 75, no. 3, pp. 559-574, 2020.

[5] P. Benckendorff and A. Zehrer, "A network analysis of tourism research," Annals of Tourism Research, vol. 43, pp. 121-149, 2013.

[6] L. Kaewkitipong, "The thai medical tourism supply chain: its stakeholders, their collaboration and information exchange," Thammasat review, vol. 21, no. 2, pp. 60-90, 2018.

[7] G. G. Binhan OĞUZ and H. H. Cruz, "Medical tourism in the time of covid-19," Global Political Trends Center (GPoT), 2020.

[8] UNWTO, "Impact assessment of the covid-19 outbreak on international tourism," 12 2020. [Online]. Available: https://www.unwto.org/impact-assessment-of-the-covid-19-outbreak-on-international-tourism. [Accessed 12 03 2022].

[9] N. Lunt, R. Smith, M. Exworthy, S. T. Green, D. Horsfall and R. Mannion, "Medical tourism: treatments, markets and health system implications: a scoping review," Organisation for Economic Co-operation and Development, Paris, 2011.

[10] N. F. Noy and D. L. McGuinness, Ontology Development 101: A Guide to Creating Your First Ontology, California: Stanford University, 2000.

[11] T. R. Gruber, "Toward principles for the design of ontologies used for knowledge sharing," International journal of human-computer studies, vol. 43, no. 5-6, pp. 907-928, 1995.

[12] C. Chandra and A. Tumanyan, "Organization and problem ontology for supply chain information support system," Data & Knowledge Engineering, vol. 61, no. 2, pp. 263-280, 2007.

[13] T. Grubic and I.-S. Fan, "Supply chain ontology: review, analysis and synthesis," Computers in Industry, vol. 61, no. 8, pp. 776-786, 2010.

[14] S. Üreten and H. K. Đlter, "Supply chain management ontology: towards an ontology-based SCM model," in In Proc. of the Fourth Int. Logistics and Supply Chain Management Congress, 2006.

[15] Y. Ye, D. Yang, Z. Jiang and L. Tong, "An ontology-based architecture for implementing semantic integration of supply chain management," International Journal of Computer Integrated Manufacturing, vol. 21, no. 1, pp. 1-18, 2008.

[16] A. L. Soares, A. L. Azevedo and J. P. D. Sousa, "Distributed planning and control systems for the virtual enterprise: organizational requirements and development life-cycle," Journal of Intelligent Manufacturing, vol. 11, no. 13, pp. 253-270, 2000.

[17] F. G. Alberti, J. D. Giusti, F. Papa and E. Pizzurno, "Competitiveness policies for medical tourism clusters: government initiatives in Thailand," International Journal of Economic Policy in Emerging Economies, vol. 7, no. 3, pp. 281-309, 2014.

[18] S. Karadayi Usta and S. SerdarAsan, "A conceptual model of medical tourism service supply chain," Journal of Industrial Engineering and Management, vol. 13, no. 2, pp. 246-265, 2020.

[19] E. Cohen, "Medical tourism in Thailand," AU-GSB e-journal, vol. 1, no. 1, pp. 24-37, 2008.

[20] S. Staab, H.-P. Schnurr, R. Studer and Y. Sure, "Knowledge processes and ontologies," IEEE Intelligent Systems, vol. 16, no. 1, pp. 26 - 34, 2001.

[21] S. Tartir, I. B. Arpinar, M. Moore, A. P. Sheth and B. Aleman-Meza, "OntoQA: metric-based ontology quality analysis," IEEE ICDM Workshop on Knowledge Acquisition for Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources, 11 2005.

[22] M. Grüninger and M. S. Fox, "Methodology for the design and evaluation of ontologies," in In Proc. of the IJCAI Workshop on Basic Ontological Issues in Knowledge Sharing, Montreal, 1995.

[23] B. M. Beamon, "Supply chain design and analysis: models and methods," International journal of production economics, vol. 55, no. 3, pp. 281-294, 1998.

[24] P. Trkman, M. I. Stemberger and J. Jaklic, "Information transfer in supply chain management," Issues in Informing Science & Information Technology, vol. 2, pp. 559-573, 2005.

[25] J. T. Mentzer, W. DeWitt, J. S. Keebler, S. Min, N. W. Nix, C. D. Smith and Z. G. Zacharia., "Defining supply chain management," Journal of Business logistics, vol. 2, no. 2, pp. 1-25, 2001.

[26] H. K. Lee and Y. Fernando, "The antecedents and outcomes of the medical tourism supply chain," Tourism Management, vol. 46, pp. 148-157, 2015.

[27] M. Ferrer and A. Medhekar, "The factors impacting on the management of global medical tourism service supply chain," GSTF Journal on Business Review (GBR), vol. 2, no. 2, pp. 206-211, 2012.

[28] J. Connell, "Medical tourism: sea, sun, sand and… surgery," Tourism Management, vol. 27, no. 6, pp. 1093-1100, 2006.

[29] H. S. Yoon, S. W. Cho and V. Sugumaran, "A service system design to support medical tourism in South Korea," in Proc. of intl conf on information resources management, Seoul, 2011.

[30] R. d. A. Falbo, "SABiO: Systematic Approach for Building Ontologies," in 1st JointWorkshop Onto.Com/ODISE on Ontologies in Conceptual Modeling and Information Systems Engineering, Rio de Janeiro, 2014.

[31] N. Guarino, "Formal ontology in information systems," Proceedings of the first international conference (FOIS'98), pp. 3-15, 6-8 06 1998.

[32] M. M. Sulaeman and M. Harsono, "Supply chain ontology: model overview and synthesis," Jurnal Mantik, vol. 5, no. 2, pp. 790-799, 2021.

[33] M. Vegetti, A. Böhm, H. Leone and G. Henning, "SCONTO: A Modular Ontology for Supply Chain Representation," in Domain Ontologies for Research Data Management in Industry Commons of Materials and Manufacturing, 2021.

[34] D. Spoladore, E. Pessot, M. Bischof, A. Hartl and M. Sacco, "Collaborative design approach for the development of an ontology-based decision support system in health tourism," Working Conference on Virtual Enterprises, vol. 629, pp. 611-620, 2021.

[35] C. Chantrapornchai and C. Choksuchat, "Ontology construction and application in practice case study of health tourism in Thailand," SpringerPlus, vol. 5, no. 1, pp. 1-31, 2016.

[36] M. Harryono, Y.-F. Huang, K. Miyazawa and V. Sethaput, "Thailand medical tourism cluster," Microeconomics of Competitiveness, MA, USA, 2006.

[37] A. Jabbari, M. Ferdosi, M. Keyvanara and Z. Agharahimi, "Stakeholders' analysis of the medical tourism industry: development strategies in Isfahan," Journal of Education and Health Promotion, vol. 2, no. 44, pp. 1-7, 2013.

[38] E. Alatrish, "Comparison some of ontology editors," Management Information Systems, vol. 8(2013), no. 2, pp. 18-24, 2013.

Stochastic Marine Predator Algorithm with Multiple Candidates

Purba Daru Kusuma, Ratna Astuti Nugrahaeni
Computer Engineering, Telkom University, Bandung, Indonesia

***Abstract*—This work proposes a metaheuristic algorithm that modifies the marine predator algorithm (MPA), namely, the stochastic marine predator algorithm with multiple candidates (SMPA-MC). The modification is conducted in several aspects. The proposed algorithm replaces the three fixed equal size iteration phases with linear probability. Unlike the original MPA, in this proposed algorithm, the selection between exploration and exploitation is conducted stochastically during iteration. In the beginning, the exploration-dominant strategy is implemented to increase the exploration probability. Then, during the iteration, the exploration probability decreases linearly. Meanwhile, the exploitation probability increases linearly. The second modification is in the prey's guided movement. Different from the basic MPA, where the prey moves toward the elite with small step size, several candidates are generated with equal inter-candidate distance in this work. Then, the best candidate is chosen to replace the prey's current location. The proposed algorithm is then implemented to solve theoretical mathematic functions and a real-world optimization problem in production planning. The simulation result shows that in the average fitness score parameter, the proposed algorithm is better than MPA, especially in solving multimodal functions. The simulation result also shows that the proposed algorithm creates 9%, 19%, and 30% better total gross profit than particle swarm optimization, marine predator algorithm, and Komodo mlipir algorithm, respectively.**

***Keywords*—Metaheuristic; marine predator algorithm; stochastic system; production planning**

I. Introduction

Optimization is a subject that is widely used and studied. Optimization is implemented in many areas, especially in operations research, such as manufacturing [1], logistics [2], transportation [3], education [4], finance [5], and so on. Optimization becomes more important because of its objective nature to maximize productivity or output or minimize resources within certain constraints and limitations. This circumstance often occurs in real-world problems, from the simple one like managing the school bus route to the complex one, such as handling the production process in manufacturing that builds products with many components, such as cars, airplanes, ships, and so on.

In general, optimization methods can be divided into two groups: exact and approximate. The exact methods have an advantage that true or global optimal is guaranteed to find. The problem is that the exact method needs excessive computation resources in solving complex and large dimension problems. On the other hand, approximate methods do not guarantee that global optimal can be found. The objective of approximate methods is to find near-optimal or acceptable optimal while avoiding local optimal [6], especially in multimodal problems. Fortunately, the approximate approach is popular because of its adaptability to computational resource constraints. The metaheuristic algorithm is a well-known and widely used method that uses an approximate approach. In metaheuristic, optimization is achieved during iteration.

Many studies have proposed new metaheuristic algorithms in this last decade. Many algorithms were inspired by nature, especially animals. These algorithms were developed based on animal behavior during foraging, such as grey wolf optimizer (GWO) [7], dragonfly algorithm (DA) [8], whale optimization algorithm (WOA) [9], and so on. Besides foraging, several algorithms were developed by mimicking animal behavior during reproduction, such as Komodo mlipir algorithm (KMA) [10], red deer algorithm (RDA) [11], Cuckoo search algorithm (CSA) [12], butterfly optimization algorithm (BOA) [13], and so on.

Besides proposing a new algorithm, many studies in metaheuristic algorithms were conducted to modify the existing algorithm. These modifications were conducted to improve the algorithm's performance or to make the algorithm more suitable to solve specific problems. Several well-known algorithms that have been widely modified or combined are genetic algorithm (GA), particle swarm optimization (PSO), simulated annealing (SA), tabu search (TS), and so on. Farag et al. [14] improved the binary-real coded genetic algorithm with k-means clustering to solve the unit commitment problem. Deb et al. [15] developed a non-dominated sorting genetic algorithm (NSGA II) derivative of a genetic algorithm to find pareto optimal in solving the multi-objective optimization problem. Sylia et al. [16] hybridized the PSO with proportional fair scheduling (PFS) to solve resource allocation in the orthogonal frequency division multiplexing (OFDM) transmission for future long-term evolution (LTE) and 5G networks.

One shortcoming metaheuristic algorithm is the marine predator algorithm (MPA). This algorithm is a metaphor-inspired algorithm that mimics the behavior of sea predators, such as shark, marlin, and swordfish, during hunting prey [17]. This algorithm is one of the rare metaheuristic algorithms that uses iteration to control the exploration and exploitation. The iteration is divided into three phases. The first phase conducts exploration [17]. The second phase conducts both exploration and exploitation [17]. The third phase conducts exploitation [17]. Exploration is mostly conducted by implementing the Brownian motion, while exploitation is conducted by implementing the Levy movement.

Even though this algorithm is new, it has been used in many optimization studies, such as in task scheduling [18], power system [19], hydrothermal scheduling [20], and so on. Moreover, studies conducted on MPA modification have been found but are still limited. Based on this circumstance, this MPA is still potential to modify. Several studies in metaheuristic, such as KMA [10], also used MPA as a performance comparison.

Based on this opportunity, this work proposes modifying and improving the basic MPA. As metaheuristic algorithm, the proposed model consists of conceptual model, algorithm in pseudocode, and the mathematical model. In this work, the proposed model is evaluated by implementing this proposed model into the simulation to solve the theoretical mathematical optimization problem and a real-world optimization problem. In this simulation, the convergence and sensitivity of the algorithm are also evaluated.

The contributions of this work are as follows.

1) The proposed algorithm replaces the static division of the iteration with the stochastic approach where the opportunity to conduct exploration or exploitation changes during the iteration.

2) This work proposes the existence of several candidates during the Brownian motion or Levy movement, where their fitness score is considered to become the prey's next move.

3) This work implements the modified version of MPA to optimize real-world production planning problem.

The remainder of this paper is organized as follows. The mechanism of the basic MPA is discussed in the second section. The model of the proposed algorithm that consists of a conceptual, algorithm, and mathematical model is explained in the third section. The fourth section explains the simulation to evaluate the proposed algorithm's performance. In this work, there are five simulations. The first to fourth simulations are conducted to evaluate the proposed algorithm's performance in solving 23 well-known optimization functions. The fifth simulation is conducted to evaluate the proposed algorithm's performance in solving a real-world production optimization problem. The more profound analysis related to the simulation result and findings is discussed in the fifth section. Finally, the conclusion and the future research potential related to this work are summarized in the sixth section.

II. Related Work

The marine predator algorithm is a metaheuristic algorithm that adopts sea predator behavior or movement during foraging or hunting prey [17]. This algorithm combines Brownian motion and Levy movement. Levy movement is a derivative of random walk movement whose characteristics are closely related to sea predators, such as shark, swordfish, or marline, during searching for prey [21]. In this algorithm, the Levy movement is combined with the Brownian motion to conduct exploration and exploitation. MPA consists of two sets of agents: predators and prey. The adoption of the Levy movement is like the Cuckoo search algorithm (CSA). CSA is developed based on the parasitism behavior of cuckoo birds during finding a nest for their eggs [12]. In CSA, the cuckoo implements the Levy movement only [12].

This algorithm divides exploration and exploitation depending on the iteration. It is very different from many common metaheuristic algorithms, such as particle swarm optimization, genetic algorithm, harmony search (HS), and so on, where the decision of running the exploration or exploitation does not depend on the iteration. MPA focuses on avoiding local optimal in the early phase and improving the solution in the later phase.

Although rare, one example of an algorithm where the exploitation-exploration decision depends on the iteration is simulated annealing (SA). In general, SA focuses on exploitation by conducting neighborhood search in every iteration [22]. Exploration is conducted by accepting a new worse solution based on some probabilistic calculation to avoid local optimal [22]. In SA, the outer loop iterates from the initial high temperature to the final low temperature. When the temperature is high, the new worse solution is easily accepted. During the decrease in temperature, accepting the worse solution becomes more difficult [22]. In the end, a new worse solution is hard to accept. This exploration-to-exploitation approach is like MPA but with different mechanism.

In MPA, the iteration is divided into three phases with the same duration. In the first phase, the process focuses on exploration by implementing the Brownian movement for all prey. The objective is that local optimal should be avoided in the early phase. In the second phase, the population is divided into two equal-size groups. The first group consists of preys that conduct exploration by implementing the Brownian movement. The second group consists of preys that conduct exploitation by implementing the Levy movement. The objective is that the algorithm focuses on improving searching quality. In the third phase, all populations focus on exploitation by adopting the Levy movement. This division is illustrated in Fig. 1.

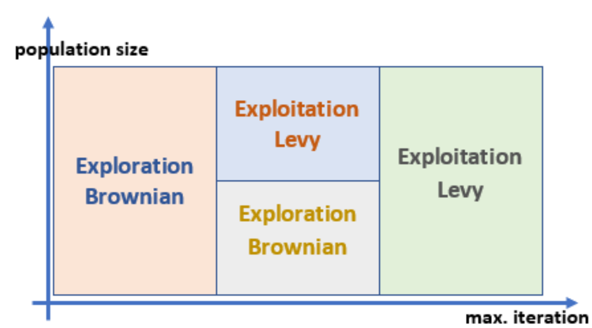


Fig. 1. Static Division in Marine Predator Algorithm.

There are several sequential processes in every iteration. The first process is the prey's movement which is explained earlier. The second process is updating the predators. The prey replaces its related predator only if its fitness is better than its related predator. The third process is the eddy formation. This eddy formation is another method used in MPA to avoid local optimal. This process is conducted stochastically based on the fish aggregating device affects that range from 0 to 1. There are two possible actions related to this process based on probabilistic calculations. If a certain generated random number is below the fish aggregating devices, the prey randomly conducts a long jump within its local problem space that narrows as iteration goes. Otherwise, this prey will move toward two randomly selected prey at a certain speed.

In the prey's guided movement, whether Brownian movement or Levy flight, only one location is considered, as shown in Fig. 2. This location may be within the path between the prey and the elite or the extended distance from the elite. The determination of this new location depends on the step size, which is determined stochastically depending on the selected movement and the gap between the predator and the prey. The fitness value is not considered.

Fig. 2. Guided Movement in Marine Predator Algorithm.

Moving toward the best solution is common in many metaheuristic algorithms. In particle swarm optimization (PSO), each agent moves toward local best and global best with a certain proportion [23]. In Komodo mlipir algorithm (KMA), female mates with the highest quality big male to produce two offspring [10]. The first offspring is close to the female, and the second one is close to the highest quality big male. Then, offspring whose fitness is better becomes the replacement. Meanwhile, the small male moves toward the big male. In the red deer algorithm, this idea is implemented during the fighting between the male commander and stag and the mating of the commander and harem [11].

Even though the MPA is proven as a competitive algorithm, there are several questions or review due to this algorithm. First, is there any possible method to conduct exploration-to-exploitation approach despite this fixed division during the iteration? Second, is there any method to improve the movement of the prey rather than the small step size?

There are several possible modifications due to this basic MPA mechanism. The first is eliminating the fixed size division of the iteration while the concept of exploration dominant in the early iteration and the exploitation dominant in the later iteration is still adopted. The second is to create several new location candidates for prey during the guided movement. Their fitness score is considered so that the prey moves to a more promising location.

III. Proposed Model

In this section, the proposed model will be discussed in detail. The model consists of a conceptual and mathematical model. The conceptual model explains the concept and the difference between the proposed algorithm and the original MPA, especially in the exploration and exploitation division and the improvement of the guided movement. The mathematical model consists of the main algorithm of SMPA-MC and the mathematical formulae following the algorithm.

Like MPA, this proposed algorithm consists of two sets of agents: preys and predators. Both preys and predators have equal population sizes. The relation between prey and predator is one-to-one. After the prey moves, then their fitness score is evaluated. If the prey's fitness score is better than the predator's fitness score, then the predator moves to the prey's location.

As a metaheuristic algorithm, SMPA-MC consists of two parts: initialization and iteration. In the initialization, the initial prey' and predators' location is generated randomly within the problem space using a uniform distribution.

As a derivative version of MPA, the iteration affects the exploration and exploitation division in this proposed algorithm. Unlike MPA, where this division is divided into fixed three phases, this division is conducted based on a stochastic approach in this proposed algorithm. In the beginning, the probability of exploration is high. Contrary, the probability of exploitation is low. During the iteration, the probability of exploration declines linearly while the probability of exploitation climbs up linearly. At the end of the iteration, the probability of exploitation is high, while the probability of exploration is low. This mechanism is illustrated in Fig. 3.

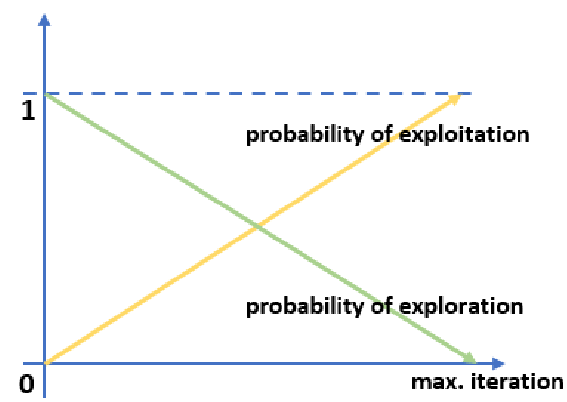


Fig. 3. Stochastic-based Exploitation and Exploration.

The guided movement of the proposed algorithm is also different from the MPA. This proposed algorithm generates multiple candidates between the prey and the elite. The inter-candidate distance is equal. One candidate whose fitness is the best among these candidates is then chosen as the best candidate. This mechanism is adopted from KMA, especially in the mating process between the highest quality male and the female [10]. The difference is that in KMA, sexual reproduction only produces two offspring. In this proposed algorithm, the guided movement generates multiple candidates. The best new generation or candidate becomes the replacement. Then, this best candidate location becomes the prey's next location. This concept is illustrated in Fig. 4.

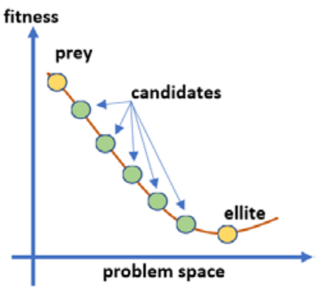


Fig. 4. Guided Movement with Multiple Candidates / Alternatives.

Different mechanism also occurs during the eddy formation. During this process, there are two possible actions. The first is that the prey moves randomly within its local problem space. In the beginning, the local problem space is wide. During the iteration, this local problem space decreases linearly too. It also reflects the transition from exploration to exploitation during the iteration. The second action is that the prey moves to a location in the middle of the prey's current location, and the other prey's location is selected randomly. Although the mechanism during eddy formation is different, the motivation is the same.

This conceptual model is then interpreted into sequential steps. These sequential steps are as follows.

- Step 1: generate initial preys and predators.
- Step 2: set initial iteration.
- Step 3: generate several candidates based on the movement that is chosen stochastically.
- Step 4: select the best candidate to replace the current prey.
- Step 5: update the related predator.
- Step 6: implement eddy formation.
- Step 7: if the maximum iteration has not been reached, go to step 3. Otherwise, iteration stops.
- Step 8: select the best predator to become the final solution.

This conceptual model is then transformed into a mathematical model. The mathematical model consists of two parts: main algorithm and formulae. The main algorithm is shown in algorithm 1. Meanwhile, several annotations used in the mathematical model are as follows.

b_l	lower bound
b_u	upper bound
c	candidate
c_{best}	best candidate
C	set of candidates
f	fitness
fad	fish aggregating devices
P	population
r	predator
R	set of predators
t	time / iteration
t_h	time threshold
t_{max}	maximum iteration
y	prey
Y	set of preys

algorithm 1: SMPA-MC main algorithm

```
//initialization
for i = 1 to n(P)
  generate (y_i)
  generate (r_i)
end
//iteration
for t = 1 to t_max
  generate (t_h)
  for i = 1 to n(P)
    if U(0, 1) < t_h then
      for j = 1 to n(C)
        generate-guided-exploitation (c_j, y_i, r_i)
      end
      c_best = select best candidate (C)
    else
      for j = 1 to n(C)
        generate-guided-exploration (c_j, y_i, r_i)
      end
      c_best = select best candidate (C)
    end if
    y_i = c_best
    r_i = update (y_i, r_i)
    if U(0,1) < fad then
      y_i = limited random move (y_i, t_h, l_i)
    else
      y_sel = U(Y)
      y_i = half move (y_i, y_sel)
    end if
  end
end
s_final = find best (R)
```

All predators' and preys' initial location is generated in the initialization process. This initial location is generated randomly within the problem space. This process is formalized by using (1) and (2). Equation (1) generates the prey's initial location, while (2) is used to generate the predator's initial location. All predators and prey are distributed randomly within the problem space.

$$y = U(b_l, b_u) \tag{1}$$

$$r = U(b_l, b_u) \tag{2}$$

The iteration process runs after the initialization process ends. At the beginning of every iteration, a time threshold is calculated. This threshold determines whether this iteration is conducted for guided exploitation or guided exploration. The time threshold is calculated by using (3).

$$t_h = \frac{t}{t_{max}} \tag{3}$$

A random number is then generated, and it follows uniform distribution as shown in algorithm 1. If this random number is less than the time threshold, guided exploitation is conducted. Otherwise, guided exploration is conducted. In both guided exploration and exploitation, several candidates are generated. Then, the best candidate is selected among these candidates. After the best candidate is selected, this candidate replaces the related prey. This process is formalized by using (4) to (6). Equation (4) is used for the guided exploitation. Equation (5) is used for the guided exploration. Equation (6) formalizes the best candidate selection.

$$c_j = r + \frac{j}{n(C)}(r - y) \tag{4}$$

$$c_j = y + \frac{j}{n(C)}(r - y) \tag{5}$$

$$c_{best} = c \in C \wedge \min\big(f(c)\big) \quad (6)$$

The predator location is then evaluated after a prey moves to its new location. If this prey's fitness score is better than the predator's fitness score, then this prey becomes the new predator. This process is formalized by using (7).

$$r' = \begin{cases} y, f(y) < f(r) \\ r, else \end{cases} \quad (7)$$

The last process in every iteration is applying the eddy formation. There are two possible actions in this process. The selection determined stochastically end depends on the fish aggregating devices value. This process can be limited to random movement or half movement. A prey will move randomly within its local problem space in the limited random movement. On the other hand, in the half movement, a prey will move to the middle between its current location and other prey selected randomly. The limited random movement is formalized using (8), while the half movement is formalized using (9).

$$y' = y + (2U(0,1) - 1)(1 - t_h)(b_u - b_l) \quad (8)$$

$$y' = y + \frac{y_{sel} - y}{2} \quad (9)$$

The complexity of this algorithm, as it is presented in big O notation, is O(t_{max}.$n(P)$.$n(C)$). This presentation means that the complexity is the multiplication between the maximum iteration, population size, and the number of candidates.

IV. Simulation and Result

The proposed algorithm is then implemented into a simulation to observe its performance. There are five simulations conducted in this work. The first simulation is conducted to evaluate its performance in solving mathematical problems. The second simulation is conducted to evaluate its performance in achieving the convergence condition. The third simulation is conducted to evaluate the proposed algorithm's performance related to the fishing aggregate devices. The fourth simulation is conducted to evaluate the performance related to the number of candidates. The fifth simulation is conducted to evaluate its performance in solving a real-world problem.

In this simulation, the proposed SMPA-MC algorithm is compared with several algorithms: PSO, HS, hide object game optimizer (HOGO), KMA, and MPA. The reason for choosing these algorithms as a comparison is that these algorithms use distinct exploration-exploitation mechanisms. PSO and HS represent the well-known old-fashioned algorithm. HOGO and KMA represent the shortcoming algorithms that hybridize many common methods. MPA is chosen to observe the performance improvement due to modifying its basic form.

PSO is a well-known algorithm that is developed based on swarm intelligence. In PSO, each agent moves to a new location depending on the weighted cumulative method among its current location, its local best, and the global best [23]. The global best represents the collective intelligence shared among agents, and it is updated every time a new local best is found [23].

HS represents the non-population-based metaheuristic algorithm. Moreover, this algorithm is the simplest one among other algorithms. The exploration-exploitation decision is conducted based on the stochastic approach [24]. A new solution can be generated from the harmony memory (exploitation) or anywhere else within the problem space (exploration) based on the harmony memory considering rate (HMCR) [24].

HOGO represents a game-based algorithm. It mimics the behavior of the old hide-object game. This algorithm is also a population-based algorithm that consists of a set of agents. The agent's movement depends on the global best, the global worst, and the randomly selected agent through a weighted cumulative method [25]. An agent tends to move toward the global best and avoid the global worst [25].

KMA represents a hybrid metaheuristic algorithm. It combines swarm intelligence and an evolution system. The males conduct the PSO-like movement by moving toward the better big males [10]. On the other hand, the evolution system is conducted by the female by mating with the highest quality big male to generate better offspring [10]. Meanwhile, exploration is conducted by parthenogenesis or asexual reproduction [10].

MPA represents an algorithm where the current iteration affects the decision to conduct exploration or exploitation. In KMA, HOGO, HS, and PSO, the iteration does not affect the decision. It also represents a population-based algorithm where an agent consists of two engaged agents: prey and predator.

There are several common parameters used in the simulation. These parameters are shown in Table I. The value of population size and maximum iteration represents the moderate computation process. The weights in PSO represent the balance movement.

TABLE I. PARAMETERS DEFAULT VALUE

Parameters	**Default Value**
population size	20
maximum iteration (except HS)	200
maximum iteration (HS)	4000
fishing aggregate devices (MPA and SMPA-MC)	0.2
current speed weight (PSO)	0.5
social weight	0.5
cognitive weight	0.5
number of candidates (SMPA-MC)	10

In the first simulation, the proposed algorithm is implemented to solve or find the global optimal of the given functions. There are 23 functions to be solved. Seven functions are unimodal functions. Six functions are multimodal functions. Ten functions are fixed dimension multimodal functions. The seven unimodal functions include Sphere, Schwefel 2.22, Schwefel 1.2, Schwefel 2.21, Rosenbrock, Step, and Quartic. The multimodal functions include Schwefel, Rastrigin, Ackley, Griewank, Penalized, and Penalized-2. The fixed dimension multimodal functions include Foxholes, Kowalik, Six Hump Camel, Branin, Goldstein-Price, Hartman 3, Hartman 6, Shekel

5, Shekel 7, and Shekel 10. The detail of the functions, which consists of formulae, dimension, problem space, and global optimal, is shown in Table II. Meanwhile, the result is shown in Table III.

Table III shows that in general, the proposed algorithm performs well and meets the metaheuristic criteria in finding the near-optimal solution and avoiding the local optimal trap. Moreover, the proposed algorithm can find the true optimal solution in solving five multimodal functions: Shekel Foxholes, Kowalik, Six Hump Camel, Branin, and Goldstein-Price. Unfortunately, its performance is not so good in solving Hartman 3 function.

Compared to other algorithms, the proposed model is competitive enough. Its performance is superior in solving 10 functions. Meanwhile, HOGO has become the most challenging algorithm due to its outstanding performance in solving 9 functions. Compared with MPA, the proposed SMPA-MC is better at solving 13 functions. Most of these are multimodal functions, especially the fixed dimension multimodal functions with narrow problem space. The proposed algorithm also outperforms at least three algorithms in solving 22 functions.

The second simulation is conducted to evaluate the performance of the proposed algorithm in achieving the convergence situation. This simulation is conducted by solving the 23 benchmark functions. There are three maximum iterations in this simulation: 50, 100, and 150. The result is shown in Table IV.

TABLE II. BENCHMARK FUNCTIONS

No	Function	Model	Dimension	Problem Space	Global Opt.
1	Sphere	$\sum_{i=1}^{d} x_i^2$	10	[-100, 100]	0
2	Schwefel 2.22	$\sum_{i=1}^{d}\lvert x_i\rvert + \prod_{i=1}^{d}\lvert x_i\rvert$	10	[-100, 100]	0
3	Schwefel 1.2	$\sum_{i=1}^{d}\left(\sum_{j=1}^{i} x_j\right)^2$	10	[-100, 100]	0
4	Schwefel 2.21	$\max\{\lvert x_i\rvert, 1 \le i \le d\}$	10	[-100, 100]	0
5	Rosenbrock	$\sum_{i=1}^{d-1}(100(x_{i+1} + x_i^2)^2 + (x_i - 1)^2)$	10	[-30, 30]	0
6	Step	$\sum_{i=1}^{d-1}(x_i + 0.5)^2$	10	[-100, 100]	0
7	Quartic	$\sum_{i=1}^{d} i\, x_i^4 + random\ [0,1]$	10	[-1.28, 1.28]	0
8	Schwefel	$\sum_{i=1}^{d} -x_i \sin\left(\sqrt{\lvert x_i\rvert}\right)$	10	[-500, 500]	-4189.8
9	Ratsrigin	$10d + \sum_{i=1}^{d}\left(x_i^2 - 10cos(2\pi x_i)\right)$	10	[-5.12, 5.12]	0
10	Ackley	$-20 \cdot exp\left(-0.2 \cdot \sqrt{\frac{1}{d}\sum_{i=1}^{d} {x_i}^2}\right) - exp\left(\frac{1}{d}\sum_{i=1}^{d}\cos 2\pi x_i\right) + 20 + exp(1)$	10	[-32, 32]	0
11	Griewank	$\frac{1}{4000}\sum_{i=1}^{d} x_i^2 - \prod_{i=1}^{d} cos\left(\frac{x_i}{\sqrt{i}}\right)+1$	10	[-600, 600]	0
12	Penalized	$\frac{\pi}{d}\left\{10\sin(\pi y_1) + \sum_{i=1}^{d-1}\left((y_i - 1)^2\left(1 + 10\, sin^2(\pi y_{i+1})\right)\right) + (y_d - 1)^2\right\} + \sum_{i=1}^{d} u(x_i, 10,100,4)$	10	[-50, 50]	0
13	Penalized 2	$0.1\left\{sin^2(3\pi x_1) + \sum_{i=1}^{d-1}\left((x_i - 1)^2\left(1 + sin^2(3\pi x_{i+1})\right)\right) + (x_d - 1)^2\left(1 + sin^2(2\pi x_d)\right)\right\} + \sum_{i=1}^{d} u(x_i, 5,100,4)$	10	[-50, 50]	0
14	Shekel Foxholes	$\left(\frac{1}{500} + \sum_{j=1}^{25}\frac{1}{j+\sum_{i=1}^{2}(x_i - a_{ij})^6}\right)^{-1}$	2	[-65, 65]	1
15	Kowalik	$\sum_{i=1}^{11}\left(a_i - \frac{x_1(b_i^2 + b_i x_2)}{b_i^2 + b_i x_3 + x_4}\right)^2$	4	[-5, 5]	0.0003
16	Six Hump Camel	$4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1x_2 - 4x_2^2 + 4x_2^4$	2	[-5, 5]	-1.0316
17	Branin	$\left(x_2 - \frac{5.1}{4\pi^2}x_1^2 + \frac{5}{\pi}x_1 - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos(x_1) + 10$	2	[-5, 5]	0.398
18	Goldstein-Price	$\left(1 + (x_1 + x_2 + 1)^2(19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2)\right).\left(30 + (2x_1 - 3x_2)^2(18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2)\right)$	2	[-2, 2]	3
19	Hartman 3	$-\sum_{i=1}^{4}\left(c_i exp\left(-\sum_{j=1}^{d}\left(a_{ij}(x_j - p_{ij})^2\right)\right)\right)$	3	[1, 3]	-3.86
20	Hartman 6	$-\sum_{i=1}^{4}\left(c_i exp\left(-\sum_{j=1}^{d}\left(a_{ij}(x_j - p_{ij})^2\right)\right)\right)$	6	[0, 1]	-3.32
21	Shekel 5	$-\sum_{i=1}^{5}\left(\sum_{j=1}^{d}(x_j - c_{ji})^2 + \beta_i\right)^{-1}$	4	[0, 10]	-10.1532
22	Shekel 7	$-\sum_{i=1}^{7}\left(\sum_{j=1}^{d}(x_j - c_{ji})^2 + \beta_i\right)^{-1}$	4	[0, 10]	-10.4028
23	Shekel 10	$-\sum_{i=1}^{10}\left(\sum_{j=1}^{d}(x_j - c_{ji})^2 + \beta_i\right)^{-1}$	4	[0, 10]	-10.5363

TABLE III. SIMULATION RESULT (MEANS)

Function	PSO	HS	KMA	HOGO	MPA	Proposed Model	Better than
Sphere	329.99	505.1458	507.3948	**4.4651e-11**	0.1662	0.1857	PSO, HS, KMA
Schwefel 2.22	3.0223e-18	**0**	0.0021	1.1295e-7	**0**	**0**	PSO, KMA, HOGO
Schwefel 1.2	1,410.4218	1,203.6542	1,677.9713	**0.1747**	0.8628	13.5649	PSO, HS, KMA
Schwefel 2.21	14.2562	16.0000	14.4289	**0.0008**	0.2600	0.8221	PSO, HS, KMA
Rosenbrock	39,750.9494	48,923.1329	58,963.9907	**8.6153**	10.5113	21.5729	PSO, HS, KMA
Step	199.4486	327.0324	457.9096	**0.0120**	2.0221	0.0776	PSO, HS, KMA, MPA
Quartic	0.1062	0.0901	0.3252	0.1152	**0.0030**	0.0046	PSO, HS, KMA, HOGO
Schwefel	**-3,342.1276**	-3,328.4096	-3,172.5575	-2,758.9118	-2,031.1057	-3,212.5081	KMA, HOGO, MPA
Ratsrigin	31.9483	24.1875	37.1191	14.6673	**0.1469**	7.1187	PSO, HS, KMA, HOGO
Ackley	7.6684	9.0253	9.4044	**0.6494**	0.3033	1.1982	PSO, HS, KMA
Griewank	3.9131	5.4470	5.9587	**0.0559**	0.1842	0.3709	PSO, HS, KMA
Penalized	15.2125	17.6490	12.2738	**0.0175**	0.7669	0.5726	PSO, HS, KMA, MPA
Penalized 2	17,158.3907	7,766.3057	6,452.3411	**0.0864**	2.7432	0.1720	PSO, HS, KMA, MPA
Shekel Foxholes	10.1559	**0.9980**	7.1567	5.9722	5.1568	**0.9980**	PSO, HS, KMA, HOGO
Kowalik	0.0102	0.0017	0.0111	0.0047	0.0027	**0.0004**	PSO, HS, KMA, HOGO
Six Hump Camel	**-1.0316**	-1.0308	-1.0278	-1.0313	-1.0280	**-1.0316**	PSO, HS, KMA, HOGO
Branin	**0.3980**	0.3984	0.4167	0.4086	0.8053	**0.3980**	PSO, HS, KMA, HOGO
Goldstein-Price	**3.0000**	**3.0000**	3.9038	3.0195	4.1111	**3.0000**	PSO, HS, KMA, HOGO
Hartman 3	-3.4724	-0.0495	-0.7724	-0.0495	**-3.7843**	-0.0495	-
Hartman 6	-3.2106	-3.2723	-3.0073	-3.2157	-2.0356	**-3.3221**	PSO, HS, KMA, HOGO
Shekel 5	-6.4334	-5.8036	-7.9029	-4.6978	-2.0375	**-10.1310**	PSO, HS, KMA, HOGO
Shekel 7	-5.7752	-7.4658	-8.0417	-5.9069	-2.4365	**-10.3892**	PSO, HS, KMA, HOGO
Shekel 10	-4.8883	-5.0071	-6.0913	-6.1175	-2.2369	**-10.5257**	PSO, HS, KMA, HOGO

Table IV shows that in general, the convergence performance of the proposed algorithm is good. It achieves convergence in the early iteration while solving all fixed dimension multimodal functions. Besides, it also achieves convergence in the early iteration in solving the Schwefel 2.22 and Shekel Foxholes functions. Otherwise, it needs a higher maximum iteration to achieve convergence.

The third simulation is conducted to observe the sensitivity of the fishing aggregate devices related to the proposed algorithm's performance. The fishing aggregate devices are chosen due to its role in determining the exploration mechanism. In this simulation, there are three values of the fishing aggregate devices: 0.25, 0.5, and 0.75. These values represent the low, moderate, and high fishing aggregate devices. The result is shown in Table V.

Table V shows that the sensitivity of the fishing aggregate devices is various depend on the problem to solve. The increase of the fishing aggregate devices worsens the proposed algorithm's performance in solving the most of unimodal functions, except Schwefel 2.22. On the other hand, the fishing aggregate devices do not affect the proposed algorithm's performance in solving most of the multimodal functions.

The fourth simulation is conducted to evaluate the sensitivity of the number of candidates related to the proposed algorithm's performance. In this simulation, there are three values of the number of candidates: 5, 10, and 15. These values represent the low, moderate, and high number of candidates. The result is shown in Table VI.

Table VI shows that in general, the number of candidates has positive relation the proposed algorithm's performance. The increase of the number of candidates tends to improve the performance. This circumstance occurs in all functions: unimodal functions and multimodal functions. In the beginning, the improvement is significant. But, after the algorithm reaches its peak performance, the improvement is less significant. In some functions, such as Schwefel 2.22 and Goldstein-Price, the peak performance is achieved in the small number of candidates.

The fifth simulation is conducted to evaluate the performance of the proposed algorithm in solving the real-world optimization problem. An algorithm test using a real-world optimization problem is needed to prove that the algorithm is good theoretically and practically. In this simulation, the proposed algorithm is implemented to optimize the production planning process in a manufacturing company.

TABLE IV. CONVERGENCE TEST RESULT

Function	Average Fitness Score		
	t_{max} = 50	t_{max} = 100	t_{max} = 150
Sphere	46.4677	3.2242	0.7576
Schwefel 2.22	0.0020	0	0
Schwefel 1.2	148.1581	57.4689	21.4743
Schwefel 2.21	4.8131	2.5311	1.3052
Rosenbrock	652.8542	131.7495	54.1151
Step	24.0461	2.0388	0.2480
Quartic	0.0145	0.0077	0.0051
Schwefel	-2,639.4568	-2,975.1690	-3,115.1478
Ratsrigin	9.5166	10.0503	7.2051
Ackley	4.4461	2.5475	1.7403
Griewank	1.3237	0.7802	0.5090
Penalized	2.6256	0.9555	0.5157
Penalized 2	9.6735	1.6905	0.2997
Shekel Foxholes	1.0821	0.9980	0.9980
Kowalik	0.0008	0.0004	0.0004
Six Hump Camel	-1.0316	-1.0316	-1.0316
Branin	0.3981	0.3981	0.3981
Goldstein-Price	3.0000	3.0000	3.0000
Hartman 3	-0.0495	-0.0495	-0.0495
Hartman 6	-3.2780	-3.3199	-3.3217
Shekel 5	-9.6085	-10.0459	-10.0976
Shekel 7	-10.1142	-10.1338	-10.1073
Shekel 10	-10.0074	-10.4680	-10.5128

TABLE V. RELATION BETWEEN FISHING AGGREGATE DEVICES AND FITNESS SCORE

Function	Average Fitness Score		
	fad = 0.25	*fad* = 0.5	*fad* = 0.75
Sphere	4.3085	7.4816	13.1763
Schwefel 2.22	0	0	0
Schwefel 1.2	40.6408	58.1769	124.8039
Schwefel 2.21	2.4441	2.9302	3.5865
Rosenbrock	131.9147	105.0508	305.1955
Step	1.5093	5.1901	10.8524
Quartic	0.0077	0.0134	0.0195
Schwefel	-2,955.7188	-3,184.1091	-3,358.1613
Ratsrigin	10.4572	15.6704	19.3374
Ackley	2.6340	2.8596	3.2112
Griewank	0.8104	0.8855	1.1047
Penalized	1.1464	0.5891	1.3944
Penalized 2	1.1628	1.7262	3.2667
Shekel Foxholes	0.9981	0.9980	0.9980
Kowalik	0.0004	0.0010	0.0009
Six Hump Camel	-1.0316	-1.0316	-1.0316
Branin	0.3981	0.3981	0.3981
Goldstein-Price	3.0000	3.0000	3.0002
Hartman 3	-0.0495	-0.0495	-0.0495
Hartman 6	-3.3202	-3.3209	-3.3211
Shekel 5	-10.0811	-10.0605	-10.0372
Shekel 7	-10.3100	-10.3351	-10.2997
Shekel 10	-10.4786	-10.4635	-10.4343

TABLE VI. RELATION BETWEEN NUMBER OF CANDIDATES AND FITNESS SCORE

Function	Average Fitness Score		
	$n(C) = 5$	$n(C) = 10$	$n(C) = 15$
Sphere	10.2168	4.2298	1.5692
Schwefel 2.22	0	0	0
Schwefel 1.2	63.8637	54.9011	41.1862
Schwefel 2.21	3.7031	2.5755	2.4947
Rosenbrock	173.8898	167.1175	98.5991
Step	4.2628	2.3263	1.2434
Quartic	0.0111	0.0017	0.0068
Schwefel	-2,584.2545	-3,054.2677	-3,016.2328
Ratsrigin	17.5424	11.0106	8.1328
Ackley	3.1073	2.3593	2.4817
Griewank	0.9317	0.6514	0.7018
Penalized	1.2564	0.8481	0.5831
Penalized 2	2.7543	0.8494	1.0848
Shekel Foxholes	1.3601	0.9980	0.9980
Kowalik	0.0008	0.0004	0.0004
Six Hump Camel	-1.0316	-1.0316	-1.0316
Branin	0.3980	0.3980	0.3980
Goldstein-Price	3.0000	3.0000	3.0000
Hartman 3	-0.0495	-0.0495	-0.0495
Hartman 6	-3.0389	-3.3195	-3.3216
Shekel 5	-9.8176	-10.1143	-10.0888
Shekel 7	-10.3370	-10.3906	-10.3829
Shekel 10	-10.3174	-10.4665	-10.4899

The simulation scenario is Muslim socks manufacturer in Bandung, Indonesia. This company produces 40 product items. Half of them are long socks, while half others are short socks. Six items are fast-moving products while the others are moderate ones. The most fast-moving products are the light brown socks, both short and long. The other fast-moving products are white socks and black socks. Each item should be produced within the minimum and maximum production ranges. On the other hand, there is a limitation in the storage and financial capacity so that all produced socks cannot surpass the total production quantity. The maximum total capacity is only 5,250 dozen. The characteristics of every item are shown in Table VII. The production quantity is presented in dozen while the price is presented in rupiah. The objective is to maximize total gross profit.

TABLE VII. PRODUCT DESCRIPTION

Product ID	Min. Production (dozen)	Max. Production (dozen)	Gross Profit (rupiah/dozen)
1	500	1000	24,000
2, 3	200	400	24,000
4-20	50	100	24,000
21	500	1000	30,000
22, 23	200	400	30,000
24-40	50	100	30,000

This optimization problem can be seen as a Knapsack optimization problem. The concept of the Knapsack problem is that there is a space with a limited capacity [26]. On the other hand, there are several products to pick up. The objective is to determine the items, and the quantity picked to minimize or maximize the objective parameters.

The proposed algorithm is compared with PSO, HS, KMA, and MPA in this simulation. Due to its characteristic as a multi-dimension problem with a large search space, the maximum iteration for PSO, KMA, MPA, and SMPA-MC is set at 300. Meanwhile, the maximum iteration for HS is set at 12.000. The result is shown in Table VIII.

TABLE VIII. REAL-WORLD PRODUCTION PLANNING SIMULATION RESULT

Algorithm	Total Gross Profit (rupiah)
proposed (SMPA-MC)	186,589,200
PSO	171,114,000
HS	188,959,800
MPA	157,104,600
KMA	143,143,100

Table VIII shows that the proposed SMPA-MC outperforms three algorithms in creating better total gross profit. Its total gross profit is 9% higher than PSO, 19% higher than MPA, and 30% higher than KMA. On the other hand, its performance is only 1% lower than HS. This result shows that although HS is inferior in solving a theoretical mathematic problem, its performance is superior in solving real-world high dimension problems.

V. DISCUSSION

In general, Table III shows that the proposed SMPA-MC algorithm is better than the original MPA in solving multimodal functions. Its superiority especially occurs in solving multimodal functions with low dimension and narrow problem space, as indicated by the last ten functions. On the other hand, MPA is better at solving unimodal functions. The proposed algorithm is better at avoiding local optimal trap (exploration), while the MPA is better at finding the near-optimal solution or more precise solution. In the context of the method used in these algorithms, it is shown that the Levy movement creates more precise solutions than a uniform random or simple random walk.

There are several notes due to the competitiveness of the proposed algorithm. All metaheuristic algorithms use iteration to improve their current solution [6]. The result in Table IV shows this circumstance. Some functions can be solved faster, while others need more iteration, such as high dimension functions or functions with large problem space. Besides, an algorithm may be better in the early iteration, which means they are better in finding the convergence. On the other hand, some algorithms may be worse in the early iteration but better in the long run.

Metaheuristic is also identic with adjusted parameters. These parameters are provided to tune the algorithm's performance in the adaptation of many optimization problems. The inferior performance of PSO, HS, and KMA in Table III may come from the adjustment. By implementing different adjustments, an algorithm may perform better or worse depending on the problem it faces. It means that competing with one algorithm with the others is not the only tool to judge the algorithm's performance.

The adjustment also affects to the performance as it is shown in Table V and Table VI. Although exploration is important to avoid the local optimal trap, targeted exploration is proven more effective rather than the fully randomized exploration, especially in the later iteration. Higher fishing aggregate devices makes the probability of the fully randomized exploration higher. It means, the searching process will restart at location somewhere in the problem space and it is not productive in the later iteration. In some circumstance, the number of candidates gives positive results. But, after the algorithm reaches its peak performance, the increase of the number of candidates does not improve the algorithm's performance significantly.

Table III also strengthens the no-free-lunch theory [27]. Although, in general, PSO and HS are inferior compared to HOGO, MPA, and the proposed SMPA-MC, they are still superior in solving several functions. PSO is superior in solving four functions, while HS is superior and can find the true optimal solution in solving three functions.

As shown in Table VIII, the real-world simulation result demonstrates that superiority in solving a high-precision mathematical problem may not work in solving real-world problems. In theoretical mathematic problems, the parameters are usually represented in floating-point numbers. Very little difference between two floating-point numbers may give a significant gap in the result. An algorithm can achieve better performance by generating a more precise floating-point number. This process can usually be conducted by making small and high-precision step sizes during the guided movement. Small step size is usually achieved by generating a more precise random number, for example by using Levy movement or normal distribution. On the other hand, uniform random is usually less precise.

On the other hand, many real-world problems do not need very precise floating-point numbers. Many of them usually use integer numbers, especially in operations research. Many studies in operations research use integer numbers, for example, to find the number of products that should be produced or ordered. It is impossible to produce goods, for example, shoes, cars, and so on, in a fractional quantity. This circumstance makes the high precision optimization algorithm, such as KMA, MPA, or HOGO, lose their advantage.

Moreover, the objective function in real-world problems is simpler than the theoretical mathematic problems. In real-world problems, especially in operations research problems, most of their objectives can be presented in multi-variate linear functions, such as minimizing total tardiness [28], production cost [29], travel distance [30], and so on. This objective can be achieved by accumulating these parameters in all dimensions, for example by accumulating all due date penalties of all executed orders or accumulating the total quantity of all unexecuted orders due to limited production or storage capacity. This objective is even simpler than Sphere and Schwefel 2.22 functions, which are the simplest among 23 benchmark functions. But in real-world problems, some optimization problems use a multi-objective model.

This circumstance is also related to the popularity of the algorithm. Many studies in optimization, especially operations research, still use old-fashioned algorithms, such as genetic algorithm, tabu search, simulated annealing, or variable neighborhood search. This phenomenon indicates that these algorithms are still well-proven and competitive enough to solve real-world problems. However, they are often beaten by the shortcoming algorithms in solving mathematical functions. Besides, the mechanism of these old-fashioned algorithms is simple so that they are easy to modify or hybridize.

VI. Conclusion

This work has demonstrated that the proposed algorithm, the stochastic marine predator algorithm with multiple candidates, has proven as a good metaheuristic algorithm. It has achieved two main objectives of metaheuristic algorithm: finding a near-optimal solution and tackling the local optimal. The simulation result shows that its performance is competitive in solving optimization problems theoretically and practically. Among 23 benchmark functions, it achieves true optimal solution in solving 5 functions. Compared with other algorithms, its performance is also superior in solving 10 functions. This algorithm also outperforms the original form of the marine predator algorithm in solving 13 functions, which means 57 percent of total functions. Practically, it is also competitive in solving real-world problems. It outperforms particle swarm optimization, marine predator algorithm, and Komodo mlipir algorithm in optimizing production planning problems. Its performance is 9%, 19%, and 30% better than these three algorithms consecutively.

This work has shown that improving the existing algorithm is also important compared to proposing a new algorithm. This improvement can be conducted by modifying the current form of the algorithm or hybridizing this algorithm with another algorithm to combine the advantage of every algorithm. In the future, modifying the marine predator algorithm is still possible and challenging. Besides, implementing this proposed algorithm to solve more real-world optimization problems is still potential, especially in solving combinatorial problems, such as scheduling, timetabling, etc.

Acknowledgment

This work was financially supported by Telkom University, Indonesia.

References

[1] V. Fernandez-Viagas, “A speed-up procedure for the hybrid flow shop scheduling problem”, Expert Systems with Applications, vol. 187, ID: 115903, pp. 1-11, 2022.

[2] T. Chargui, A. Bekrar, M. Reghioui, and D. Trentesaux, “Scheduling trucks and storage operations in a multiple-door cross-docking terminal considering multiple storage zones”, International Journal of Production Research, vol. 60, no. 4, pp. 1153-1177, 2022.

[3] K. Huang, L. Xu, Y. Chen, Q. Cheng, and K. An, “Customized bus route optimization with the real-time data”, Journal of Advanced Transportation, vol. 2020, ID: 8838994, pp. 1-9, 2020.

[4] P. D. Kusuma and A. S. Albana, “University course timetabling model in joint courses program to minimize the number of unserved requests”, International Journal of Advanced Computer Science and Applications, vol. 12, no. 10, pp. 121-127, 2021.

[5] A. Mahmoudi, L. Hashemi, M. Jasemi, and J. Pope, “A comparison on particle swarm optimization and genetic algorithm performances in deriving the efficient frontier of stocks portofolios based on a mean-lower partial moment model”, International Journal of Finance & Economics, vol. 26, no. 4, pp. 5659-5665, 2020.

[6] H. R. Moshtaghi, A. T. Eshlaghy, and M. R. Motadel, “A comprehensive review on meta-heuristic algorithms and their classification with novel approach”, Journal of Applied Research on Industrial Engineering, vol. 8, no. 1, pp. 63-89, 2021.

[7] S. Mirjalili, S. M. Mirjalili, and A. Lewis, “Grey wolf optimizer”, Advances in Engineering Software, vol. 69, pp. 46-61, 2014.

[8] S. Mirjalili, “Dragonfly algorithm: a meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems”, Neural Computing and Applications, vol. 27, no. 4, pp. 1053-1073, 2016.

[9] S. Mirjalili and A. Lewis, “The whale optimization algorithm”, Adcances in Engineering Software, vol. 95, pp. 51-67, 2016.

[10] Suyanto, A. A. Ariyanto, and A. F. Ariyanto, “Komodo mlipir algorithm”, Applied Soft Computing, vol. 114, pp. 1-17, 2022.

[11] A. M. Fathollahi-Fard, M. Hajjaghaei-Keshteli, and R. Tavakkoli-Moghaddam, “Red deer algorithm (RDA): a new nature-inspired metaheuristic”, Soft Computing, vol. 24, pp. 14637-14665, 2020.

[12] A. H. Gandomi, X.-S. Yang, and A. H. Alavi, “Cuckoo search algorithm: a metaheuristic approach to solve structural optimization problems”, Engineering with Computers, vol. 29, pp. 17-35, 2013.

[13] S. Arora and S. Singh, “Butterfly optimization algorithm: a novel approach for global optimization”, Soft Computing, vol. 23, pp. 715-734, 2019.

[14] M. A. Farag, M. A. El-Shorbagy, I. Eldesoky, and A. A. El-Sawy, “Binary-real coded genetic algorithm based k-means clustering for unit commitment problem”, Applied Mathematics, vol. 6, no. 11, pp. 1873-1890, 2015.

[15] K. Deb, A. Pratap, S. Agarwai, and T. Meyarivan, “A fast and elitist multiobjective genetic algorithm: NSGA II”, IEEE Transactions on Evolutionary Computation, vol. 6, no. 2, pp. 182-197, 2002.

[16] Z. Sylia, C. Gueguen, B. Lamine, T. Larbi, and K. Abdelkrim, “New strategy for resource allocation using PSO-PFS hybrid”, International Journal of Wireless and Mobile Computing, vol. 18, no. 2, pp. 175-182, 2020.

[17] A. Faramarzi, M. Heidarinejad, S. Mirjalili, and A. M. Gandomi, “Marine predators algorithm: a nature-inspired metaheuristic”, Expert Systems with Applications, vol. 152, ID: 113377, 2020.

[18] M. Abdel-Basset, R. Mohamed, M. Elhoseny, A. K. Bashir, A. Jolfaei, and N. Kumar, “Energy-aware marine predators algorithm for task scheduling in IoT-based fog computing applications”, IEEE Transactions on Industrial Informatics, vol. 17, no. 7, pp. 5068-5076, 2021.

[19] M. A. Sobhy, A. Y. Abdelaziz, H. M. Hasanien, and M. Ezzat, “Marine predators algorithm for load frequency control of modern interconnected power systems including renewable energy sources and energy storage units”, Ain Shams Engineering Journal, vol. 12, no. 4, pp. 3843-3857, 2021.

[20] G. Chen, Y. Xiao, F. Long, X. Hu, and H. Long, "An improved marine predators algorithm for short-term hydrothermal scheduling", IAENG International Journal of Applied Mathematics, vol. 51, no. 4, pp. 1-14, 2021.

[21] N. E. Humphries, N. Queiroz, J. R. M. Dyer, N. G. Pade, M. K. Musyl, K. M. Schaefer, D. W. Fuller, J. M. Brunnschweiler, T. K. Doyle, J. D. R. Houghton, G. C. Hays, C. S. Jones, L. R. Noble, V. J. Wearmouth, E. J. Southall, and D. W. Sims, "Environmental context explains Levy and Brownian movement patterns of marine predators", Nature, vol. 465, pp. 1066-1069, 2010.

[22] T. Guilmeau, E. Chouzenoux, and V. Elvira, "Simulated annealing: a review and a new scheme", IEEE Statistical Signal Processing Workshop (SSP), Rio de Janeiro, Brazil, 2021.

[23] D. Freitas, L. G. Lopes, and F. Morgado-Dias, "Particle swarm optimisation: a historical review up to the current developments", Entropy, vol. 22, pp. 1-36, 2020.

[24] M. Dubey, V. Kumar, M. Kaur, and T.-P. Dao, "A systematic review on harmony search algorithm: theory, literature, and applications", Mathematical Problems in Engineering, vol. 2021, ID: 5594267, pp. 1-22, 2021.

[25] M. Dehghani, Z. Montazeri, S. Saremi, A. Dehghani, O. P. Malik, K. Al-Haddad, and J. M. Guerrero, "HOGO: hide objects game optimization", International Journal of Intelligent Engineering & Systems, vol. 13, no. 4, pp. 216-225, 2020.

[26] L. E. Schafer, T. Dietz, M. Barbati, J. R. Figueira, S. Greco, and S. Ruzika, "The binary knapsack problem with qualitative levels", European Journal of Operations Research, vol. 289, no. 2, pp. 508-514, 2021.

[27] S. P. Adam, S. A. N. Alexandropoulos, P. M. Pardalos, and M. N. Vrahatis, "No free lunch theorem: a review", Approximation and Optimization, Springer Optimization and Its Applications, vol. 145, pp. 57-82, 2019.

[28] B. Mor, G. Mosheiov, and D. Shabtay, "Minimizing the total tardiness and job rejection cost in a proportionate flow shop with generalized due dates", Journal of Scheduling, vol. 24, no. 6, pp. 553-567, 2021.

[29] H. M. Al-Hattami, J. D. Kabra, and M. Lokhande, "Reducing cost in manufacturing firm by using target costing technique", International Journal of Business Excellence, vol. 22, no. 1, pp. 69-82, 2020.

[30] P. D. Kusuma and M. Kallista, "Dispatching model in online taxi system by using stable marriage and first price sealed bid auction models", International Journal of Intelligent Engineering & Systems, vol. 14, no. 5, pp. 166-176, 2021.

A Novel High-Speed Key Transmission Technique to Avoid Fiddling Movements in e-Commerce

A.B. Hajira Be[1]
Ph.D Research Scholar
Department of Computer Science, Mother Teresa Women's University, Kodaikanal, Tamilnadu, India

Dr. R. Balasubramanian[2]
Professor & Dean
Department of Computer Applications, Karpaga Vinayaga College of Engineering and Technology,Maduranthagam Taluk, Tamilnadu, India

***Abstract*—To prevent fraud in e-shopping, the High-Speed Key Transmission Technique (HSKT) is primarily focused on safe and effectual transactions of payments and product content. The privacy of users, traders, trader information, product content, and the payment procedure are all protected by this framework. High speed key transmission is also committed to providing an effective, sensible approach to privacy in order to minimize the complexity of applications and the load on consumers. This paper proposes a new key transmission technique that allows users to conduct multi-banking account transactions from a single location. The proposed system is devoted to preventing unwanted people from gaining access. The secret key is used to represent access mechanisms, allowing authorized users to verify the transaction if and only if its characteristics meet secret key access requirements. Finally, the proposed high-speed key transmission technique (HSKT) improves the chances of success and throughput by reducing the time it takes to decrypt and encrypt, the amount of energy it uses, and the average delay.**

Keywords—Secret key; encryption; decryption; transaction; HSKT

I. Introduction

Payment issues plague e-commerce applications including electronic transactions using Visa otherwise debit cards, net banking, PayPal, or other tokens take more inconsistency issues since it is more vulnerable to being targeted. To prevent fraud in E- shopping, the high-speed key transmission method (HSKT) is primarily focused on safe and efficient payment processes and product content [1-5]. The privacy of users, traders, trader information, product content, and the payment procedure are all protected by this framework. HSKT is also committed to providing an effective, sensible approach to privacy in order to minimize the complexity of applications and the load on consumers. This method allows users to conduct multi-banking account transactions from a single location. This system is devoted to preventing unwanted people from gaining access [6-10].

The secret key used to represent access mechanisms, allowing authorized users to verify the transaction characteristics meet secret key access requirements. Finally, HSKT reduces decryption and encryption time, energy consumption, average delay, and increases achievement rate and throughput compared to other sites, which incur greater consequences if information is lost or modified. Although, the Indian government took positive steps encouraging rapid expansion of E-commerce enacting cyber-laws, by lowering framework taxes. People are hesitant to make online purchases due to concerns about security and installation methods [11]. Additionally, there are fraudulent debit cards and credit cards that happens to anybody when shopping online.

E-commerce applications suffer from payment problems; for instance, automated transactions using Debit cards or Visa, PayPal or Net banking, or methods by the way of other tokens partake added consistency issues and are more likely to be targeted than other sites because they suffer more consequences if information is lost or modified. Although, the Indian government has made significant steps to promote the rapid growth of E-commerce by enacting cyber legislation, lowering framework taxes, and so on. People are hesitant to make online purchases due to concerns about security and installation methods.

A. *Objectives*

The resulting ones are the main objectives of this work:

- To create a High-Speed Key Transmission (HSKT) technique that will allow for safe and quick financial transactions while preventing fraudulent behavior.
- To use a privacy method to keep track of traders and consumer content profiles.
- To provide an E-Commerce information retrieval system that is both effective and secure for both merchants and consumers.

To enhance the suggested method's success rate by reducing encryption and decryption time, average delay, and energy usage when compared to current methods.

II. Literature Review

The term "cryptography" comes from the Greek and means "the process of making data unintelligible." This prevents an unauthorized user from getting their hands on sensitive information. To put it another way, it's a method for concealing data in transit [12]. As soon as the transmitter uses cryptographic methods and a particular key to turn the data into cypher text, it is sent to the recipient. Encryption is the name given to this technique. To decrypt a message, the receiver uses a known key to decipher the ciphertext. Decryption is the term for this step.

There are two types of cyphers known as symmetric (or secret) and asymmetric (or public). In order to accomplish security, each of them makes use of a discrete and distinct set of techniques. Symmetric key cryptography focuses on the structure of simple repeated cryptographic processes, while asymmetric cryptography relies on the complexity of a mathematical problem [13].

To save energy, network designers should employ symmetric key cyphers to encrypt data sent by sensor nodes. Due to resource limits, standard cryptographic techniques cannot be used on sensor nodes. For these networks, a key management method provides the best security. Before exchanging information, nodes must safely exchange keys. The key management is a multi-operational approach where the initial key is produced, then disseminated and traded across the nodes, then utilized by the sender and recipient, and finally abandoned and renewed. Thus, key management system essential processes are: creation, distribution, exchange, usage, abolish, and refresh. WSNs are managed in several ways [14].

Conventional public-key cryptosystems have a major drawback in that in order to provide the high degree of security required, the key size must be suitably big [15]. In the third place, we have hybrid key cryptography, which brings together symmetric and public key elements. It combines the best features of the two approaches. Public-key cryptography (PKC) has been demonstrated to be a viable method of encrypting data[16].PKC is gaining traction in WSNs because it is capable of solving two basic and challenging challenges, namely authentication and symmetric key distribution, using DSA and Diffie-Hellman key exchange [17][18]. For sensor nodes with limited resources, ECC is an excellent choice since it has less overhead than RSA, but ECC operations are still hefty [19]. ECC is superior to RSA in terms of security [20].

III. Proposed Methodology

It concentrated on the cryptography algorithms. The reader can get a clear comparison of the key length and the analytical view of the cryptography methods to be utilized in the real-time applications utilizes the model. Where, AES, DES and Blowfish algorithm are discussed details with their key generation, key comparisons, encryption and decryption process. All conventional method is evaluated on regular postponement, energy intakes, throughput, encryption and decryption time. Blowfish algorithm is more secure for increasing the key size. It compared to other symmetric key algorithms and offered less processing time and rounds [21-25].

The secure and efficient payment transaction (SEPT) approach aims to prevent hackers from tracing one's online transaction. Initially, entering account numbers are very simple. Hence, the system gathers account numbers with allotted password details. Next, due to a multi encryption method, consumer will receive interim password for activating the account. Finally, Aadhar number should be entered to validate the consumer profile verifications and activate the account for further transactions [26-30].

A. Implementation Preprocessing Steps for SEPT

Implementation preprocessing steps are as follows Authentication process, Purchase Request, Authorization Demand, Authorization Response, Cardholder Authentication Demand, Cardholder Authentication Response and Ending Payment and Secure and Efficient Payment Transaction (SEPT) Approach.

The system has highly concentrated on the secure and efficient payment transactions to avoid fraudulent action. The study's primary goal is to provide an effective smart method to protect privacy while avoiding the complexity of apps and consumer burden. The study explains about HSKT algorithm with their system workflow, implemented modules, the implementation details, mathematical proof, operational details, and dataflow diagram. It explains the implementation, logical, analytical, and view of HSKT.

The HSKT method is used to get a clear concept of the logical and the analytical view of the proposed HSKT method for implementing in the real-time e-commerce applications. The proposed methodology ensures secrecy and efficiency in payment transactions of product content to avoid fraudulent activity in e-commerce.

B. Implementation Pre-Processing Steps for EHSKT

Implementation preprocessing steps are as follows Admin, Product Upload Module, Product Update and Product History, User Authentication, Product Search, View Shopped Products, Order Process and Payment Process. For safe and quick financial transactions, as well as to prevent fraudulent behavior, a High-Speed Key Transmission (HSKT) Technique is suggested. The suggested method generates a unique id for financial transactions using a UUID (Universally Unique Identifier). To prevent fraud, the proposed system produces a Unique ID (Identifier) that consists of a mix of numbers, alphabets, and special characters. This method allows users to conduct multi-banking account transactions from a single location. The suggested system is devoted to preventing fraud and unauthorized users.

The approach enhances the RSA methodology by using a 4096-bit key length, which makes it more successful for creating and distributing secret keys in hazardous settings. The secret key of authorized user is used to represent access mechanisms in this case, allowing transaction to be verified if and only if transaction features satisfy secret key access criteria. Vertical or horizontal partitioning of data is often anticipated. In case of horizontally partitioned data, several places collect the same set of information about distinct entities. The proposed system created security based on data attributes. The suggested method reduces the time required for encryption, decryption, and key complexity. The suggested design approach is as follows: Setup:

The method accepts a collection of numbers, alphabets, and a special character k as an input parameter and reverses public key PK and user secret key, USK. For the creation of unique transaction IDs, PK is used.

1) Secret key generation: Generating two large random prime numbers h and k and it has approximately equivalent size to item N = hk necessitated 4096 bits length. Computing N=secret key exponent SE, 1<SE<ϕ becomesiSE=1modϕ. The private key is (SE, h, k) and public key is (N, i). Maintain all the values of SE, h, k and ϕ secretly. When using SE, the private key is sometimes expressed as you need as the value of N. We could write the key pair as ((N, I SE) at other instances.

Modulus is a term that refers to the number N. The exponent is also known as the public exponent, encryption exponent, or simply exponent. The secret key exponent (SE) is also known as the decryption exponent.

This method will utilize the access tree structure h and k as well as the user secret key USK as inputs. This method uses a secret key or signature S that allows the user to verify transactions for specific users. Only an authorized user may generate using USK. Transactions Validations:

The user has the ability to make modifications, such as generating a message digest of the material to be shared. Exponent DE between 1 and N-1 is used to represent this message or content digest. Calculates the signature S=DESE mod N using the private key (N, SE) and sends it to the recipient. On the receiver's side, calculate integer x=Si mod N using the sender's public key exponent (N, I Calculates the message or content digest of the data that has been signed independently. It calculates the expected representative integer x′ by encoding the expected message digest if x=x′ , the signature is correct.

It takes user input together with their signature S in order to initiate the key access structure's payment validation process. The suggested method validates payment transactions if and only if the set of payment characteristics matches the user signature in the access tree.

2) Encryption: The user may get the receiver's public key and represent the plaintext content as a positive integer PI with 1PIN, compute the ciphertext CT=PIi mod N, and send the ciphertext to the receiver in the following methods.

3) Decryption: Based on the size of SE and N, the receiver uses his/her private key (N, SE) to compute PI=CTSE mod N using his/her private key (N, SE). It retrieves plain text from a message or a PI that represents content.

C. Working with EHSKT Technique System Constraints

H and K are two important prime numbers with bit widths of 4096 that are kept hidden. The size difference between the two sizes must be large enough in the security scenario.

N = H *K, ϕ= (H −1) (K −1), and ϕ is kept secret. Our software utilizes a 4096 bit modulus EHSKT implementation.PI can use only positive integer, but then great general separation with respect to N and ϕshould bw 1. SE * i ≡ 1 mod ϕ where i is an integer.

PK (PI, N) referred as public key PK, whereas SK (SE, N) is the private key. Considering PT equals plaintext, (PTPI) SE ≡ PT mod N ≡ PT mod H, PT mod K.

Process

Encryption: CT =Enc (PI, PT) = PTPI mod N.

Decryption: PT = Dec (SE, CT) = CTSE mod N = CTSE mod H, CTSE mod K.

Operations

Choose two big prime numbers and calculate N and ϕ.To speed up encryption select PI=3. Depended on PI and N then SE can be estimated. “Calculate PTkey mod H, PTkey mod K, where the key can be PI as well as SE.

Forϕ= (H–1) (K–1) is even, PI may be tried from 3. As a result, PI can be raised by 2 once.

The following is an example of a PI computation algorithm:

```
BigIntegerComputePI(BigIntegerϕ)
{
BigInteger PI=3; BigIntegerdivision=0;
While (GCD( ϕ,PI)!=1) // Here Greatest Common Divisor is the function to calculate the greatest common division of two integers, and it can be executed with EuclidAlgorithm.
{ PI=i+2;
}
Returni;
}
GeneratePublicKey (BigInteger PI)
{
BigInteger one = BigInteger.ONE; BigInteger two = one.add(one); BigInteger key = one.add(two); while(PI.GCD(key).compareTo(one) != 0)
{
key = key.add(two);
}
Return key;
}
GeneratePrivateKey (BigInteger N, BigInteger DE)
{
return DESE mod N; // SE is secret key exponent
}
publicBigInteger Encryption(BigInteger□)
{
powerModule(□)
Encryption CT= PTPI mod N
}
publicboolean Decryption(BigInteger CT, BigInteger SK)
{
BigInteger VS = powerModule(SK); if(VS.compareTo(SK) == 0) return true; Decrption PT=CTSE mod N;
return false;
}
publicbooleanVerifySignature(BigInteger CT, BigInteger SK)
{
BigInteger VS = powerModule(SK); if(VS.compareTo(SK) == 0) return true; return false;
}
}
```

IV. Result and Discussion

A. Comparative Analysis of Closest Conventional Approaches

For safe payment in financial transactional processes, DES, AES, and BLOWFISH are used. As shown in Table I, it also supports the suggested technique for reducing delay, encrypting time, energy usage, decryption time, and increasing throughput, and it was discovered that blowfish has the best score for each given constraint for the corresponding parameter.

The suggested method is evaluated based on average delay, throughput, encryption time, energy consumption, and decryption time, as shown in Fig. 1. Based on average delay, throughput, encryption time, energy consumption, and decryption time, Blow Fish (BF) is calculated using Data Encryption Standard (DES) and Advanced Encryption Standard (AES) methods. The nearest competitor is AES. The data confidentiality and integrity were supplied by AES. AES fails to reduce key complexity, and it readily compromises key privacy. Blow Fish (BF) has enhanced security for safe and fast financial transactions while also preventing fraud. Blow Fish (BF) improves 7.02 percent by 0.40 AD (Average Delay), 2.45 EC (Energy Consumption), 0.52 ET (Encryption Time), and 0.68 DT (Decryption Time) (Throughput). Finally, the article argues that the Blow Fish (BF) algorithm is the best.

B. *Comparative Analysis of Closest Conventional Approaches using RSA*

The RSA (1024) and RSA (2048) key length is used for secure and efficient payment transaction process. It also supports proposed methodology for minimizing the average delay, encryption time, decryption time as shown in Table II seen that RSA (2048) shows best score on every specified constraint for the respective parameter.

TABLE I. ENCRYPTION TIME, DECRYPTION TIME FOR E-COMMERCE AVERAGE DELAY, THROUGHPUT, ENERGY CONSUMPTION

Algorithm	**Encryption Time (s)**	**Decryption Time (s)**	**Average Delay (s)**	**Throughput (Kbps)**	**Energy Consumption (Joules)**
AES	0.314	0.321	49.1367	5.27	72.087
DES	0.434	0.451	48.7766	28.13	83.087
BF	0.262	0.253	48.7349	35.2	85.544

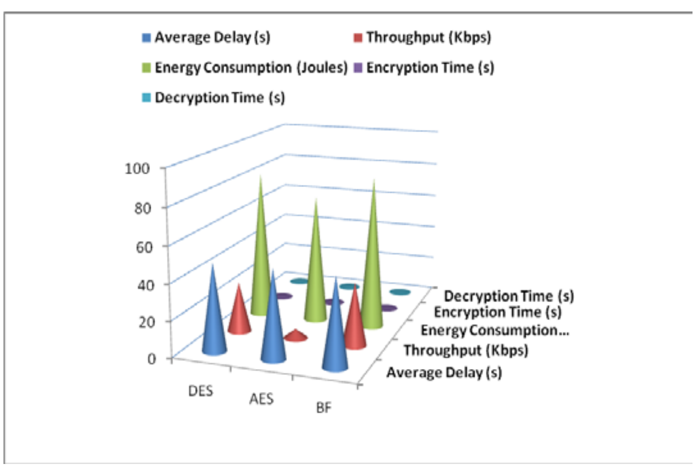


Fig. 1. Encryption Time, Decryption Time for e-Commerce Average Delay, Throughput, EnergyConsumption.

Based on the observation of Fig. 2 is estimated RSA algorithm utilizing key length of 1024 and 2048 bits. The RSA algorithm set the public exponent keys 1 to 3. Since, it offers better performance in key validations. The public exponent is generally utilized in constrained surroundings anywhere various validations have to occur. The results utilizing other exponents, key generations are more proficient, and validations are efficient. Finally, the performance of the RSA algorithm is good for security.

The cryptography method is used for the protection of information. These methods are utilized for decryption and encryption process over information. The Secure and Efficient Transaction System through RSA algorithm set the public exponent keys and it offers better performance in key validations. The public exponent is generally utilized in constrained surroundings, anywhere various validations have to incur security.

It concentrates on the secure and efficient payment transaction system and subsequently formulates new criterion functions for secure payment framework. It also prohibits hackers from tracing one's online transaction. The proposed system works based on Aadhar for secure payment transaction systems. Here, Aadhar act as a unique identification number (UID). The design narrated above in this chapter, is utilized to get a clear idea of the logical and analytical view of the proposed Encryption and Decryption process of Secure and Efficient Transaction System through Aadhar approach for real-time applications. The study explains the secure payment system through Aadhar ID for internet or web-based business using secure and efficient payment transaction with RSA algorithm process flow, mathematical steps. The system workflow diagrams and security of payment system details are displayed. It explains the implementation, logical, analytical, and mathematical view of secure and efficient payment transaction algorithm. Here RSA algorithm uses 2048 key size. It enhances protection for secure payment process.

TABLE II. DECRYPTION TIME, AVERAGE DELAY AND ENCRYPTION TIME

Algorithm	**AverageDelay (s)**	**EncryptionTime (s)**	**Decryption Time (s)**
RSA(1024)	2.5	4.52	3.12
RSA(2048)	2.9	5.13	4.13

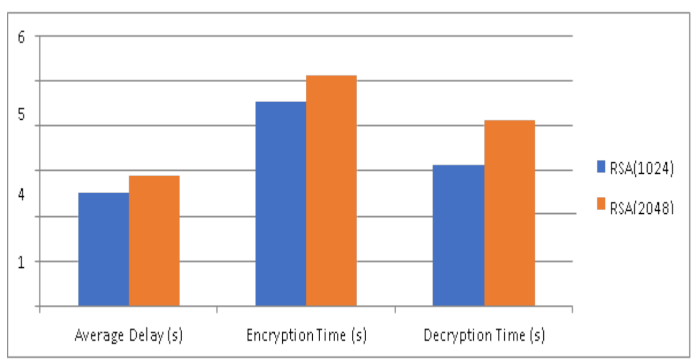


Fig. 2. Decryption Time, Average Delay and Encryption Time.

C. Simulation Result

The proposed technology represents a mathematical model to estimate the performance of Customer Success Rate, Mean of Transaction Time (MTT), and Reliability (R) to find the effectiveness of Secure and Efficient Payment Transaction with Aadhaar technology. The proposed technology expresses the payment process in secure and efficient way in e- commerce applications.

1) Customer success rate (CSR): The customer success rate refers to the person's discernment that utilizing a specific framework to deal with. It considered one of the most influential features behalf of new technology adoption like as the customer success rate. It has an effect on the attitude, self-viability, and instrumentality with utility. The impact of the perceived customer success rate has been exhibited in various database forums with various setting. The customer success rate is depending upon the attitude and aim whose details are explained in equation (1).

$$\text{CSR} = \frac{T_{F_R}+T_{J_u}}{T_F+T_{I_1}+FT_F+FT_{I_I}.} \tag{1}$$

Where, TF Transaction Fail, TT is Transaction True, FTF is False Transaction fail and FTT is False Transaction True. This derivation notifies that CSR mostly based on some features condition.

2) Mean of transaction time (MTT): The mean of transaction time is estimated by separating the total value of all exchanges by the number of exchanges or sales. This can be estimated on a day-day, month to month or yearly premise. The mean of transaction time is exhibited in equation (2).

$$MTT = \frac{\text{Total Transaction Value}}{\text{Number of Salele.}} \tag{2}$$

3) Reliability (R): A client's total number of partnerships with an organization and its products. Client experience management is a fundamental component of client relationship management. The client's overall impression of the organization and contributions is reflected in the overall experience. In equation, the suggested approach provides a mathematical model for Reliability (R) (3).

$$\text{Reliability} = \text{Performances of Process} - \text{Exceptation of Products} \tag{3}$$

Table III represents the Customer Success Rate, Mean of Transaction Time (MTT) and Reliability (R) with payment data for secure and efficient payment transactions. The proposed technology displays their average values for respective parameters with the respective data. The SEPT with Aadhaar technology is evaluated with the existing technology; namely, POS (Point of Sale), Mobile-based technology, Micro-ATM, Kiosk. According to Table III, it noticed that SEPT approach is best method compare than other existing method that have the best result on every specific aspect for the respective parameter.

Based on Table III, the proposed technique SEPT with Aadhaar ID computes Customer Success Rate, Mean of Transaction Time and Reliability for identifying the effectiveness of the technology. Fig. 3 shows the payment processing of CSR,R and MTT values. The proposed method is evaluated with POS payment technology, Mobile payment technology, Micro-ATM payment technology and Kiosk payment technology existing method on behalf of Customer Success Rate, Mean of Transaction Time, and Reliability. The POS technology depended on resolutions with or without Smart Card for payment process. A customer to offer their payment data in the transaction for a product or service utilizes it. However, there are many forms of POS frameworks utilized in numerous company types and it have privacy problems. The proposed method is providing efficient schemes for secure payment transaction process. Micro ATMs permit clients to perform fundamental economic exchanges utilizing only their unique number and fingerprint as unique evidence (according to a Bank Identification Number for inter-bank exchanges). However, it has manual access to all submitted document works. Kiosk payment technology is banking resolutions utilizing little bandwidth Internet for connectivity. However, it fails to maintain customer success rate and transaction time of payment process. The mobile payment technology is the nearest competitor on overall parameters. Mobile-based technology for transaction achievement will be able to meet their requirements and expectation much better than other delivery resolutions or platforms. However, it fails to maintain security and transaction time for all most cases. Proposed technique reduces 0.6 MTT (Mean of Transaction Time) and improves 0.9% (Reliability) and 0.9% of CSR (Customer Success Rate). Finally, the study concluded that the suggested SEPT technique outperforms all other evaluation matrices and input characteristics. As a result, the SEPT is the best technique in terms of overall characteristics.

TABLE III. AVERAGE CUSTOMER SUCCESS RATE (CSR), RELIABILITY (R) AND MEAN OF TRANSACTION TIME (MTT) FOR 50 USER PAYMENT PROCESSING WITH VARIOUS SIZE OF TRANSACTION RECORD SIZE

Technology	CSR	MTT	Reliability
POS	4.05	2.6	3.6
Mobile	1.8	1.5	2.24
Micro-ATM	3.4	2.4	3.2
Kiosk	2.9	2.1	2.51
SEPT	4.95	0.9	4.5

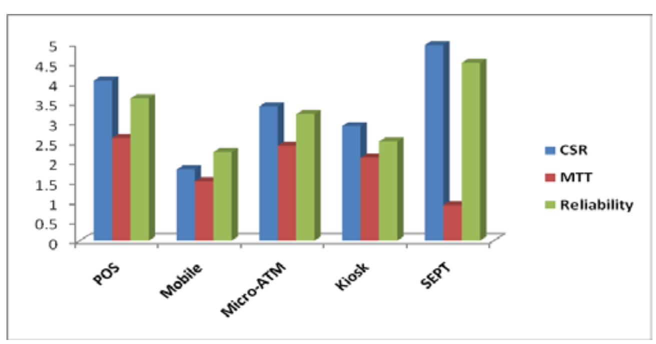


Fig. 3. Customer Success Rate (CSR), Reliability (R) and Mean of Transaction Time (MTT) for Payment Processing.

D. HSKT Technique

To prevent fraud, the HSKT investigates performance indicators to enhance security for safe and fast financial transaction processes in e-commerce applications. It provides evaluation metrics such as average delay, throughput, energy usage, encryption time, and decryption time.

1) Average delay: The timing difference between current data packets received as well as preceding data packets transmitted is evaluated using the HSKT method in the stages. In equation, the suggested approach is defined as a mathematical model for the average delay in equation (4). The average delay (AD) is calculated as follows:

$$\text{Average Delay} = \frac{\text{Pkt Received Time-Pkt Sent Time}}{\text{time}} \quad (4)$$

Where Pkt is Packet and Received time is received time.

2) Throughput: The throughput is an average of successful finance transactions and client record maintenances. The proposed method is described as a mathematical model for throughput in equation (5). Throughput is evaluated as:

$$\text{Throughput} = \frac{\sum_0^n \text{ Packets Received } (n)_* \text{ Packet Size}}{1000} \quad (5)$$

Where n is number of packets.

3) Energy consumption: Over the simulation period, the energy usage is really the difference between the new and residual energy. The energy consumption graph depicts the total amount of energy used in the Tx and Rx modes throughout the whole operation. Tx mode is used when one process transmits data to another process. Tx Energy of a process state is the amount of energy needed to transmit a data packet. Tx energy is calculated depending on the size of the data packet (in bits). When a process state receives the packet from some other process state, it is in Rx mode. RX energy is the amount of energy required to receive a data packet. In equation (6), the suggested approach is defined as a mathematical model (7). The following is an estimate of the Energy Consumption (EC):

Energy Tx = (330*data Size)/2*106 (6)

EnergyRx=(230*data Size)/2*106 (7)

Where Tx is initial energy and Rx is residual energy.

4) Encryption time (ET): The encryption time finds to changes over the plaintext into the ciphertext. The encryption time depended on the message chunk size and the key size and illustrated in milliseconds. It has direct impact on the execution of the HSKT technique. The HSKT technique has the less encryption time, to create the encryption technique responsive and quick. The proposed methodology is defined as a mathematical model in equation (8). The Encryption Time (ET) is estimated as:

$$EP = (I, CT\{EP_i\}_{i \in I}) \quad (8)$$

Where EP is an encoding process, and I is an attribute set. CT is ciphertext.

5) Decryption time (DT): The decryption time is a total time required to recuperate the plaintext from the ciphertext. For this purpose of HSKT technique is quick and responsive, it is attractive that the time taking for decoding is similar to the encryption time, and it is measured in milliseconds. A mathematical model in equations is used to describe the suggested approach (9). The following is an estimate of the Decryption Time (DT):

$$EP(EPi, ski) = EP(g, g)^{pi(0)s} \quad (9)$$

EP is an encoding process, and sk is a secret key.

With traditional methods, Table IV shows the Average Delay, Throughput, Energy Consumption, Encryption Time, and Decryption Time for various input restrictions. The suggested technique compared current methods to different types of algorithms such as DES, AES, and BF. On the basis of Table IV, it can be concluded that the suggested approach outperforms current methods. The method calculates Average Delay, Throughput, Energy Consumption, Encryption Time, and Decryption Time, as well as displaying average values for each parameter with specific data.

As shown in Fig. 4, the HSKT is computed using the average delay, throughput, encryption time, energy consumption, and decryption time. The suggested HSKT is computed using Data Encryption Standard (DES), Advanced Encryption Standard (AES), and Blow Fish (BF) techniques based on average delay, throughput, encryption time, energy consumption, and decryption time. The method that comes closest is BF. AES ensured the secrecy and integrity of the data. AES, on the other hand, fails to decrease key complexity and often compromises the secrecy of critical keys. The framework provides details about the programming environment, execution steps for the experimental process, tabular result, and comparative analysis of the proposed HSKT technique. It discovers the complete overview of how the proposed systems works, what is the necessary software required to implement this methodology. How these approaches are best when compared to other methods. The average delay, throughput, encryption time, energy consumption, and decryption time of the HSKT method are all assessed. The findings of the HSKT approach are compared to the nearest traditional methods in tabular and graphical form. HSKT enhances privacy for safe and fast financial transactions while also preventing E-commerce fraud. The HSKT improves by 59.8% and lowers 4.73 AD (Average Delay), 23.91 EC (Energy Consumption), 0.95 ET (Encryption Time), and 0.85 DT (Decryption Time) (Throughput). Finally, the study argues that the suggested HSKT technique outperforms all other evaluation matrices and input parameters.

TABLE IV. AVERAGE DELAY, THROUGHPUT, ENERGY CONSUMPTION, ENCRYPTION TIME AND DECRYPTION TIME FOR 50 USERS WITH VARIOUS DATA SIZES

Algorithm	**Average Delay (s)**	**Throughput(Kbps)**	**Energy Consumption (Joules)**	**EncryptionTime (s)**	**DecryptionTime (s)**
DES	48.7766	28.13	83.087	0.434	0.451
AES	49.1367	5.27	72.087	0.214	0.221
BF	48.7349	35.2	85.544	0.262	0.253
EHSKT	13.334	95.16	48.168	0.119	0.136

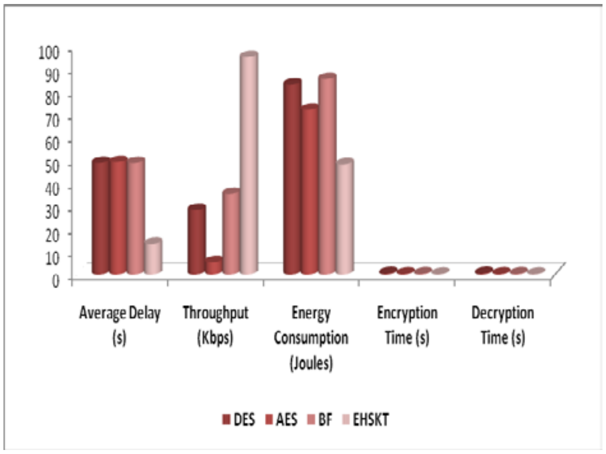


Fig. 4. Average e-Commerce Delay, Throughput, Energy Consumption, Encryption Time and Decryption Time.

V. CONCLUSION

The World Wide Web is the most important resource in contemporary business, since it opens up new business possibilities. E-commerce refers to an online store where goods is sold or purchased electronically. The High-Speed Key Transmission (HSKT) Technique was created to provide secure financial transactions, protect product content privacy, and prevent fraud. The privacy system keeps track of and controls the content profiles of merchants and clients. To prevent fraudulent behavior, the proposed approach increases throughput of safe besides effectual payment transactions, as well as the privacy of product content. The proposed method shows better results in performance parameters. In future, the method needs to focus more on transmission delay.

REFERENCES

[1] Rajan, M. T., Vincent, A. J., Prakash, G., Prakash, N., & JU, R. M. Computerized RTBS System. International Journal of Emerging Engineering Research and Technology Vol. 2, No. 2, pp. 25-29, 2014.

[2] Dass, R., “Unique Identity Project in India: A Divine Dream or a Miscalculated Heroism?” Indian Institute of Management, 2011.

[3] Yadav, V., “Unique identification project for 1.2 billion people in India: can it fill institutional voids and enable 'inclusive' innovation?” Contemporary Readings in Law and Social Justice, Vol. 6, No. 1, pp. 38, 2014.

[4] Chander, S., & Kush, A., “Unique Identification Number and E-Governance Security," International Journal of Computing and Business Research, Vol. 1, No. 1, 2010.

[5] Jadav, M. B., Desai, M. A., Patel, M. F., & Patel, M. R., “Cloud Computing E-Voting: A Technical Review," Int. J. Res. Emerg. Sci. Technol, Vol. 2, pp. 8-13, 2015.

[6] Bahuguna, A., “FIRe: Firefox for Computer Security Incident Reporting and Coordination," IITM Journal of Management and IT, Vol. 6, No. 1, pp. 3-11, 2015.

[7] Varshney, D., &Goyal, D., “UID based Mobile Money Implementation in Rural Areas of India," International Journal of Research in Engineering & Advanced Technology, Vol. 1, No. 6, 2014.

[8] Chatterjee, P., &Nath, A., “Smart Computing Applications in Railway Systems-A case study in Indian Railways Passenger Reservation System," International Journal, Vol. 3, No. 4, 2014.

[9] AnandShende, OmkarGurav, SwapnilShirode, PiyushGovekar, and S.N.Zaware, “Secure Unique Identification using Encrypted Storage in NoSQL Database," International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, No. 4, 2016.

[10] Henderson-Sellers, B., Firesmith, D. G., Bock, C., & Odell, J., “ESIGN," Journal of Object-oriented Programming, Vol. 11, 1998.

[11] Sahu, S. K., &Kushwaha, A., “Performance Analysis of Symmetric Encryption Algorithms for Mobile Ad hoc Network," International Journal of Emerging Technology and Advanced Engineering, Vol. 4, No. 6, pp. 619-624, 2014.

[12] G. R. Patel and K. Panchal, “Hybrid Encryption Algorithm”, International Journal of Engineering Development and Research (IJEDR), vol. 2, iss. 2, pp. 2064-2070, 2014.

[13] Xueying Zhang,et al.,“Energy efficiency of encryption schemes applied to wireless sensor networks”, Security and Communication Networks, John Wiley & Sons, Ltd., 2011.

[14] R.Sharmila and V.Vijayalakshmi, “Hybrid Key Management Scheme for Wireless Sensor Networks”, International Journal of Security and Its Applications, Vol.9, No.11, pp.125-132, 2015.

[15] S. Gajbiye, et al., “A Survey Report on Elliptic Curve Cryptography”, International Journal of Electrical and Computer Engineering, vol.1, no.2, pp.195-201, 2011.

[16] D. Malan, et al., “A public-key infrastructure for key distribution in TinyOS based on elliptic curve cryptography”, in Proc. 1st IEEE Int. Conf. Sensor Ad Hoc Communication Network, pp. 71-80, 2004.

[17] Y. Liu,et al., “PKC based broadcast authentication using signature amortization for WSNs”, IEEE Trans. Wireless Communications, vol. 11, no. 6, pp. 2106-2115, 2012.

[18] P. Porambage, et al.,“Pauthkey: A pervasive authentication protocol and key establishment scheme for wireless sensor networks in distributed applications” International Journal of Distributed Sensor Network, vol. 2014.

[19] D. Kim and S. An, “PKC-based DoS Attacks-Resistant Scheme in Wireless Sensor Networks”, IEEE Sensors Journal, 2016.

[20] M. Panda, “Security in Wireless Sensor Networks using Cryptographic Techniques”, American Journal of Engineering Research (AJER), vol.03, iss.01, pp-50-56, 2014.

[21] Akhilesh, K. B., & Srinivasan, R., “Driving the economy through innovation and entrepreneurship: Emerging agenda for technology management”, 2013.

[22] C. Mukhopadhyay, A. Gurtoo, P. Ramachandran, P. P. Iyer, M. Mathirajan, & M. H. B. Subrahmanya (Eds.). Springer India. AUTHENTICATION, UIDAI, 2011.

[23] Choi, Younsung. "Cryptanalysis on Privacy-aware two-factor Authentication Protocol for Wireless Sensor Networks." Indonesian Journal of Electrical Engineering and Computer Science, Vol. 8, no. 2, pp. 296-301, 2017.

[24] Singh, Pooja, and R. K. Chauhan. "A Survey on Comparisons of Cryptographic Algorithms Using Certain Parameters in WSN." International Journal of Electrical and Computer Engineering, Vol. 7, no. 4, pp. 2232, 2017.

[25] Gayathri, P., Syed Umar, G. Sridevi, N. Bashwanth, and Royyuru Srikanth. "Hybrid Cryptography for Random-key Generation based on ECC Algorithm." International Journal of Electrical and Computer Engineering (IJECE), Vol. 7, no. 3, pp. 1293-1298, 2017.

[26] Xiao, Y. Compensation Method of Electronic Commerce Data Transmission Delay Based on Fuzzy Encryption Algorithm. *Mobile Netw Appl* (2022).

[27] Hassan, M.A.; Shukur, Z.; Hasan, M.K. An Efficient Secure Electronic Payment System for E-Commerce. *Computers* 2020, *9*, 66.

[28] Alam, S.S.; Ali, M.H.; Omar, N.A.; Hussain, W.M.H.W. Customer satisfaction in online shopping in growing markets: An empirical study. Int. J. Asian Bus. Inf. Manag. 2020, 11, 78–91.

[29] Noor Ardiansah, M.; Chariri, A.; Rahardja, S.; Udin, U. The effect of electronic payments security on e-commerce consumer perception: An extended model of technology acceptance. Manag. Sci. Lett. 2020, 10, 1473–1480.

[30] Satar, N.S.M.; Dastane, O.; Ma'arif, M.Y. Customer value proposition for E-Commerce: A case study approach. Int. J. Adv. Comput. Sci. Appl. 2019, 10, 454–458.

Machine Learning Algorithms for Document Classification: Comparative Analysis

Faizur Rashid[1]
Department of Computer Science
Haramaya University
Almaya, Ethiopia

Suleiman M. A. Gargaare[2]
Department of Computer Science
University of Hargeisa
Somaliland

Abdulkadir H. Aden[3]
Department of Computer Science
Bule Hora University
Ethiopia

Afendi Abdi[4]
Department of Software Engineering
Haramaya University
Almaya, Ethiopia

***Abstract*—Automated document classification is the machine learning fundamental that refers to assigning automatic categories among scanned images of the documents. It reached the state-of-art stage but it needs to verify the performance and efficiency of the algorithm by comparing. The objective was to get the most efficient classification algorithms according to the usage of the fundamentals of science. Experimental methods were used by collecting data from a sum of 1080 students and researchers from Ethiopian universities and a meta-data set of Banknotes, Crowdsourced Mapping, and VxHeaven provided by UC Irvine. 25% of the respondents felt that KNN is better than the other models. The overall analysis of performance accuracies through various parameters namely accuracy percentage of 99.85%, the precision performance of 0.996, recall ratio of 100%, F-Score 0.997, classification time, and running time of KNN, SVM, Perceptron and Gaussian NB was observed. KNN performed better than the other classification algorithms with a fewer error rate of 0.0002 including the efficiency of the least classification time and running time with ~413 and 3.6978 microseconds consecutively. It is concluded by looking at all the parameters that KNN classifiers have been recognized as the best algorithm.**

Keywords—Document classification; machine learning algorithms; text classification; analysis

I. Introduction

Document classification is a vital research area or topic since the establishment of digital documents used widely [1]. This is one of the major parts of the manual effort. More tech companies outsource the job and business processes people have to sort the documents of the packages manually. There is the availability of the hundreds of documents types. Nowadays, text classification is an important task because of the very large amount of text documents required to deal with day-to-day activities. In general, document classification can be classified as topic-based document classification and document genre-based classification. Topic-based document categorization can be classified documents according to their topics [2]. Also, texts can be written in many different genres, for example, academic papers, advertisement updates, political news, and movie reviews. Genre referred to the way a text is made, the way it was modified, the identification of language used, and the type of listeners to whom it is addressed. Existing studies on genre classification found that task differs from the categorization of topic-based [3]. Commonly, most data based on genre classification were collected from the newspaper, web, noticeboards, and live broadcasts.

The classification is an information retrieval from the metadata, manually classified, or via an automatic classifier retrieving information from the content. As manually classifying documents can be a time-consuming and inconsistent task. It is usually not beneficial on a larger scale. Instead, automatic Document Classification is suggested to solve the categorization of retrieved information, because of the automatic process for larger systems [4].

Although some degree of automation is achieved that helps to search through keywords and expressions the accuracy and efficiency of such solutions are questionable and not satisfactory. An approval of efficiency matters to show the user's satisfaction that needs assessing and analyzing machine learning algorithms. It increases the satisfaction of automated document classification. The document classification contains many concepts as Fig. 1 shows.

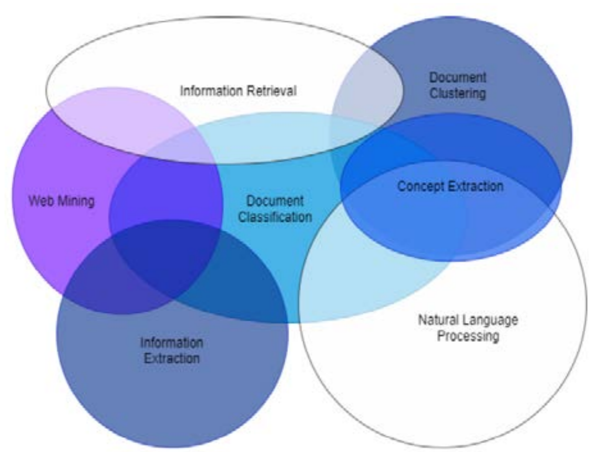


Fig. 1. By [5] Venn Diagram of the Text Mining Area.

This study is organized as follows, In Section 1, Introduction is presented. Section 2 presents different types of Classification Algorithms. Section 3 describes the related works. Section 4 introduces the Methodology. Section 5 shows the Results and Discussions of the study. Conclusions and recommendations are discussed in Section 6 and Reference is cited in Section 7.

II. Classifcation of Algorithm

A. Logistic Regression

According to [6], the binary outcomes either something happens or nothing happens, yes or no, pass or fail, living or dead are calculated using logistic regression. In addition, two factors were described: independent and dependent variables, which were examined to determine the binary outcome based on whether the outcomes were numeric or categorical. The independent variables might be categorical or numeric, but the dependent variable was categorical all the time and stated in equation (1):

$$P\ (A=1|B) \text{ or } P\ (A=0|B) \quad (1)$$

Whereas, A and B calculate the probability of the dependent variable and independent variable consequently.

Positive or negative connotation {0, 1} word or tree, grass and flower which was common object contained in a photo calculated by probability of each object between 0 and 1 R. Wolf, 2021.

B. K-Nearest Neighbor (KNN)

The objects in KNN were categorized as [7], defined, and filled by selecting numerous labeled training examples with the shortest distance from each other. The k-nearest neighbor classification method stood out with its simplicity and commonly used techniques for text categorization, Even if classifying items took longer when a large number of training samples were provided. Even with multi-categorized documents, this strategy works well for regulating categorization jobs. KNN should be chosen manually, and some of them can be determined by calculating the distance between each test object and all of the training samples.

C. Decision Trees

A decision tree is a tree in which internal nodes are labeled with terms, according to [8]. The branch was labeled by numerical data, while the leaf was labeled by categories. The "divide and conquer" method was employed in decision tree concepts. [8], further stated that each node in a tree corresponded to a collection of cases. Terms should check whether these were under the same label or not, according to this entire training example. Then, if not the same label, select partitioning terms from the pooled classes of documents with similar values for the term and place each of them in a distinct subtree.

D. Random Forest

The random forest algorithm was the enlargement of the decision tree, as mentioned in [6]; constructing the actual world an axis of decision tree from training data in the real world. It essentially normalizes data so that it bonds to a nearby tree on the data scale. Random forest prototypes are important because they solve the decision tree's problem of unnecessarily "pushing" data points into a category.

E. Naive Bayes Algorithm

A Naive Bayes classifier [9] is a "simple probabilistic classifier based on Bayes' Theorem and strong independence assumptions." It calculates the document's subsequent probability of being assigned to multiple classes and assigns the document to the class with the highest subsequent probability, employing the autonomous feature as the probability model. As a result, the existence of one feature in a classification task has no bearing on the existence of other features.

F. Perceptron

A threshold function serves as an activation function in a perceptron, which is an artificial neuron. Assume an artificial neuron with input signals x_1, $x_2 \ldots$, x_n and associated weights w_1, $w_2 \ldots$, w_n with w_0 for constant, [10]. If the output of a neuron is assumed to be a perceptron, the equation is as follows in (2).

$$O(\mathrm{x1}, \mathrm{x2}, \ldots, \mathrm{xn}) = \begin{cases} 1, \text{ if Wo } + \text{ W1X1 } + \ldots + \text{ WnXn } > 0 \\ -1, if\ W0 + W1x1 + \cdots + WnXn \leq 0 \end{cases} \quad (2)$$

G. Support Vector Machines

Support Vector Machines (SVM) are utilized for test classification which is a supervised classification algorithm. Text classifiers must cope with a large number of features [11] with high-dimensional input space. It is also capable of handling vast feature spaces and employed in place of adequate protection that is not based on the number of features. Furthermore, the majority of text categorization issues are linearly separable. Therefore, SVMs were developed to locate and search for linear separators.

H. Gaussian Naive Bayes

Gaussian Naïve Bayes is also known as Gaussian NB which is based on the Gaussian Normal Distribution and supports continuous data, as specified by [12]. It's a Naive Bayes variation to compute continuous data, continuous values were often associated with each class and distributed according to the normal distribution.

III. Related Research

Automated document classification is the machine learning fundamental that refers to assigning automatic categories among scanned images of the documents. KNN is the most researched topic by most researchers, with the most important aspect being to perform a detailed study over survey applications that were performed by implementing introductory data mining books and survey reports that were documented by [13], which proposed many improvements of KNN algorithms for implementing data classification. Another noteworthy study by [14] was on the weighted KNN classification method based on various symbolic characteristics, in which the distance was measured and calculated before being depicted in the form of tables to produce real-valued distances from symbolic domains that also represent features. The authors claim that the suggested method outperforms existing algorithms such as KNN because it was tested on three different application areas, with the key advantage being training speed and ease of implementation.

[15], an optimization function based on the "leave-out-out cross-validation" technique and "greedy hill-climbing technique" and introduced three major "decision tree algorithms" was published that focused on an adjustment of weight while implementing KNN for identifying optimum weighted vector by using an optimization function that was based on the "leave-out-out cross-validation" technique and "greedy hill-climbing technique" and introduced three major "decision tree Many other studies conducted extensive surveys of applications based on various decision tree algorithms in the fields of machine learning and data mining technologies [16].

SVMs are a type of classification algorithm that works by examining a feature space and attempting to build a hyperplane to divide data points belonging to distinct classes [16][17]. It worked by employing a kernel function to translate data onto a higher-dimensional space and defining the hyperplane. Although SVMs are binary classifiers by nature, they can be adjusted for multiclass issues by employing pairwise classification, which treats a problem as a series of binary problems. The illustration in support vector machines is comprised of self-learning kernels, where the survey was demonstrated in SVMs and their mathematical foundation. Major parts of SVMs were applied to the implementation of text categorization in other works [18].

Many scholars used Decision Trees to solve the challenge of automatic affect identification. This was a simple classifier that used data observations and map observations to make class ownership decisions [16] [17]. It works by repeatedly querying a test instance for more information about the classes to which it may belong using a set of if-then rules.

IV. Methodology

A. Data Collection

The interviews were taken as a data collection tool from nine first-generation Ethiopian universities either face-to-face or via the medium of wire. We focused to question from research students, senior teachers, and individual researchers.

Our objective was to get the most efficient classification algorithms according to usage and level of understanding of the fundamentals of science. 720 research students, 225 teachers, and 135 individuals of computing domain as the sum of 1080 were questioned using the Question: “which algorithm is better for document classification, in terms of simplicity and accuracy?” We gave them eight different classification algorithms including; KNN, SVM, Naive Bayes,

Gaussian NB, Perceptron, Random Forest, Logistic Regression, and Decision Tree.

All of the interviewees did not answer the question, which makes a response rate of 92% (appx). 25% of the respondents felt that KNN is the most used algorithm and 20% of the respondents said that SVM is better in their experience. 15% agreed that Perceptron is the most popular and used, another 15% suggested that Gaussian NB is easy for them and the rest have recommended as described in Fig. 2.

Therefore, the study was compared with four different classification algorithms that the respondents felt were the most used classifiers with a response rate of 15% or above are, KNN, SVM, Perceptron, and Gaussian NB which were included in the analysis.

B. Experiments

This study was used on a whole meta-data set called Banknotes, Crowdsourced Mapping, and VxHeaven provided by UC Irvine [19] holds information about the different categories of data. We used the Irvine dataset to get the exact performance of every algorithm. Overall, functions and commands were used in the platform of Python 3.7.

The detail of the data is mentioned in Table I which is employed to perform the k-Fold cross-validation technique for training and testing the datasets were used. Here, the value is assumed as k=10 which shows the unbiased result of the datasets. KFold() Scikit-learn class used with an argument of the number of splits whether to shuffle the sample. We created an instance that splits a dataset into k folds split returned each group of train and test sets after calling the split() function. Index of a way returned into the original data samples of observation to use for train and test sets on each repetition.

Fig. 3 illustrates the first 10 rows in the banknotes dataset that were used to train and test the algorithm. Similarly, all the datasets were trained and tested.

1) Datasets: All the files are images in which banknotes are pictures of various banknotes and measuring different properties of currency and in particular, they categorized each of these banknotes as either counterfeit banknotes or not counterfeit (authentic).

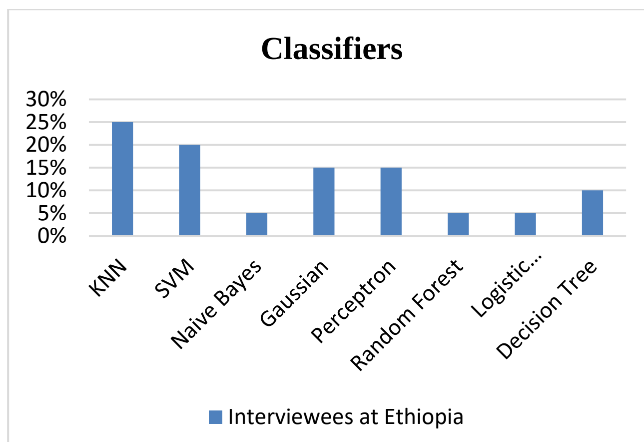


Fig. 2. Classifiers for Interviewees in Ethiopia.

TABLE I. Datasets And Data Size

Datasets	Data Size	Text Size
Banknotes	1372	10-fold Cross-Validation
Crowdsourced Mapping	10546	10-fold Cross-Validation
VxHeaven	2598	10-fold Cross-Validation

Crowdsourced Mapping Images were taken through satellites of different land cover classes of the farm, forest, orchard, and water being categorized. VxHeaven is the detection of malware observed through various features of the text in the Operating System.

2) Machine learning algorithms were evaluated using a resampling procedure called k-Fold cross-validation. A dataset is split into o number of groups known as k-Fold. The score with high variance may change the idea based on data to fit the algorithm or overestimate the skill of the algorithm due to biased data if the value of k is selected poorly. Here, k=10, which is less biased of the algorithm than another method to split the train or test result.

Each row of the dataset represents the banknote and has four different input values. These inputs have an output value of 0 or 1. 0 means a genuine (authentic) bill and 1 means a counterfeit bill.

So, this study used supervised learning to begin to predict some sort of function that can take four values as input and predict the output would be the algorithm was built using Python language, Jupyter Notebook as a compiler, and libraries including, Pandas, Scikit-Learn, and Matplotlib.

V. Result And Discussion

In evaluations, data were divided into two different sets, training and testing datasets using the k-Fold model. Training data sets were used to build the algorithm (classifier) and then tested the classifier for the prediction of the goal.

A. Confusion Matrix

A confusion Matrix is also known as an Error Matrix that is used to define the analyses of classification algorithms on a set of test data for which the true values are well known. It is a table that has two dimensions; actual value and predicted value as Table II illustrates.

TABLE II. Truth Table of Confusion-Matrix

	Predicted		
Actual		***No***	***Yes***
	No	*TN*	*FP*
	Yes	*FN*	*TP*

TN: True-Negative, TP: True-Positive, FN: False-Negative, FP: False-Positive.

```
df = pd.read_csv("banknotes.csv")
print(df.head(10))
```

```
   variance  skewness  curtosis  entropy  class
0  -0.89569   3.00250 -3.606700 -3.44570      1
1   3.47690  -0.15314  2.530000  2.44950      0
2   3.91020   6.06500 -2.453400 -0.68234      0
3   0.60731   3.95440 -4.772000 -4.48530      1
4   2.37180   7.49080  0.015989 -1.74140      0
5  -2.21530  11.96250  0.078538 -7.78530      0
6   3.94330   2.50170  1.521500  0.90300      0
7   3.93100   1.85410 -0.023425  1.23140      0
8   3.97190   1.03670  0.759730  1.00130      0
9   0.55298  -3.46190  1.704800  1.10080      1
```

Fig. 3. Sample of First 10 Rows of Banknotes.

So, the confusion matrixes of mentioned classification algorithms are illustrated in the below matrixes.

Confusion Matrix 1.1 KNN $\begin{matrix} 363 & 1 \\ 0 & 322 \end{matrix}$

Confusion Matrix 1.2 SVM $\begin{matrix} 389 & 7 \\ 0 & 290 \end{matrix}$

Confusion Matrix 1.3 Perceptron $\begin{matrix} 382 & 1 \\ 9 & 294 \end{matrix}$

Confusion Matrix 1.4 Gaussian NB $\begin{matrix} 348 & 39 \\ 60 & 239 \end{matrix}$

The evaluation metrics are classified in Accuracy, Precision, Recall, Error Rate, and F-Score which are described in paragraphs (1), (2), (3), (4), and (5) consecutively below that what variable function for their parameters calculated. The summary of all the evaluations of the analyzed algorithm is shown in Table III.

1) Accuracy: It is close to the measured value to the actual (true) value.

$$\text{Accuracy} = \left(\frac{\text{TP+TN}}{\text{Total Tuples in Test Dataset}}\right) \tag{3}$$

The accuracy of these four algorithms; KNN, SVM, Perceptron, and Gaussian NB were scored 99.85%, 98.98%, 98.54%, and 85.57% respectively.

2) Precision: It is an evaluation analysis technique that finds the closeness of the measured values to each other.

$$\text{Precision} = \left(\frac{\text{TP}}{\text{Predicted Yes}}\right) \tag{4}$$

Performance of precisions of algorithms KNN, SVM, Perceptron, and Gaussian NB were 0.996, 0.976, 0.996, and 0.856 respectively.

3) Recall: The ratio of all correctly predicted positive predictions was measured using Recall.

$$\text{Recall} = \left(\frac{\text{TP}}{\text{Actual Yes}}\right) \tag{5}$$

The algorithms, KNN, SVM, Perceptron, and Gaussian NB reached 100%, 100%, 97%, and 79% respectively.

4) Error rate: It is an evaluation analysis technique that calculates the number of all incorrect predictions divided by the total number of the datasets. The worst error rate is 1.0 and the best error rate is 0.0.

$$\text{Error Rate} = 1 - \text{Accuracy} \tag{6}$$

KNN, SVM, Perceptron, and Gaussian NB algorithm errored 0.002, 0.011, 0.015, and 0.145 respectively of error rate declares.

5) F-Score: It is an evaluation analysis technique that calculates the harmonic mean of precision and recall.

$$\text{F-Score} = \left(\frac{2(\text{P}*\text{R})}{\text{P+R}}\right) \tag{7}$$

Where *P* is Precision and *R* is Recall.

TABLE III. EVALUATED TECHNIQUES WITH PARAMETERS FOR THE ALGORITHM

Evaluation Technique	Analyzed Algorithm			
	KNN	SVM	Perceptron	Gaussian NB
Accuracy	99.85%	98.98%	98.54%	85.57%
Precision	0.996	0.976	0.996	0.856
Recall	100%	100%	97%	79.90%
Error Rate	0.002	0.011	0.015	0.145
F-Score	0.997	0.987	0.982	0.826

All algorithms: KNN, SVM, Perceptron, and Gaussian NB were marked 0.997, 0.987, 0.982, and 0.826 respectively in the F-Score evaluation shows.

B. Evaluation Time

Time is one of the important parameters to check the efficiency of any algorithm. It helps to test a comparative algorithm. Timing of classification of document and execution were tested and analyzed using datasets as shown in Table IV and Table V. Performance of execution was measured in python using the timeit() function. Although python is slower than C++ libraries.

The optimization Table IV shows the least classification time and running time of KNN is ~413 and 3.6978 consecutively which is the lowest of another algorithm.

TABLE IV. EXECUTION TIME OF CLASSIFICATION TIME OF THE ALGORITHM

File Size (MB) (Datasets)	Classification Time			
	KNN	*SVM*	*Perceptron*	*Gaussian-NB*
0.22	~400	~409	~413	~411
0.4	~422	~428	~437	~438
0.32	~417	~419	~433	~434
Average	~413	~418.67	~427.6667	~427.6667

TABLE V. EXECUTION TIME OF RUNNING TIME OF THE ALGORITHM

File Size (MB) (Datasets)	Running Time			
	KNN	*SVM*	*Perceptron*	*Gaussian-NB*
0.22	3.6827	3.6828	3.7055	3.7076
0.4	3.7155	3.716	3.7188	3.7189
0.32	3.6952	3.6954	3.6967	3.6966
Average	3.6978	3.69807	3.707	3.7077

VI. CONCLUSION

In this comparative study, we compared the performance and efficiency of various machine learning classification algorithms. KNN, SVM, Perceptron, and Gaussian NB using a meta-data set created by UC Irvine as an experiment. Authors estimate that the used data set is the meta-size of the data whose training and testing procedure was taken using k-Fold methods and experiments. K-Fold cross-validation technique for training and testing of the datasets was used where the value of k=10 was assumed. In addition, the study focused on identifying a better algorithm for document classification that executed well on different meta-data sets. However, it was assessed that the accuracies of the tools depend on the data set used. Also, noted that the classifiers of a special group did not perform with equal accuracies in terms of the overall performance accuracies algorithm. KNN performed better than the other classification algorithms with a fewer error rate of 0.0002 the efficiency of the least classification time and running time with ~413 and 3.6978 microseconds consecutively. It is concluded that KNN classifiers have been recognized as the best algorithm for document classification with a percentage accuracy of 99.85%, recall value of 100%, and f-Score of 0.997. There would be platform variations of the algorithm that might be the case of study in the future.

ACKNOWLEDGMENT

This paper is written by us using meta-data. We thank each other for the collaboration, collection, and execution of data that we passed. Authors are working in the area of Artificial Intelligence and substitute areas like machine learning, and deep learning including data sciences. These sciences are heavily dependent on the tools and algorithms used and tested in the data of different fields. Authors considered carrying out more into the domain. This work is not supported by any funder or our affiliated organization.

REFERENCES

[1] M. Ikonomakis, S. Kotsiantis, and V. Tampakas, "Text Classification using Machine Learning Tanique", WSEAS TRANSACTION on COMPUTERS, vol. 4, no. 8, pp. 966-974, 2005.

[2] Y. Yang, "An Evaluation of Statistical Approaches to Text Categorization", Journal of Information Retrieval, vol. 1, pp. 67-88, 1999.

[3] R.Johnson, and T Jhang, "Supervised and Semi-Supervised test Categorization using lstm for region Embedding", preprint, arXiv:1602.02373, 2016.

[4] Goller, Loning, Will, and Wolf, "Automatic Document Classification", International Symposium for Information, pp. 145-162, 2000.

[5] M.T.U.U. Hugo Moriz, "A Comparative Study of Machine Learning Algorithm for Document Classification", Uppsala Unpublished. 2020.

[6] K. Kowsari, K Jafari Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, "Text Classification Algorithm: A Survey Information" vol. 10, no 4, pp. 150, 2019.

[7] Tam Santoso, and R. Setiono, "A Comparative Study of Centroid Based, Neighborhood-based and Statistical Approaches for Effective Document Categorization", ICPR 02 Proceedings of the 16th International Conference on Pattern Recognition (ICPR'02), vol. 4, no. 4, 2002, pp. 235-238, 2002.

[8] Mou R. Men, G. Li, Y. Xu, L.Zhang, R. Yan, and Z.Jin, "Natural Language inference by tree-based convolutional and heuristic matching", arXiv:1512.08422, 2015.

[9] I. Rish, "An Empirical Study of the Naïve Bayes Classifier", Proc. Of the IJCAI-01, Workshop in Empirical Methods in Artificial Intelligence, CiteULike-article-id-352583, 2001.

[10] V. N. Krishnachandran, "Machine Learning", Vidya Center for Artificial Intelli9gence Research, India, 2018.

[11] Pratiksha P. Pawar, and S.H. Gawande, "A Comparative Study on Different Types of Approaches to Text Categorization", International Journal of Machine Learning and Computing, vol. 2, pp. 423-426, 2021.

[12] S. Agarwal, B. Jha, T. Kumar, M.Kumar, and P. Ranjan, "Hybrid of Naïve Bayes and Gaussian Naïve Bayes for Classifiaction: A Map Reduace Approach", vol. 8, no. 6S3, pp. 266-268, 2019.

[13] R. Agrawal, T. Imielinski, and A. N. Swami, “Database Mining: A Performance Perspective”, IEEE Trans, Knowledge and Data Engineering, vol. 6, pp. 914-925, 1993.

[14] Jiawei Han, Jian Pei, and Micheline Kamber, “Data Mining, Concepts and Techniques”, 2nd Edition: Morgan Kaufmann, 2006.

[15] H.Z., “A Sort Introduction to Data Mining and its Application”, IEEE, 2011.

[16] B. Rama, P. Parveen, H. Sinha, and T. Chaudhary, “A Study on Causal Rule Discovery with PC Algorithm”, International Conference on Infocom Technologies and Unmanned System (ICTUS), Dubai 2017.

[17] S. Nagaparameshwara Chary, and B. Rama, “A research Travelogue on Classification Algorithms using R. Programming”, International Journal of Recent Technologies and Engineering (IJRTE), vol. 8, no. 4, pp. 9155-9158, 2019.

[18] P. Praveen, B. Rama, and Uma N. Dulhare, “A Study on monthelic Divisive Hierarchical Clustering Method”, International Journal of Advanced Scientific Technologies Engineering and Management Sciences, vol. 3, 2017.

[19] L. Volker, “Bank Notes Database- University of Applied Sciences”, Ostwestfalen-Lippe: Lemgo: UC Irvine, 2012.

A Novel Evolving Sentimental Bag-of-Words Approach for Feature Extraction to Detect Misinformation

Yashoda Barve[1]
Suryadatta College of Management
Information Research & Technology
Pune, India

Jatinderkumar R. Saini[2]*
Symbiosis Institute of Computer Studies and Research
Symbiosis International (Deemed University)
Pune, India

Kaushika Pal[3]
Sarvajanik College of Engineering & Technology
Surat, India

Ketan Kotecha[4]
Symbiosis Centre for Applied Artificial Intelligence
Symbiosis International (Deemed University)
Pune, India

***Abstract*—The state-of-the-art misinformation detection techniques mainly focus on static datasets. However, a massive amount of information is generated online and the websites are flooded with this legitimate information and misinformation. It is difficult to keep track of this changing information and provide up-to-date accurate status of webpages giving either legitimate information or misinformation. Therefore, to keep the features up-to-date, authors have proposed evolving sentimental Bag-of-Words approach. This involves, updating sentimental features every time the new or changed web contents are read. This process accumulates the sentimental features at different time intervals that can be utilized to detect misinformation in URLs and upgrade the status of the webpage with timely information. Apart from sentimental features, other state-of-the-art features viz. syntactical, Part-Of-Speech Tagging (POST), and Term-Frequency (TF) are updated in a timely manner and utilized to detect misinformation. The model performed well with the support vector machine showing an accuracy of 80% while the decision tree classifier showed less accuracy of 56.66%.**

Keywords—Misinformation detection; healthcare; incremental learning; sentiment analysis; bag-of-words

I. Introduction

In an era of the information explosion, the volume of information available on the web has increased rapidly. Social media and the web as a whole have become an important source of information for people due to its easy access, low cost, availability, and popularity [1], [2]. At every moment millions of people access the internet to share information and interact over social media. This information has many hidden patterns inside. In one of the surveys in 2012 from the USA, it was found that 49% of people have connected to social media and the web to disseminate information, while in 2016 over 62% have accessed social media regularly to browse news articles [3], [4]. Therefore, the data disseminated on the web and social media have become the topic of interest for many researchers.

This large volume of proliferated data can cause the spread of misinformation. The misinformation or false information is inaccurate or incorrect information that is confirmed with existing evidence [4]. The false information may appear in the form of fake news, rumor, satire news, hoaxes, misinformation, disinformation, and opinion spam [3]. Thus, a massive amount of false information has been observed to spread in an uncontrolled fashion over the web.

The misinformation propagated via the web gets spread quickly, resulting in a widespread impact on people, business, healthcare, politics, and all other aspects. For example, in the 2016, USA presidential elections, the spread of misinformation resulted in public shootings. Also, rumors about the outbreak of the Zika virus were portrayed as a bioweapon and that the Zika vaccine is developed to depopulate the earth had been the considerable activity on Twitter and Facebook [5]. Hence, the persistent effects of misinformation are harmful and can contribute to violent conflicts which may lead to the death of people, especially in the healthcare domain. A substantial quantity of fictitious and fabricated information passes through the web creating an untrustworthy feeling among the public. Thus to raise the reliability of the web and social networks and reduce the detrimental effects of misinformation, detecting and combating false information has become the need of time and is highly recommended [6], [7].

Detecting misinformation is a challenging task due to three key reasons. First, to verify information with factual data; the second is the unavailability of structured data in a specific domain; third, adapting to the changing web data [4], [8]. The researchers have been keen on finding the answers to these challenges and detecting misinformation from raw text data. The detailed process of detecting misinformation is explained in the related work section of the paper.

The sentimental features are exclusively used in the literature to detect misinformation, fake news, rumors, etc.

*Corresponding Author

The studies have observed that the sentiment polarity features in misinformation and true information are totally different. For misinformation, there are more negative words that may be due to guilt or tension while writing wrong information leading to very strong negative sentiment [9]. In a study of identifying prominent features to detect rumor propagation authors found that rumors are dominated by sentiments like anxiety, and uncertainty [10]. In [11], authors considered sentiment analysis as one of the important factors to assess information credibility.

Although there is a sufficient machine learning-based model in the literature to detect misinformation using sentiment features, the authors did not find any work dealing with newly emerging data in the healthcare domain. As the data change gradually during the period of time the sentimental words may also alter i.e. increase or decrease and it may affect the percentage of misinformation. As information in the real world is time-critical and newly emerging data needs to be adapted accordingly by the model, the efforts are required to furnish the newly arriving data, thus making it a challenging task [4], [12]. Hence, a technique like incremental learning is most suitable to handle this new chunk of data arriving at different time intervals due to its less memory requirement and short training time [13].

Therefore, in this research authors have used incremental learning to record the sentimental changes occurring in the data by generating sentimental “Bag-of-Words” in the healthcare domain and thus detect the change in percentage of misinformation at different time intervals. The aim is to find the URLs showing legitimate content initially and which later spread misinformation at different time intervals and vice versa and detect a change in the percentage of misinformation.

Hence, the objectives of this study are as follows: 1) Generate a sentimental “Bag-of-Words” for healthcare domain, 2) Incrementally evolve the sentimental “Bag-of-Words” using sentiment relevance score function, 3) Generate fact-check dataset of 100 URLs related to the healthcare domain, 4) Extract features to detect misinformation 5) Use state-of-the-art machine learning classifiers to classify the URLs as legitimate or non-legitimate, 6) Identify and study the change in URLs at Time T1 and T2 in terms of misinformation and sentiments.

The remaining sections of the paper are structured as follows: Section 2 defines the literature survey in the field of misinformation detection using sentiment analysis and incremental learning in the healthcare domain, Section 3 explains the methodology, Section 4 discourses the results and discussion and Section 5 describes the conclusion and Section 6 elaborates on limitations and future enhancements.

II. Literature Survey

This section discusses the outline of related work, which are focused on: (i) misinformation detection techniques, (ii) Incremental Learning and sentiment analysis approaches and their applications, (iii) application of incremental learning in the healthcare domain (iv) application of sentiment analysis in the healthcare domain. Although tasks (ii), (iii), and (iv) are effectively examined independently, there is a shortage of research that would merge them and come up with a sentiment analysis-based approach for detecting misinformation, which would be capable of controlling changes in newly arriving data. In other words, there is no implementation of a model which would detect misinformation in the healthcare domain with incremental learning and sentiment analysis on a real-world dataset. Similar efforts by [14], incorporate behavioral features with linguistic features to detect misinformation. However, the proposed research categorically focuses on using incremental learning to accommodate newly arriving information into the model along with the sentiment and linguistic features so that the real-time changes are incorporated in the model.

A. *Misinformation Detection Techniques*

The researchers are keenly driving in with a lot of efforts to detect misinformation. Several supervised, unsupervised and semi-supervised machine learning and deep learning-based algorithms exist which can detect misinformation. Machine learning is the component of Artificial Intelligence that can acquire knowledge and progress with past experiences [3], [4].

The basic process of detecting misinformation using machine learning techniques contains the following steps: data collection, data pre-processing, manual classification in case of the gold standard dataset using expert knowledge, feature extraction, and classification. The latest machine learning models from the literature have used features like content/syntactical features, user-specific features, sentiment features, grammatical features, linguistic features, image-specific features. Upon identifying and extracting the required feature set, the significant machine learning classifiers are used to perform the classification viz. Support Vector Machine, Naïve Bayes, Random Forest, Decision Tree, Logistic Regression, k-Nearest Neighbors [3]. Feature Extraction is a significant aspect of misinformation detection as the effectiveness of machine learning algorithms is mainly dependent on feature extraction. According to [11] sentimental features are considered to be crucial in detecting the credibility of contents. Hence, authors have considered sentimental features as central features in this study.

B. *Incremental Learning and Sentiment Analysis*

Sentiment analysis is the process of detecting sentiment polarity in terms of positive and negative words in a text. However, the user perception of data in terms of sentiment polarity may change as time passes. The static models designed require updates to adapt to the changes. The incremental learning affords to adapt to these changed instances of data with the efficient cost of computations. Hence, incremental learning and sentiment analysis together can tackle the problem of sentiment shifts and new word generations [15], [16]. In the literature, researchers have used sentiment analysis to detect commuter emotions using incremental learning by computing emotion density at different time intervals [17]. Sentiment analysis with incremental learning benefits to make a more granular analysis of the user’s opinion. In another research, lexicons are built dynamically and evolved incrementally for Arabic text. Thus, incremental learning helps to build and expand the lexicons dynamically [18].

C. Incremental Learning in Healthcare

Incremental learning in healthcare was used in literature to enhance the predictive accuracy and decrease the computation time of classification many disease diagnosis systems have used incremental learning techniques, incremental SVM, and incremental PCA. Incremental learning was applied to the numeric datasets to classify diabetes disease using a combination of support vector machine, Expectation-Maximization, and Principal Component Analysis [19]. An incremental knowledge-based health management information system is designed to fetch and process the electronic medical records and expert guidance to accommodate the knowledge system [20]. In another approach, for newly arriving chunks of data, the network was trained using incremental learning by remembering the earlier patterns of data. This approach helped the researchers to discover heart rate changeability patterns in mobile healthcare services [21]. Hence, the newly arriving chunk of data of electronic medical records is captured using incremental learning techniques.

D. Sentiment Analysis in Healthcare Domain

The sentiment analysis determines the dominant sensitive views from the text to know about the author's approach towards sentiments as either positive, negative, or neutral. In the state-of-the-art techniques, [22] the researchers collected medical terms from the Unified Medical Language System (ULMS) and clustered similar documents with medical terms using Latent Dirichlet Allocation (LDA) to form condition topics. The sentiment analysis helped the user to identify the feelings of other users with similar conditions. The researchers have been using sentiment analysis to discover the tendency of sentiments about vaccination like measles vaccine, and HPV vaccine through social media platforms. Sentiment libraries like VADER, Google Cloud Sentiments are used for analysis. Machine learning techniques such as TF-IDF, K-means clustering, and topic modeling using Latent Dirichlet Allocation (LDA) are applied for sentiment classification [23]–[25]. These investigations help the health experts or researchers to identify the various reasons for vaccine scares or hesitancy and take corrective actions to improve vaccine uptake. Sentiment analysis is proved as efficient and consistent while dealing with huge amounts of data [11]. However, the existing sentiment polarity libraries are restricted to general-purpose sentiments and not towards domain-based sentiments. To acquire improved accuracy of sentiment analysis, domain-based sentimental libraries are essential. However, due to the scarcity of such domain-based sentiment libraries in the healthcare domain, it is essential to generate sentimental Bag-of-Words in the healthcare domain [18].

E. Research Contributions

With the proposed system platform following are the research contributions.

1) Generation of sentimental Bag-of-Words (BoW) using incremental learning.

2) A novel sentiment-based incremental machine learning model to detect misinformation from web URLs.

3) Analyzing the performance of the proposed model.

III. Methodology

Fig. 1 shows the diagrammatic representation of the proposed model. The detailed techniques are discussed in the following section.

A. Data Collection and Pre-Processing

This section explains the methodology used to collect the dataset for misinformation detection. To fetch the web URLs, a Google search engine is used. Google is extensively used for collecting information online. According to the statistics, 60% of desktop users and 90% of mobile or tablet users use Google while searching [11]. Initially, the authors identified and shortlisted 25 keywords related to the healthcare domain by identifying the most frequently occurring words in the healthcare domain. Table I displays the 25 keywords used for data collection. The various combinations of these keywords are used in the query for the Google search engine and the top 100 URLs are fetched every time, resulting in a group of 500 URLs. These 500 URLs were a mixture of healthcare URLs, non-healthcare URLs, and duplicate URLs. Thus, the authors manually segregated the URLs into two categories healthcare and non-healthcare, and also removed duplicate URLs and the final set of 100 URLs was used for analysis. Further, authors along with domain experts performed fact-checking of URLs and manually categorized 100 URLs into legitimate and non-legitimate based on contents, contact information available on the website. Also, the authors have considered URLs from the CoAID [26] dataset after verifying them with domain experts. The authors observed that maximum legitimate URLs are from .gov, and .edu domains while non-legitimate URLs are from other domains. Further, URLs are pre-processed to get cleaned data. The process begins with the grouping of data from web URLs and generated words. Regular expressions are used to remove punctuations, and numbers from the data. The pre-processing technique sanitization was used to transform the words into lower case, followed by the stop-word removal technique which is applied to reduce the corpus size by removing unnecessary words not providing any useful information. Also, all the single characters and duplicate data are removed to further minimize the corpus size and build quality information. In addition, 2 wordlists are developed manually containing Bag-of-Words (BoW) with sentiment polarity of 452 positive words, and 341 negative words by using healthcare URLs after carefully understanding the meaning of each word, especially the core medical terms by referring to MeSH (Medical Subject Headings) dictionary which explains the biomedical terminologies [27] and taking expert opinion about the usage of those words. Table II displays the sample words list of manually labeled sentiments of words.

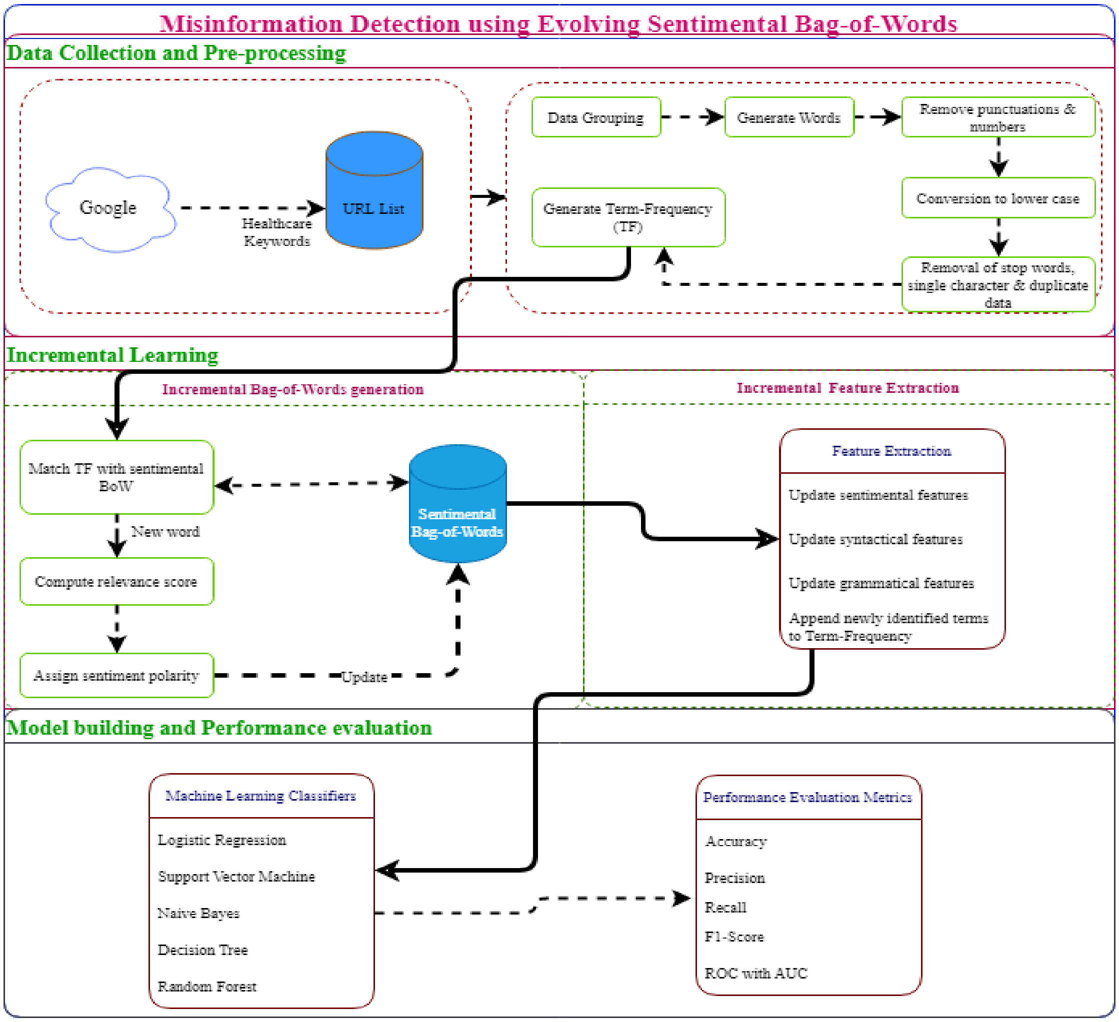


Fig. 1. Process of Healthcare Related Misinformation Detection Model.

TABLE I. LIST OF KEYWORDS FOR GOOGLE SEARCH

Sr. No.	Keyword Name	Sr. No.	Keyword Name
1	Doctor	14	Consultation
2	Medical	15	Ambulance
3	Pharmacy	16	Medicare
4	Healthcare	17	Medi-claim
5	Chronic	18	Surgery
6	Disease	19	Damage
7	Treatment	20	Specialist
8	Prevention	21	Kill
9	Therapy	22	Chemical
10	Medication	23	Organ
11	Illness	24	Infection
12	Clinic	25	Vaccine
13	Cancer		

TABLE II. SAMPLE OF MANUALLY DEVELOPED WORD LIST

Sr. No.	Positive Words	Negative Words
1	Survivors	Abnormal
2	Lifesaving	Scary
3	Protect	Depression
4	Painless	High-risk
5	Specialists	Illegal
6	Safeguard	Fatigue
7	Relieving	Crime
8	Efficacy	Die
9	Support	Funeral
10	Raised	Ruth
11	Inspiring	Fear
12	Careful	Threatening
13	Cure	Damaged
14	Ability	Abolished
15	Better	Harm

B. Incremental Learning

Upon getting cleaned data, the term-frequency is generated and matched every term with the sentimental "Bag-of-Words". Here, each term is referred to as a word. These newly arriving words are assigned sentiment polarity using the sentiment relevance score function; thus, updating the sentimental bag-of-words incrementally. This function generates synonyms and antonyms of the newly arriving word using the WordNet library of the NLTK tool. To find the closeness between newly arriving words and their synonyms and antonyms authors have used Jaccard similarity co-efficient. The Jaccard similarity coefficient of sets A and B is given by $|A \cap B| / |A \cup B|$. It is the fraction of the intersection of the number of elements of A and B to the number of elements of their union. It is used in various applications like mail filtering [28] to find the co-occurrence of values of all pairs of words in an email. Further, the threshold value was chosen after carefully analyzing the synonyms and antonyms generated. For example, for the word "good" the relevant words above 0.5 were honourable=0.9, beneficial=1.0, effective=1.0, etc. while below 0.5 were like satisfactory=0.4, fair=0.33, etc. Hence, the threshold of 0.5 was decided manually to assign sentiment polarity to a new word. However, for the first batch of 10 URLs, it was observed that Bag-of-Words was upgraded with the most relevant words, but during the second batch of 10 URLs, there was an addition of a few irrelevant words. Hence, authors again sensibly tried identifying the best applicable threshold value which could fetch only relevant synonyms or antonyms of the word. This was possible with a change in the threshold of 0.95 instead of 0.5. After using a new threshold value of 0.95, for the next batches of URLs, it showed a good collection of relevant words. Therefore, choosing the right threshold value can improve the collection of the most relevant sentimental "Bag-of-Words" in the healthcare domain. This process continues in an incremental fashion. The relevance score function is defined as shown in Fig. 2. Next, incremental learning is adopted to update features. This is achieved by fetching the URLs at time T2. Thus, URL TF has computed again at time T2 and these terms are matched with bag-of-words to get the final list of terms. These newly arriving terms are appended to TF. Also, based on changed bag-of-words the sentimental, syntactical, and grammatical features are extracted and updated. The syntactical features consist of word count and sentence count. Grammatical features contain a number of verbs, nouns, adverbs, adjectives, and pro-noun. While sentiment-based features include a number of positive, negative words, the percentage of positive and negative words. The final set of 11 features is as listed in Table III.

C. Model Building and Performance Evaluation

To build the model, five different state-of-the-art supervised classification algorithms Logistic Regression (LR), Support Vector Machine (SVM), Naïve Bayes (NB), Decision Trees (DT), and Random Forest (RF) are used. Further, the performance of the model is evaluated based on confusion matrix, accuracy, precision, recall, f1-score, ROC, and AUC.

```
Input: word
Output: sentiment score for the word
Algorithmic Steps:
1. Procedure relevance_score (word)
2. Generate synonyms and antonyms for a given word
3. For every synonym 'w'
4.     jd = Jaccard Similarity Coefficients (word, w)
5.       if jd <= 0.95, append w with "word" sentiment
6.      else if jd > 0.95, append w with "word" antonym
7. For every antonym w
8.      id = Jaccard Similarity Coefficients (word, w)
9.        if jd <= 0.95, append w with "word" antonym
10.  elseif jd > 0.95, append w with "word"
           sentiment
11. End Procedure
```

Fig. 2. Function to Compute a Relevance Score.

TABLE III. DESCRIPTION OF EXTRACTED FEATURES

Feature Type	Feature	Description
Syntactical Features	w_count	Total number of words
	s_count	Count the number of sentences in the document
Sentiment Features	pw_count	Positive Words Count
	nw_count	Negative Words Count
	perc_pos	Percentage of Positive Words
	perc_neg	Percentage of Negative Words
	sen_net_score	Sentence Net Score
Grammatical Features (POST)	verb_count	Count of verbs in the document
	nouns_count	Count of nouns in the document
	adverbs_count	Count of adverbs in the document
	adj_count	Count of adjectives in the document
	pronoun_count	Count of Pronouns in the document

IV. Experimental Results and Discussion

This section discusses the experiments accomplished to evaluate the performance of the newly developed features. The results of Classification with the combination of feature categories are shown in Table V, along with the combination of features set. From Table IV, the syntactical grammatical and sentiment features when applied together, the models Support Vector Machine and Logistic Regression showed 80% and 76.6% of accuracy. Fig. 3 shows the performance of all the models based on precision, recall, and F1-score. Fig. 4 represents the percentage of misinformation in terms of sentiment features for 30 URLs of test data. It is observed that URL 2, URL 4, URL 8, URL12, URL18, URL19, and URL26 showed the maximum percentage of misinformation. Fig. 5 displays the word cloud for objectionable words on a cancer web page. It is seen that objectionable words like symptoms, smoking, chemotherapy, disease, death, risk factor play a vital role in detecting misinformation. Also, Fig. 6 represents the word cloud for legitimate words which can be used to detect true URLs. The words like covid, vaccine, treatment, vaccine trials sound positive and refer to legitimate information. Fig. 7 displays the Receiver Operating Characteristic (ROC) curve with 10-fold cross-validation and Area Under the Curve (AUC). Algo_1 represents the first fold, Algo_2 represents the second fold, and so on.

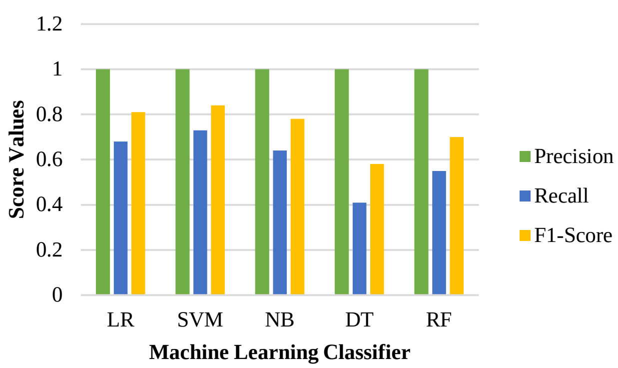


Fig. 3. Performance of Classification Models based on Precision, Recall and F1-Score (%).

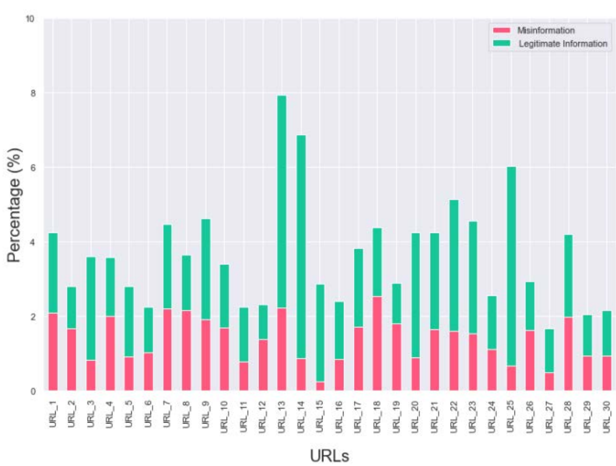


Fig. 4. Percentage of Legitimate Information and Misinformation in Healthcare Domain.

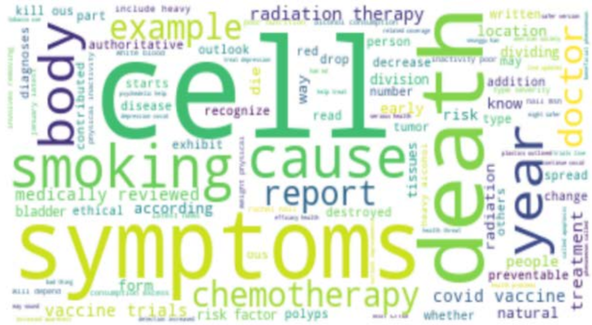


Fig. 5. Word Cloud for Objectionable Words.

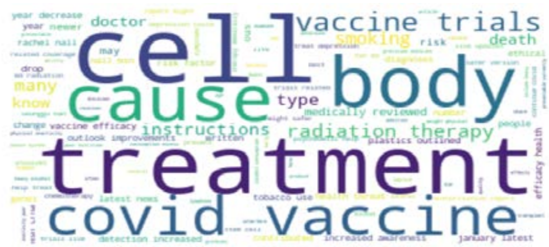


Fig. 6. Word Cloud for Legitimate Words.

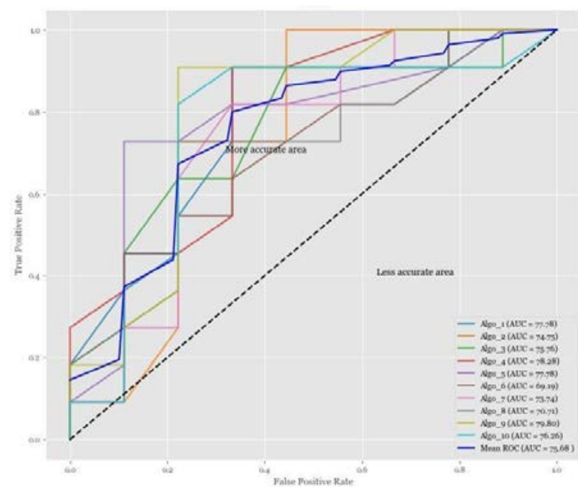


Fig. 7. Receiver Operating Characteristic (ROC) Curve with 10-Fold Cross-Validation and Area Under the Curve (AUC).

The analysis of the percentage of URLs updated with positive or negative words and the URLs which remain consistent shows that 23% of URLs changed in the textual contents after the second iteration whereas 77% of URLs remained unchanged. From among the 23% URLs which got updated in the second iteration it was observed that in 43% of URLs the percentage of misinformation has changed, i.e. either the misinformation has increased or decreased. In 57% of URLs, although there was a change in textual contents, the percentage of misinformation remained consistent. Fig. 8 shows the statistics of misinformation changed from among 43% of URLs. It can be observed that four URLs (2, 4, 6, and 9) had negative change, which means the percentage of misinformation has reduced. However, six URLs (1, 3, 5, 7, 8, and 10) have positive changes, depicting an increase in the percentage of misinformation.

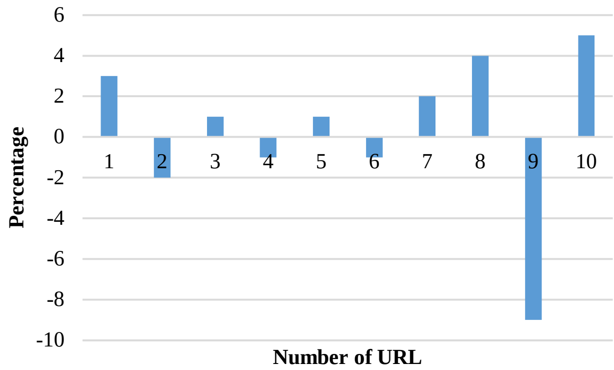


Fig. 8. Fluctuation in the Percentage of Misinformation in Updated URLs.

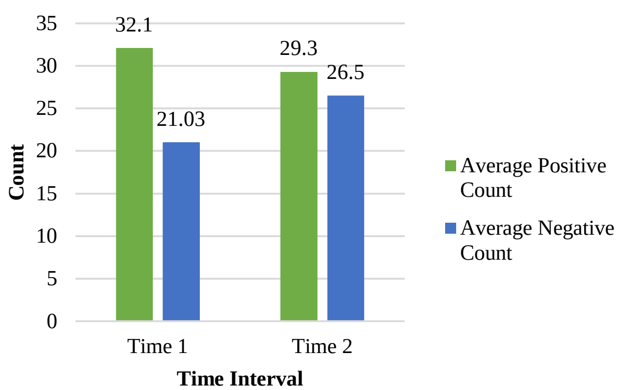


Fig. 9. An Example Legitimate URL showing Average Positive and Negative Word Count after First (TIME 1) and Second (TIME 2) Iteration.

Fig. 9 depicts the updated positive and negative word count in the URL. The experimental results show that the URL has more positive words and fewer negative words in the first iteration whereas showed an increase in the number of negative words in the second iteration and a decrease in the number of positive words. The experimental results among updated URLs i.e. URLs with changes in textual contents in the second iteration show 61% of average positive count and 39% of the average negative count. It is observed that 39% of negative changed URLs showed a change in misinformation. Fig. 8 shows this fluctuated change in misinformation. Fig. 10 and Fig. 11 display the sentiment median values and standard deviation of sentiment values for legitimate and non-legitimate URLs respectively. It is observed that in legitimate URLs the positive sentiment values are more compared to negative sentiment values. In non-legitimate URLs, the negative sentiment values count is superior to positive values. Fig. 12 depicts the increase in Bag-of-Words for the first batch of 10 URLs. Fig. 13 shows the increase in Bag-of-Words after 10 batches of 10 URLs each. Also, the same URLs were scraped again on 3rd April 2021 to observe the increase or decrease of words in sentimental "Bag-of-Words". Thus, it can be seen from Fig. 14 that the Bag-of-Words evolved over time reaching up to 1069 words at T1 and 1663 words at T2. Fig. 15 shows the total number of positive and negative words generated at time T1 and dT2, respectively.

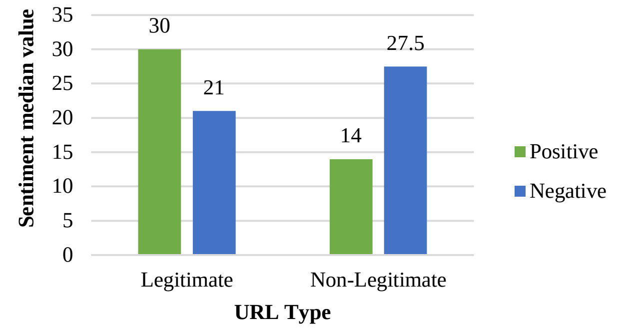


Fig. 10. Sentiment Median Values for Legitimate and Non-Legitimate URLs.

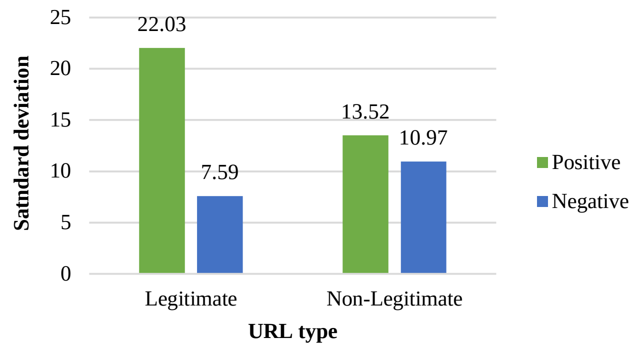


Fig. 11. Standard Deviation of Sentiment Values for Legitimate and Non-Legitimate URLs.

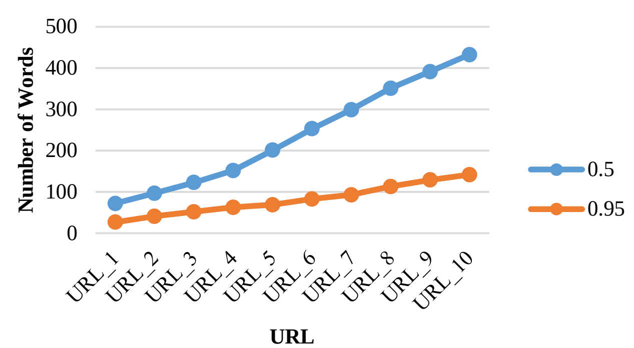


Fig. 12. Increase in Bag-of-Words based on Threshold Value.

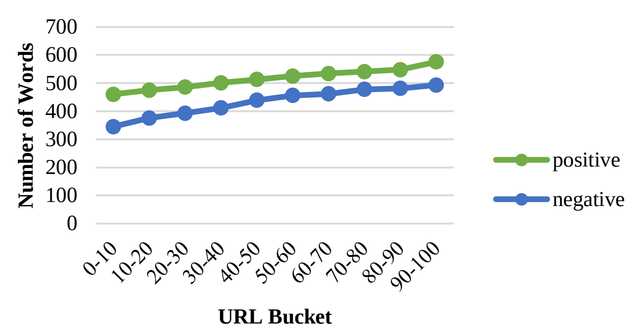


Fig. 13. Increase in Sentiment Words per 10 URLs.

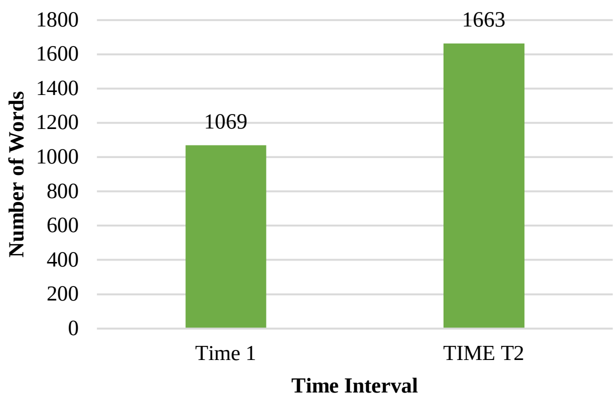


Fig. 14. Total Number of Words Generated at Time T1 and T2.

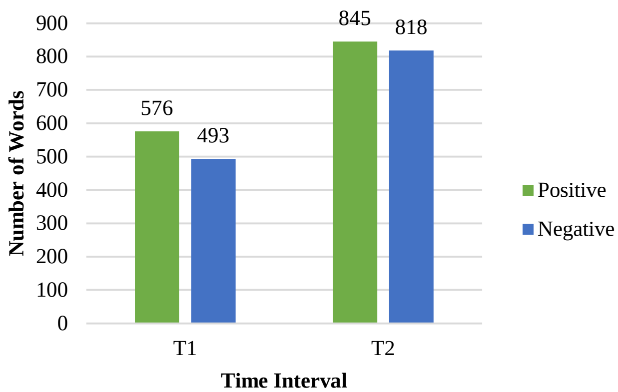


Fig. 15. Total Number of Positive and Negative Words at Time T1 and T2.

A. Theoretical Implications

While there is a reasonable volume of earlier research to detect misinformation on the web in the healthcare domain, these methods do not adapt to the recurrently changing data and are thus unable to encounter the sentimental changes occurring in the data. Theoretical Implications of the study are as follows. First, the study describes a facet of incremental learning and sentiment analysis to generate evolving sentimental "Bag-of-Words" in the healthcare domain. This theoretically generated knowledge can be suitably implemented in the other domains like politics and business which are more susceptible to misinformation generation. Hence, this is an important theoretical contribution. Secondly, the study defines a new aspect by developing sentiment analysis and incremental learning models for false information detection. The study presents a collaborative process of incremental learning and sentiment analysis to classify healthcare-related web URLs as legitimate or non-legitimate. The authors have experimentally verified and confirmed the importance of sentimental features and incremental learning in false information detection. These results have exposed new factors of false information change and its detection, proposing an incremental learning and sentiment features-based model. These observations have certainly augmented the knowledge base of misinformation detection in the healthcare domain.

B. Practical Implications

Practically, results indicate that this evolving Bag-of-Words can be used to find the sentiment intensity of the articles in the healthcare domain. The sentiment shifts of the patients can be studied using "Bag-of-Words" in the healthcare domain. Thus, these outcomes could help Researchers, Scientists, Doctors, Patients, and Organizations to employ them to detect false misinformation and mitigate it.

C. Comparison with Existing Research with SVM Model

Table IV shows the comparison of studies in healthcare domain. It can be seen that proposed model have outperformed the existing techniques.

TABLE IV. COMPARISON OF PROPOSED MODEL WITH EXISTING STUDIES

Ref. No.	Domain	F1-score
[29]	Healthcare	0.75
[30]	Healthcare	0.64
Proposed Model	**Healthcare**	**0.84**

TABLE V. CLASSIFICATION RESULTS OF MODELS BASED ON ACCURACY (IN %)

Features Combination	Logistic Regression	SVM	Naïve Bayes	Decision Tree	Random Forest
Syntactical Features	37.10%	37.10%	40.10%	51.61%	56.00%
Syntactical Features + Sentiment Features	73.33%	76.66%	73.33%	53.33%	63.33%
Syntactical Features +Sentiment Features+ Grammatical Features	76.6%	80%	73.33%	56.66%	66.66%

V. Conclusion and Future Enhancements

In this research, we have proposed a method to detect healthcare-related misinformation using sentiment-based incremental learning approaches. This framework can help to analyze and identify misinformation using central features such as sentiments and other features consisting of syntactic and grammatical features. A relevance score function assigns the sentiment to the newly arriving words using the Jaccard distance measure. The model was built with five machine learning classifiers viz. LR, SVM, NB, DT, and RF. The results show that SVM has outperformed other classifiers with an accuracy of 80%. The decision tree classifier showed the lowest accuracy of 56.66%. Thus, it was observed that incremental learning plays a crucial role in determining the consistency of correct information by detecting changes in the percentage of misinformation at different time intervals.

VI. Limitations and Future Enhancements

The research has several benefits in terms of finding misinformation in the healthcare domain, but it does have certain limitations too. Firstly, though authors have used the BoW approach, the n-gram approach can be used to correctly predict the sequence of occurrence of words. The second limitation is the dataset size. It is a need to increase the dataset to thousands to verify the change in accuracy.

In the future, the authors are willing to dynamically assign threshold to the sentiment relevance score which may lead to more accurate Bag-of-Words, and later build the model using incremental clustering techniques to handle the real-time data.

References

[1] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media," *Big Data*, vol. 8, no. 3, pp. 171–188, 2020, doi: 10.1089/big.2020.0062.

[2] X. Zhou and R. Zafarani, "A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities," *ACM Comput. Surv.*, vol. 53, no. 5, 2018, doi: 10.1145/3395046.

[3] P. Meel and D. K. Vishwakarma, "Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities," *Expert Syst. Appl.*, vol. 153, 2020, doi: 10.1016/j.eswa.2019.112986.

[4] A. Habib, M. Z. Asghar, A. Khan, A. Habib, and A. Khan, "False information detection in online content and its role in decision making: a systematic literature review," *Soc. Netw. Anal. Min.*, vol. 9, no. 1, 2019, doi: 10.1007/s13278-019-0595-5.

[5] Y. Wang, M. McKee, A. Torbica, and D. Stuckler, "Systematic Literature Review on the Spread of Health-related Misinformation on Social Media," *Soc. Sci. Med.*, vol. 240, 2019, doi: 10.1016/j.socscimed.2019.112552.

[6] S. Lewandowsky, U. K. H. Ecker, C. M. Seifert, N. Schwarz, and J. Cook, "Misinformation and Its Correction: Continued Influence and Successful Debiasing," *Psychol. Sci. Public Interes. Suppl.*, vol. 13, no. 3, pp. 106–131, 2012, doi: 10.1177/1529100612451018.

[7] C. Carvalho, N. Klagge, and E. Moench, "The persistent effects of a false news shock," *J. Empir. Financ.*, vol. 18, no. 4, pp. 597–615, 2011, doi: 10.1016/j.jempfin.2011.03.003.

[8] I. Ahmad, M. Yousaf, S. Yousaf, and M. O. Ahmad, "Fake News Detection Using Machine Learning Ensemble Methods," *Complexity*, vol. 2020, 2020, doi: 10.1155/2020/8885861.

[9] Y. Yang, L. Zheng, J. Zhang, Q. Cui, Z. Li, and P. S. Yu, "TI-CNN: Convolutional Neural Networks for Fake News Detection," 2018, [Online]. Available: http://arxiv.org/abs/1806.00749.

[10] S. Kwon, M. Cha, K. Jung, W. Chen, and Y. Wang, "Prominent features of rumor propagation in online social media," *Proc. - IEEE Int. Conf. Data Mining, ICDM*, pp. 1103–1108, 2013, doi: 10.1109/ICDM.2013.61.

[11] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on Twitter," *Proc. 20th Int. Conf. Companion World Wide Web, WWW 2011*, no. January, pp. 675–684, 2011, doi: 10.1145/1963405.1963500.

[12] A. Choudhary and A. Arora, "Linguistic feature based learning model for fake news detection and classification," *Expert Syst. Appl.*, vol. 169, 2021, doi: 10.1016/j.eswa.2020.114171.

[13] Y. Barve and P. Mulay, "Bibliometric Survey on Incremental Learning in Text Classification Algorithms for False Information Detection," *Libr. Philos. Pract.*, vol. 2020, no. November, pp. 2388–2392, 2020.

[14] Y. Zhao, J. Da, and J. Yan, "Detecting health misinformation in online health communities: Incorporating behavioral features into machine learning based approaches," *Inf. Process. Manag.*, vol. 58, no. 1, 2021, doi: 10.1016/j.ipm.2020.102390.

[15] T. Doan and J. Kalita, "Sentiment analysis of restaurant reviews on yelp with incremental learning," in *Proceedings - 2016 15th IEEE International Conference on Machine Learning and Applications, ICMLA 2016*, 2017, pp. 697–700, doi: 10.1109/ICMLA.2016.158.

[16] S. Mishra, J. Diesner, J. Byrne, and E. Surbeck, "Sentiment analysis with incremental human-in-the-loop learning and lexical resource customization," in *HT 2015 - Proceedings of the 26th ACM Conference on Hypertext and Social Media*, 2015, pp. 323–325, doi: 10.1145/2700171.2791022.

[17] D. Nallaperuma *et al.*, "Online Incremental Machine Learning Platform for Big Data-Driven Smart Traffic Management," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 12, pp. 4679–4690, 2019, doi: 10.1109/TITS.2019.2924883.

[18] S. Sanagar and D. Gupta, "Unsupervised Genre-Based Multidomain Sentiment Lexicon Learning Using Corpus-Generated Polarity Seed Words," *IEEE Access*, vol. 8, pp. 118050–118071, 2020, doi: 10.1109/ACCESS.2020.3005242.

[19] M. Nilashi, O. Bin Ibrahim, A. Mardani, A. Ahani, and A. Jusoh, "A soft computing approach for diabetes disease classification," *Health Informatics J.*, vol. 24, no. 4, pp. 379–393, 2018, doi: 10.1177/1460458216675500.

[20] T. Ali *et al.*, "The Intelligent Medical Platform: A Novel Dialogue-Based Platform for Health-Care Services," *Computer (Long. Beach. Calif).*, vol. 53, no. 2, pp. 35–45, 2020, doi: 10.1109/MC.2019.2924393.

[21] T. H. N. Vu, N. Park, Y. K. Lee, Y. Lee, J. Y. Lee, and K. H. Ryu, "Online discovery of Heart Rate Variability patterns in mobile healthcare services," *J. Syst. Softw.*, vol. 83, no. 10, pp. 1930–1940, 2010, doi: 10.1016/j.jss.2010.05.074.

[22] F. C. Yang, A. J. T. Lee, and S. C. Kuo, "Mining Health Social Media with Sentiment Analysis," *J. Med. Syst.*, vol. 40, no. 11, 2016, doi: 10.1007/s10916-016-0604-4.

[23] V. Raghupathi, J. Ren, and W. Raghupathi, "Studying public perception about vaccination: A sentiment analysis of tweets," *Int. J. Environ. Res. Public Health*, vol. 17, no. 10, 2020, doi: 10.3390/ijerph17103464.

[24] X. Luo, G. Zimet, and S. Shah, "A natural language processing framework to analyse the opinions on HPV vaccination reflected in twitter over 10 years (2008 - 2017)," *Hum. Vaccines Immunother.*, vol. 15, no. 7–8, pp. 1496–1504, 2019, doi: 10.1080/21645515.2019.1627821.

[25] J. Du, J. Xu, H. Y. Song, and C. Tao, "Leveraging machine learning-based approaches to assess human papillomavirus vaccination sentiment trends with Twitter data," *BMC Med. Inform. Decis. Mak.*, vol. 17, no. Suppl 2, 2017, doi: 10.1186/s12911-017-0469-6.

[26] L. Cui and D. Lee, "CoAID: COVID-19 Healthcare Misinformation Dataset," pp. 1–10, 2020, [Online]. Available: http://arxiv.org/abs/2006.00885.

[27] S. Schulz, P. Daumke, M. Romacker, and P. López-García, "Representing oncology in datasets: Standard or custom biomedical terminology?," *Informatics Med. Unlocked*, vol. 15, 2019, doi: 10.1016/j.imu.2019.100186.

[28] S. Temma, M. Sugii, and H. Matsuno, “The Document Similarity Index based on the Jaccard Distance for Mail Filtering,” 2019, doi: 10.1109/ITC-CSCC.2019.8793419.

[29] Y. Liu, K. Yu, X. Wu, L. Qing, and Y. Peng, “Analysis and detection of health-related misinformation on Chinese social media,” IEEE Access, vol. 7, pp. 154480–154489, 2019, doi: 10.1109/ACCESS.2019.2946624.

[30] S. N. Qasem, M. Al-Sarem, and F. Saeed, “An ensemble learning based approach for detecting and tracking COVID19 rumors,” Comput. Mater. Contin., vol. 70, no. 1, pp. 1721–1747, 2021, doi: 10.32604/cmc.2022.018972.

Data Security Awareness in Online Learning

Rosilah Hassan[1]
Center for Cyber Security
Faculty of Information Science and Technology
Universiti Kebangsaan Malaysia
Bangi, Malaysia

Wahiza Wahi[2]
School of Liberal Studies (CITRA-UKM)
Universiti Kebangsaan Malaysia
Bangi, Selangor

Nurul Halimatul Asmak Ismail[3]
Department of Computer Science and Information Technology, Applied College
Princess Nourah bint Abdulrahman University
Riyadh, Kingdom of Saudi Arabia

Samer Adnan Bani Awwad[4]
Deanship of Information and Communication Technology
Imam Abdulrahman Bin Faisal University
Dammam, Kingdom of Saudi Arabia

***Abstract*—The Covid-19 pandemic has intensified the online learning activities, which have becoming the most crucial platforms for learning sessions. With these online learning activities, a major concern arises particularly on the security of educators and students' data. Every element in an online learning system can be a potential target of hacking or attack. Therefore, this paper examines students' awareness of data security in online learning. Google forms have been used to gather the students' feedback. Forty-two (42) students, involving the secondary school, pre-university and university students, responded to the survey questions. The results show that most of the respondents are well aware of how to keep their data safe in online learning. It is discovered that the level of awareness about data security in online learning amongst the students is commendable. The findings of this study indicate there are various ways to secure students' data especially during online learning.**

***Keywords*—Covid-19; IR4.0; education 4.0; online learning; data security**

I. Introduction

The massive spread of Covid-19 virus has disrupted many human activities all around the globe. One common daily activity, which has been significantly affected by Covid-19, is education. The Covid-19 pandemic has caused UNICEF, WHO, and IFRC to appeal when the virus situation spreads rapidly, schools must be closed in the 'Prevention and Control of Covid-19' spreading in schools [1]. On 8th April 2020, it was reported that 220 million post-secondary students across 175 countries [2] had experienced severe disruptions to their education given that their colleges and universities were forced to shut down in order to stop the spread of the virus [3]. This represents 13 percent of the total number of students affected globally [4]. Hence, most educational institutions have opted for online learning platforms to ensure that learning sessions continue to operate within the pandemic [5, 6].

Online learning is "a type of delivery method used in distance education that allows synchronous and asynchronous exchanges of resource over a communication network" [7] by using various devices with internet access such as mobile phone, laptops, and computer [8, 9]. Online learning is seen as a tangible form of technological development that is not restricted to the current 4.0 Industrial Revolution (4IR) only [10]. Haseeb (2018) denotes that online learning is one of the important aspects for future education or Education 4.0 to produce highly creative graduates [11]. The Education 4.0 is a global framework introduced by the World Economic Forum (2020) to ensure a high-quality education within the context of the 4IR and the future of the new global economy and society [12]. For example, the 4IR has observed that learners play a critical part in their learning activities, as well as a flexibility in the method and ambience of learning provided to them in many online classes given by higher learning institutions [13].

It is apparent that the current and future of education relies greatly on the use of advanced technology and automation whereby online learning plays a central role [14]. Fig. 1 illustrates the major trends of Education 4.0.

In fact, the Covid-19 pandemic has brought about tremendous challenges on students having to go through online learning sessions [15]. According to Chung et al. [16], the pandemic's quick shifts have had a significant impact on students and lecturers at higher education institutions critically. Many lecturers and students in schools all around the world have expressed their frustrations with adopting online learning platforms for teaching and learning. The ramifications of this epidemic are unexpected, according to Shahzad et al. [17], and it has transformed the education system into a new teaching and learning paradigm.

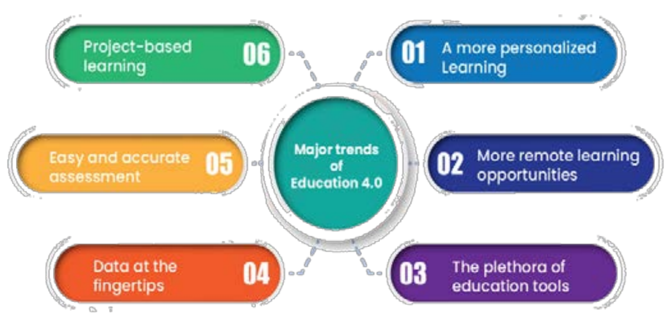


Fig. 1. Major Trends of Education 4.0.

This paper is granted under the Dana Impak Perdana UKM DIP-2018-040.

Indeed, various aspects of the new education system such as the curriculum, the role of educators, student positions as well as assessments have been modified to cope with current trends [18, 19]. For instance, educators need to fill in students' data and update them constantly. Students, on the other hand, need to access their online learning platform persistently to obtain information and instructions related to their assignments and exercises. When everything is done online, the data security of educators and students has becoming a crucial concern. This is especially true when the users commit malicious acts and access the data illegally. These malicious acts include instilling viruses and malicious code such as worms and Trojan horses into the device with the intent to steal users' information. Every component of an online learning system is vulnerable to hackers or attacks.

Data security is to protect valuable and sensitive personal data such as email and bank information. Valuable data comprise of educational information that collected, stored and managed by universities as well as confidential information such as financial data, phone numbers, and sensitive personal information about students' personal life and family data are the goldmine for hackers. In data security, the processes and technologies should be used to safeguard the data is a crucial element in protecting personal data at best. The created, collected, stored, and exchanged data is a valuable asset. Safeguarding data from corruption and unauthorized access by internal or external hackers protects individual's data from being stolen and used without his or her consent and knowledge.

Researchers have devised a number of remedies and methods to increase security in online learning in response to rising dangers, particularly during current epidemic. This paper reports on a study that investigated students' understanding on data security as they are taking online learning during the Covid-19 pandemic. The reminder of this paper is organized as follows; Section II presents related works on online learning and data security, Section III describes methodology, Section IV depicts results and discussion, and Section V concludes the paper.

II. Related Work

A. Online Learning

Online learning refers to the teaching and learning method carried out in a web-based environment [20]. With access to the internet, the learning process can be conducted anywhere and anytime with the use of gadgets or online platforms which are accessible by any users. There are various teaching and learning applications available on the internet such as Google Classroom, Google Meet, Microsoft Teams, Skype, WhatsApp, and Zoom. Most educators and students depending on their own teaching and learning needs utilize these applications.

The occurrence of the Covid-19 pandemic has instigated the implementation of online learning on a vast scale worldwide. Interactions between educators and their students have now changed from a face-to-face to screen mode. This unexpected transition has led to many challenges encountered by both the educators and students. Some students are facing internet connectivity problems and lack of digital devices for their virtual learning [21]. Likewise, most lecturers are experiencing various challenges including financial, physical, and mental issues [22]. Most of them are grappling with the demands of online learning lessons and the advanced technologies in their attempt to ensure that learning takes place without disruption [20].

B. Security in Online Learning

Security in online learning relates to the safeguarding of resources against deliberate or unintentional usage [23]. According to previous research, there are three main needs for security: confidentiality, integrity, and availability [24], as shown in Fig. 2.

Confidentiality refers to the protection of sensitive information against unauthorized access as well as the absence of illegal dissemination [25]. Many online learning environments have a huge number of users (among them students, visitors, instructors, tutors, and administrators). To protect access to the proper user, both a login mechanism and a robust delimitation designating registered users and user groups are necessary [24]. Security precautions such as authentication and encryption are widely used to secure personal information.

Integrity, a critical element of security, refers to "the protection of data from intentional or accidental unauthorized changes" [24]. It also denotes "the absence of improper system alterations" [25]. It assures that "information and data have not been accidentally or maliciously modified or destroyed, and are in accurate, correct, and complete original form" [26]. Access control is the key to maintaining integrity in the online learning environment [24].

Availability means the readiness for correct service [24]. It connotes that any authorized users whenever needed can access an online learning system [24]. Additionally, it assures that "information and communication resources are readily accessible and reliable in a timely manner by authorized persons" [26]. Availability can be destroyed mainly by denial of service and/or loss of data processing capabilities [24].

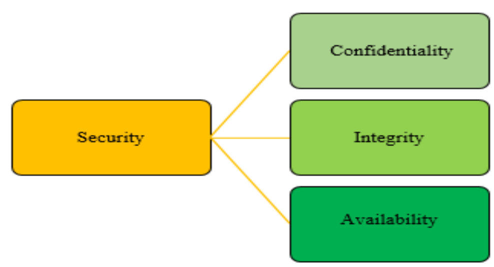


Fig. 2. Basic Requirements of Security.

III. Methodology

Google forms questionnaire was used to acquire feedback from students in this study. The questions were designed to measure the level of awareness among students on their understanding regarding data security in online learning. A total of 42 students responded the survey. The respondents include secondary school students, pre-university students, and university students. The methodology for conducting this research is shown in Fig. 3.

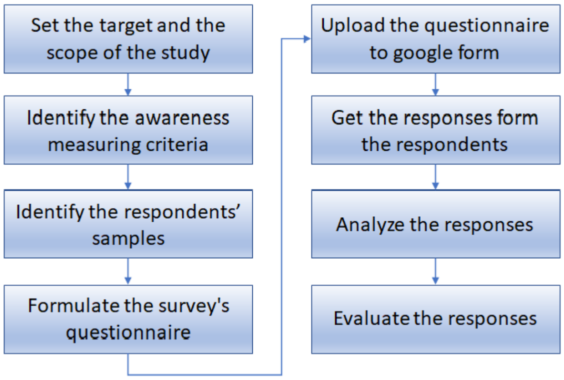


Fig. 3. Methodology.

This paper targets to measure the students' awareness level and practice of data security while they use online learning. The students' awareness level has been measured based on the following indicators:

- General awareness about online security.
- Using anti-virus on the device that they use for online learning.
- Using two factor authentications when logging in to the online learning application.
- Practicing logging out of online learning applications after using them.
- Using their own personal device individually or sharing the device with other students.
- Awareness about the confidentiality of the data that the student share in online learning.
- Using common information such as birthday, identity card (IC) number, matric number as the passwords for their online learning software.
- Using online learning platform on public computers or Wi-Fi such as in the library.
- Sharing the information in trusted website only.

IV. Result and Discussion

Forms Questionnaire. Fig. 4 shows the gender of the respondents. As the figure depicts, 71.4% of the respondents are male while the remaining 28.6% are female.

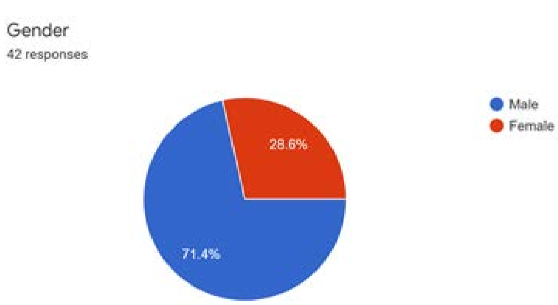


Fig. 4. Gender of the Respondents.

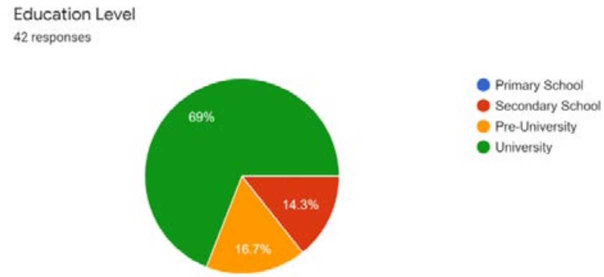


Fig. 5. Educational Level of the Respondents.

Fig. 5 exhibits the education level of the respondents whereby 69% of the respondents are university students, 16.7% are pre-university students, and 14.3% are secondary school students.

Fig. 6 displays the respondents' awareness about online security. The highest percentage of students, 50% of them, is well aware about online security. On the other hand, the lowest percentage, 2.5% of them, represented respondents who are unaware about online security. Meanwhile, the chart shows that almost 4.8% of the respondents have little awareness, 16.7% have awareness, and 26.2% have excellent awareness about online security.

As can be depicted from the figure, the majority of respondents are aware about online security, which indicates a positive result. However, it will be a lot better if all the respondents are very well aware about online security since this will prevent them from any malicious actions such as data theft.

Fig. 7 shows whether or not the respondents have practiced online security in online learning. The highest percentage of the respondents, almost 42.9% of them, has practiced online security well. On the other hand, the lowest percentage of them, almost 2.4%, are unaware about practicing online security. Meanwhile, the chart shows that also 4.8% have practiced a little, 19% have practiced, and 31% have very well practiced the online security. An overall result indicates that most of the respondents have at least used an anti-virus in their electronic devices in order to protect their devices from virus such as Trojan [27]. However, it is speculated that half of the respondent have never practiced online security. For example, they do not used two factor authentications to secure third party platforms.

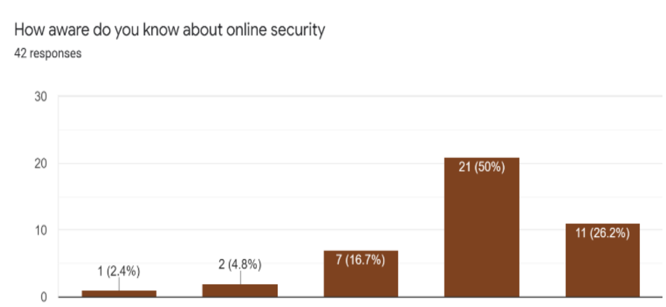


Fig. 6. Awareness about Online Security.

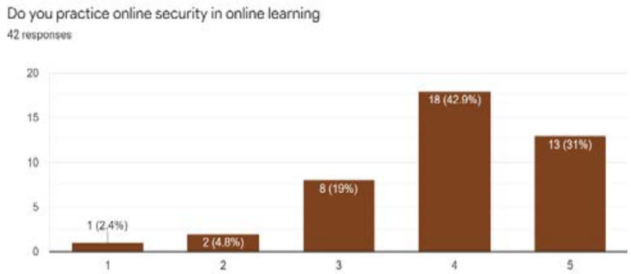


Fig. 7. Respondents Practice Online Security in Online Learning.

Fig. 8 shows that 83.3% of the respondents have used antivirus software during online learning. On the other hand, 16.6% respondents have not used any anti-virus software and that they are aware of the importance of it particularly in online learning. The pie chart clearly depicts that majority of the respondents have used anti-virus. This is due to the fact that anti-virus is a common software that can be easily installed in most devices.

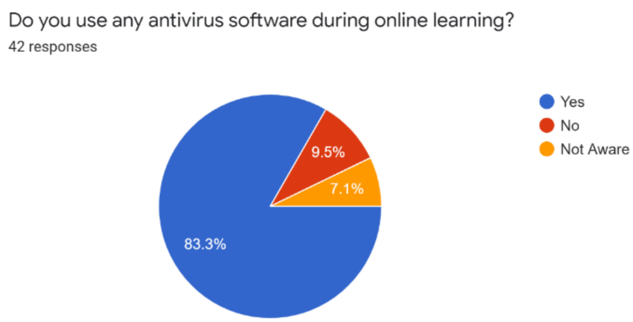


Fig. 8. Use Antivirus Software during Online Learning.

Fig. 9 shows whether the respondents log out of the online learning applications software after using them. It is obvious that half of the respondents log out of the applications software after using them while 40.5% of them do not. 9.5% of the respondents are not sure whether they log out after using the applications software or not.

It is important to note that 50% of the respondents, who log out from the applications software after using them, are those who are very much aware of online security and thus, logging out after using the applications has become a common practice to them. These respondents understand the risks of not logging out of the applications software. For example, they are aware of the risks when someone else uses their devices and gets an access to their personal information like e-mails. It can be concluded that another 50% of the respondents, who do not log out of the applications software, are less mindful and unconcerned of the effects of their actions.

Fig. 10 illustrates the respondents' answers on whether or not they have used two factor authentications during their online learning. The results depict that 52.4% of the respondents have used two factor authentications for their online learning platform. On the other hand, 35.7% of them have not used two factor authentications for online platform while the remaining 11.9% are not aware about using two factor authentications for online learning platform.

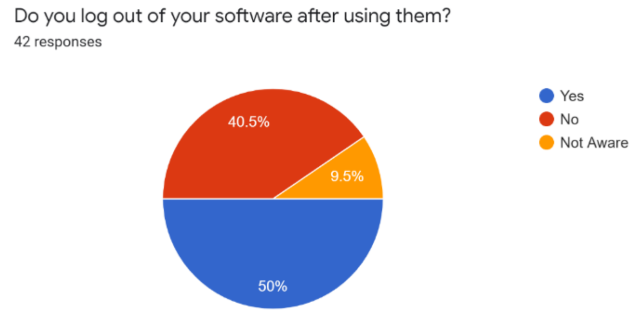


Fig. 9. Log out of Software after using them.

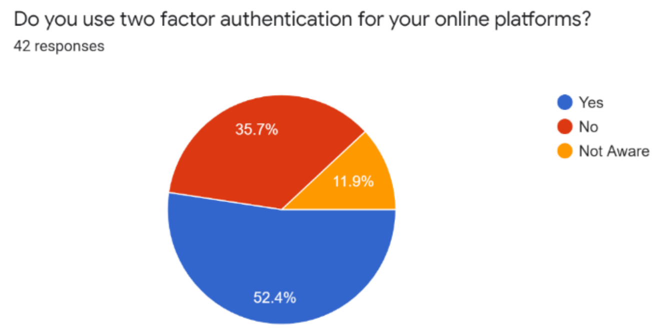


Fig. 10. Use Two Factor Authentications for Online Platforms.

This indicates that some respondents are totally not aware of the function of two factor authentications. This is rather dangerous as other users using their e-mail's accounts or passwords can easily access their accounts.

The respondents were asked whether they have individually used their own devices or they have shared their own devices with others during online learning. Fig. 11 presents the results in which approximately 69% of the respondents used their own personal devices, whilst, 23.8% indicated that they shared their devices with others during online learning. The remaining 7.1% of the respondents are not sure whether they have used their own device or they have shared devices with others during online learning.

The results indicate that it is better for the respondents to use their own personal devices instead of sharing the devices with others as this can avoid from sharing their data with others. By having their own devices, the respondents can also manage their files more strategically and efficiently. If the respondents share their devices with others, there is a possibility that they may unintentionally expose their confidential information.

Fig. 12 shows whether the respondents are aware about the data that they share in online learning including the privacy of the data like personal information, pictures and location. Majority of the respondents, 66.7%, are aware about the data their share in online learning. Conversely, 33.3% of the respondents are not aware about the of the privacy data they share in online learning.

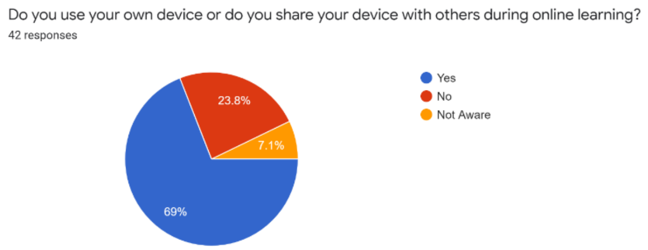


Fig. 11. Use Own Device or Share Device with others during Online Learning.

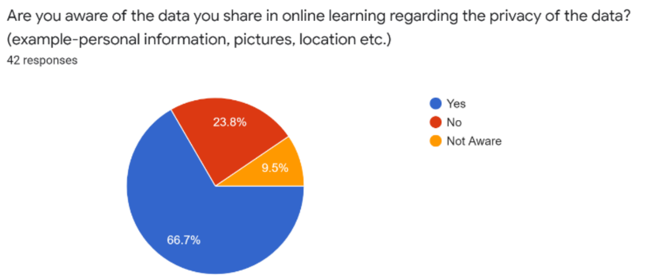


Fig. 12. Aware of the Data you Share in Online Learning.

The respondents should be aware of the data that they share in online learning especially when it involves confidentiality of the data. This is because hacker could either use the respondents' information to blackmail them or sell their data to other parties. It is recommended that 33.3% of the respondents to take classes about data security or use self-reading to gain more information about the importance of protecting data especially during this pandemic.

Fig. 13 shows whether the respondents use common information such as birthday, identity card (IC) number, matric number as the passwords for their online learning software. 59.5% of the respondents have been using common information for their online learning software's password. This is probably because it is easier for the respondents to remember their password whenever they want to log in into their accounts. 26.2% of the respondents have not used common information for their online learning software's password.

This indicates that the respondents are alert that most hackers these days can easily guess users' password particularly their personal data such as birthdate or real name as their passwords. Next, 14.3% of the respondents are not aware or unsure if they use common information for their online learning software's password.

Fig. 14 shows whether the respondents have ever used online learning platform on public computers or Wi-Fi such as in the library. The figure shows that 54.8% of the respondents have not used online learning platform on public computers or Wi-Fi such as in the library. This is generally because of the pandemic that forces the students to do their online learning at home [28]. 38.1% of the respondents have used online learning platform on public computers or Wi-Fi such as in the library or even McDonald's.

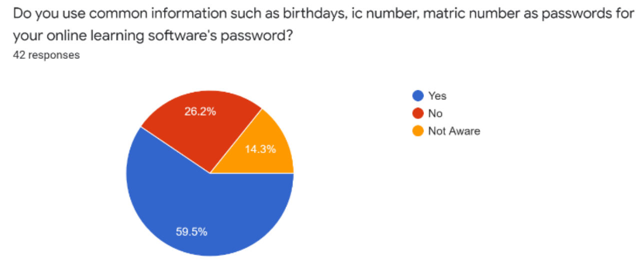


Fig. 13. Use Common Information as Passwords for Online Learning Software Password.

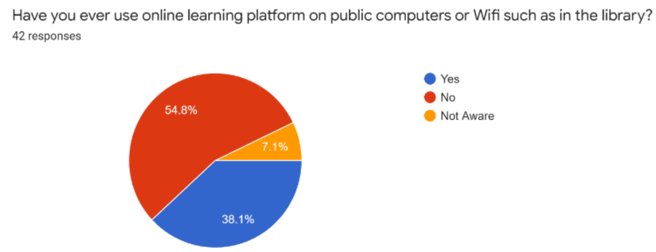


Fig. 14. Use Online Learning Platform on Public Computer or Wi-Fi.

This indicates that some of the respondents have been using the university Wi-Fi or even library during this pandemic. The remaining 7.1% of the respondents are not aware whether they have ever used online learning platform on public computers or Wi-Fi in public places such as in the library.

Fig. 15 presents the respondents' answers on whether or not the website in which they share their information is trusted and legitimate. The respondents' awareness helps to avoid sharing important information with irresponsible people.

The results show that 66.7% of the respondents are aware that the websites which they share information is secure and legit, while 33.3% of the respondents are not aware about it. This result is disturbing given that generally 1 out of 3 respondents is not aware about the danger of the websites that can be potentially a scam to steal important information, money, and private information.

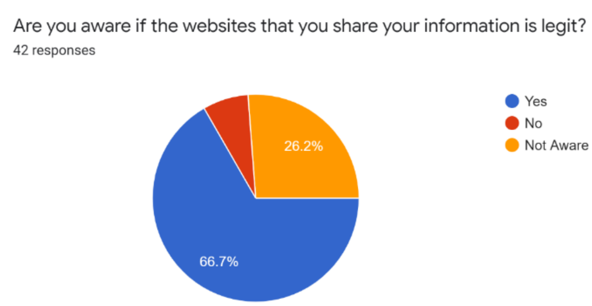


Fig. 15. Awareness if the Websites that you Share your Information is Legit.

V. Conclusion

The demands of the use of technology, especially during the Covid-19 pandemic, have led to the growth of end user devices and the widespread of online learning around the globe. Indeed, these massive applications of online learning bring along with them various security risks to many users. Nevertheless, it is safe to say that online learning is still at its premature and impulsive phase because most educators and students alike are still struggling in getting themselves familiar with its applications and operations. Concurrently, its providers are still taking the security risks of online learning lightly although it has become the most vital means of learning in the world today. It is important to note that most providers had to opt for an impulsive execution of the information and communication technology without taking into consideration relevant security concerns such as two-factor authentications. This was deliberately done to ensure that learning continues regardless the challenges experienced by students and educators during the pandemic. Given this scenario, it is discovered that most of them are not concerned about the importance of online security. This is evidenced in the study in which students at various levels of education were found to be insensitive towards the crucial needs to protect data security when conducting online learning. It is recommended that the stakeholders, such as the Ministry of Higher Education and the university, to enforce on data security in online learning. Online talks and forums on the most effective ways of online learning security should be conducted for educators and students to enhance their awareness. Therefore, data security awareness among students is essential.

Acknowledgment

We would like to acknowledge the Network and Communication Technology Research Group for the assistance and support they have provided. This survey has been conducted by UKM Students from TK2133-Data Communication and Telecommunication class.

References

[1] E. Budiman, U. Hairah, M. Wati, and H. Haviluddin, "Sensitivity analysis of data normalization techniques in social assistance program decision making for online learning," Advances in Science, Technology and Engineering Systems Journal, vol. 6, no. 1, pp. 49-56, 2021, doi: 10.25046/aj060106.

[2] N. Azman and D. Abdullah, "A critical analysis of Malaysian Higher Education Institutions ‘response towards Covid-19: Sustaining academic program delivery," Journal of Sustainability Science and Management, vol, 16, no. 1, pp. 70-96, 2021, doi:10.46754/jssm.2021.01.008.

[3] A. Obeidat, R. Obeidat, and M. Al-Shalabi, “The effectiveness of adopting e-Learning during Covid-19 at Hashemite University,” International Journal of Advanced Computer Science and Applications (IJACSA), vol. 11, no. 12, pp. 96- 104, 2020.

[4] UNESCO, 2020. COVID-19 educational disruption and response. Retrieved from https://en.unesco.org/themes/educationemergencies/ coronavirus-school-closures. Accessed on 28 December 2021.

[5] D. F. Murad, R. Hassan, Y. Heryadi, and B. D. Wijanarko, "The impact of the Covid-19 pandemic in Indonesia (Face to face versus Online Learning)," 2020 Third International Conference on Vocational Education and Electrical Engineering (ICVEE), pp. 1-4. 2020, doi:10.1109/ICVEE50212.2020.9243202.

[6] Q. Kharma, K. Nairoukh, A. Hussein, M. Abualhaj, and Q. Shambour, “Online learning acceptance model during Covid-19: an integrated conceptual model” International Journal of Advanced Computer Science and Applications (IJACSA), vol. 12, no. 5, 2021, http://dx.doi.org/10.14569/IJACSA.2021.0120561.

[7] B. H. Khan, "Web-based instruction (WBI): An introduction," Educational Media International, vol. 35, no. 2, pp. 63-71, 1998, doi: 10.1080/0952398980350202.

[8] S. Dhawan, "Online learning: A panacea in the time of Covid-19 crisis," Journal of Educational Technology Systems, vol. 49, no. 1, pp. 5-22, 2020, doi: 10.1177/0047239520934018.

[9] V. Singh and A. Thurman, "How many ways can we define online learning? A systematic literature review of definitions of online learning (1988-2018)," American Journal of Distance Education, vol. 33, no. 4, pp. 289-306, 2019, doi: 10.1080/08923647.2019.1663082.

[10] B. M. Batubara, "The problems of the world of education in the middle of the Covid-19 pandemic," Budapest International Research and Critics Institute (BIRCI-Journal): Humanities and Social Sciences, vol. 4, no. 1, pp. 450-457, 2021, doi:10.33258/birci.v4i1.1626.

[11] A. Haseeb, 2018. Higher Education in the Era of IR 4.0. New Straits Times. Retrieved from https://www.nst.com.my/education/2018/01/ 323591/higher-education-era-ir-40. Accessed on 28 December 2021.

[12] World Economic Forum. 2020. Schools of the Future, Defining New Models of Education for the Fourth Industrial Revolution. Geneva: World Economic Forum. Retrieved from https://www3.weforum.org/ docs/WEF_Schools_of_the_Future_Report_2019.pdf. Accessed on 4 January 2022.

[13] D. F. Murad, R. Hassan, W. Wahi, and B. D. A. Wijanarko, "User-item collaborative filtering system to predict online learning outcome," Advances in Science, Technology and Engineering Systems Journal, vol. 5, no. 5, pp. 117-121, 2020, doi: 10.25046/aj050516.

[14] Creatrix Campus, 2020. Why should higher education institutions focus on Education 4.0? Retrieved from https://www.creatrixcampus.com/ blog/Education-4.0. Accessed on 28 December 2021.

[15] E. J. Thandevaraj, N. A. A. Gani, and M. K. M. Nasir, "A review of psychological impact on students online learning during Covid-19 in Malaysia," Creative Education, vol. 12, no. 6, pp. 1296-1306, 2021, doi:10.4236/ce.2021.126097.

[16] E. Chung, N. M. Noor, and V. N. Mathew, "Are you ready? An assessment of online learning readiness among university students," International Journal of Academic Research in Progressive Education and Development, vol. 9, pp. 301-317, 2020, doi:10.6007/IJARPED/v9-i1/7128.

[17] A. Shahzad, R. Hassan, R., A. Y. Aremu, A. Hussain, and R. N. Lodhi, "Effects of Covid-19 in e-learning on higher education institution students: the group comparison between male and female," Quality & Quantity, vol. 55, pp. 805-826, 2020, doi:10.1007/s11135-020-01028-z.

[18] F. M. Kamaruzaman, N. A. Sulaiman, and N. A. N. Shaid, "A study on perception of student’s readiness towards online learning during Covid-19 pandemic," International Journal of Academic Research in Business and Social Sciences, vol. 11, no. 7, pp. 1536-1548, 2021, doi:10.6007/IJARBSS/v11-i7/10488.

[19] S. J. Daniel, "Education and the Covid 19 pandemic," Prospects, vol. 49, pp. 91-96, 2020, doi:10.1007/s11125- 020-09464-3.

[20] M. N .O. Sadiku, P. O. Adebo, and S. M. Musa, "Online teaching and learning," International Journals of Advanced Research in Computer Science and Software Engineering, vol. 8, no. 2, pp. 73 – 75, 2018.

[21] M. Mahyoob, "Challenges of e-learning during the Covid-19 pandemic experienced by efl learners," Arab World English Journal (AWEJ), vol. 11, no. 4, pp. 351 – 362, 2020, doi:10.24093/awej/vol11no4.23.

[22] C. T. Vu et al., "Dataset of Vietnamese teachers’ perspectives and perceived support during the COVID-19 pandemic," Elsevier, Data Brief 2020, vol. 31, p. 105788, 2020, doi:10.1016/j.dib.2020.105788.

[23] A. Adams, A. Blandford, "Security and online learning: To protect or prohibit," Usability Evaluation of Online Learning Programs, pp. 331-359, 2003, doi:10.4018/978-1-59140-105-6.ch018.

[24] A. Serb, C. Defta, N. M. Iacob, and M. C. Apetrei, "Information security management in e-learning," Knowledge Horizons, vol. 5, no. 2, pp. 55-59, 2013.

[25] E. Weippl and M. Ebner, "Security privacy challenges in e-learning 2.0," In World Conference on E-Learning in Corporate, Government,

Healthcare, and Higher Education, vol. 2008, no. 1, pp. 4001-4007, 2008.

[26] R. Raitman, L. Ngo, N. Augar, and W. Zhou, "Security in the online e-learning environment," In 5th Advanced Learning Technologies, 2005 (ICALT), pp. 702-706, 2005, doi:10.1109/ICALT.2005.236.

[27] K. G. Liakos, G. K. Georgakilas, S. Moustakidis, P. Karlsson, and F. C. Plessas, "Machine learning for hardware trojan detection: a review," 2019 Panhellenic Conference on Electronics & Telecommunications (PACET), 2019, pp. 1-6.

[28] N. A. N. Shaid, F. M. Kamruzaman, and N. A. Sulaiman, "Online Learning During Ongoing Covid-19 Pandemic: A Survey of Students' Satisfaction," International Journal of Academic Research in Business and Social Sciences, vol. 11, no. 7, pp. 924-937, 2021.

KP-USE: An Unsupervised Approach for Key-Phrases Extraction from Documents

Lahbib Ajallouda[1], Fatima Zahra Fagroud[2], Ahmed Zellou[3], Elhabib Ben Lahmar[4]
SPM Research Team, ENSIAS, ENSIAS, Mohammed V University, Rabat, Morocco[1, 3]
LTIM–FSBM, FSBM Hassan II University, Casablanca, Morocco[2, 4]

***Abstract*—Automatic key-phrase extraction (AKE) is one of the most popular research topics in the field of natural language processing (NLP). Several techniques were used to extract the key-phrases: statistical, graph-based, classification algorithms, deep learning, and embedding techniques. AKE approaches that use embedding techniques are based on calculating the semantic similarity between a vector representing the document and the vectors representing the candidate phrases. However, most of these methods only give acceptable results in short texts such as abstracts paper, but on the other hand, their performance remains weak in long documents because it is represented by a single vector. Generally, the key phrases of a document are often expressed in certain parts of the document as, the title, the summary, and to a lesser extent in the introduction and the conclusion, and not of the entire document. For this reason, we propose in this paper KP-USE. A method extracts key-phrases from long documents based on the semantic similarity of candidate phrases to parts of the document containing key-phrases. KP-USE makes use of the Universal Sentence Encoder (USE) as an embedding method for text representation. We evaluated the performance of the proposed method on three datasets containing long papers, namely, NUS, Krapivin2009, and SemEval2010, where the results showed its performance outperforms recent AKE methods which are based on embedding techniques.**

Keywords—Key-phrase extraction; natural language processing; semantic similarity; embedding technique; universal sentence encoder

I. INTRODUCTION

The explosive growth in the number of digital documents has prompted researchers to find effective ways to analyze and summarize all these documents [1]. AKE is one of the best document content analysis solutions. Researchers have used several techniques to improve the performance of key-phrase extraction [2]. Sentence embedding [3] is a recent technique that has been used to select key-phrases using a similarity measure [4]. Most of these methods perform less well, especially in long documents, because the document contains a large amount of information, which negatively affects its representation by a single vector. In general, most documents contain the title, abstract, introduction, and conclusion. These are the parts likely to contain key-phrases. The calculation of the similarity between the candidate phrases and the document must therefore take this factor into account.

The objective of this article is to propose KP-USE, a new, unsupervised method for key-phrases extraction from documents. KP-USE divides the document into five main parts: title, abstract, introduction, body, and conclusion. It represents each part by a USE technique [5] as a sentence embedding technique, in which the semantic similarity between candidate key-phrases and document is based on the phrases proximity to these parts, giving preference to parts that often contain key-phrases.

The rest of the article is organized as follows. In Section 2, we discuss related works. The USE sentence embedding technique is presented in Section 3, and then we present the proposed method for key-phrase extraction in Section 4. We empirically evaluate KP-USE in Section 5. Finally, we conclude the article in Section 6.

II. RELATED WORK

In this section, we will talk about the most important related work to automatic key-phrases extraction, and in addition, we will introduce sentence embedding techniques.

A. Key-phrases Extraction Approaches

Over the past twenty years, many automatic methods of key-phrase extraction have been proposed. Siddiqi et al. in [6] classify these approaches into three sets, supervised, Semi-Supervised and unsupervised methods. Generally, the extraction by supervised approaches can be considered as a binary classification problem, where the candidate phrases are classified either as a key-phrase or a non-key phrase. While unsupervised methods rely on ordering candidate phrases based on calculating the score from one or more weights.

While Papagiannopoulou et al. in [2] suggested classifying these methods according to the approved techniques, methods based on statistics such as [7], [8], [9]. Graph-based methods such as [10], [11], [12]. Methods that use binary classification such as [13], [14]. Methods based on deep learning as [15],[16], [17] and methods based on embedding techniques such as [18], [19], [20].

B. Sentence Embedding Technics

Sentence embedding is an efficient way to convert textual data into fixed-length multidimensional vectors. Sentence embedding methods can be classified into two categories: (i) non-parameterized methods such as SIF [21], uSIF [22], and GEM [23] which rely on a simple technique, by encoding the words that make up the sentence and averaging the resulting vectors as a vector representing the sentence. However, this technique neglects information about word order and sentence semantics. (ii) parameterized more complex methods and generally perform better than unparameterized models [24].

The popularity of encoders has led to a great evolution of these methods. Conneau et al. propose in [25] InferSent is a technique based on a supervised RNN model predicting the semantic relations between pairs of sentences. Universal Sentence Encoder [5] is a trained and optimized technique for short sentences and paragraphs, it is based on two encoder models, Transformer and Deep Average Network (DAN). Subramanian et al. propose in [26] a more complex model for learning sentence embeddings in a multitasking configuration. Reimers and Gurevych propose in [27] Sentence-BERT, which also uses a Siamese network to create BERT-based sentence embeds [28]. Generally, Parameterized methods outperform non-parameterized ones on many tasks, but they are computationally expensive and require more training time.

III. Universal Sentence Encoder

Universal Sentence Encoder (USE) is a sentence embedding model that encodes a sentence or paragraph into a 512-dimensional vector. This vector encodes the meaning of the sentence and can therefore be used as input for NLP tasks such as document classification, key-phrase extraction, and textual similarity analysis.

A. USE Process

The idea of USE is to encode sentences into 512-dimensional vectors, via an encoder. These vectors are used in many NLP tasks. Depending on the errors made in these tasks, USE again reproduces vectors for these sentences. Fig. 1 shows the USE process.

The USE process begins with a tokenization operation, where sentences are converted to lowercase and tokenized into tokens. In the second step, USE encodes the sentence.

B. Encoder

USE offers two architectures for encoding the sentence. The first is based on a transformer consisting of six layers. Each layer has a feedback network preceded by a self-attention module that takes into account word order and context when generating each word representation. Fig. 2 shows USE architecture with transformer encoder.

The second is based on Deep Average Network (DAN) proposed by Iyyer et al in [29], where the vectors of the words in the sentence are averaged. The resulting vector is then passed through a 4-layer deep neural network (DNN) to obtain a vector of 512 dimensions. Fig. 3 shows USE architecture with DAN.

Generally, the result obtained through the transformer is very precise but requires more calculation time. It is therefore difficult to use for long texts, whereas DAN generates text encoding in less time but with less accuracy than a transformer.

C. Multi-task Learning

After the encoding of the sentences, either by transformers or by DAN. USE relies on multi-task learning (MTL) [30] to exploit commonalities and differences across tasks. This improves its learning efficiency and the accuracy of its predictions in the vector representation of sentences. Also, the authors of USE exploited a set of sources as training data. In addition to Stanford Natural Language Inference (SNLI) corpus [31], some web resources such as Wikipedia, web news, question and answer web pages, and discussion forums were also used.

Once USE has been trained, it can be used to represent any text, whether it is a phrase, a sentence, or a paragraph, with a 512-dimensional vector.

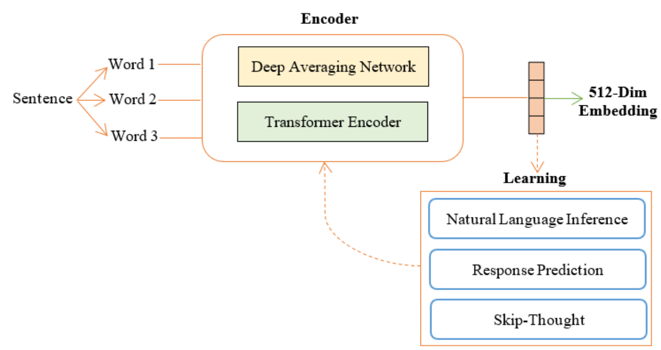


Fig. 1. Universal Sentence Encoder Process.

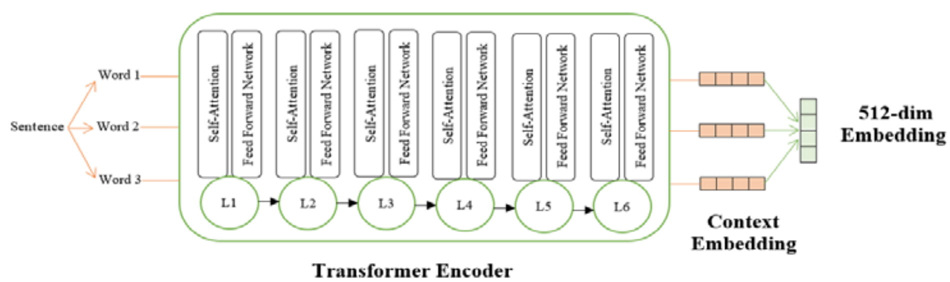


Fig. 2. USE Architecture with Transformer Encoder.

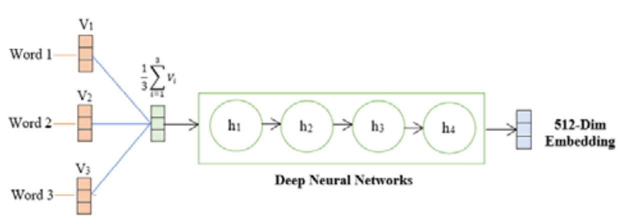


Fig. 3. USE Architecture with Deep Average Network.

D. Evaluation of USE

To show that embedding USE provides a better text representation. The authors used semantic similarity in the different tasks of SentEval [32]. They also preferred using angular distance (Formula 2) as a measure of similarity rather than cosine (Formula 1) because, in their opinion, it gave better results.

$$sim(p,q) = \frac{\sum p_i q_i}{\sqrt{\sum p_i^2}\sqrt{\sum q_i^2}} \quad (1)$$

$$ang\_sim(p,q) = 1 - \frac{1}{\pi} arccos\big(sim(p,q)\big) \quad (2)$$

- DataSet: Binary classification and sentence similarity tasks in SentEval are used to evaluate the quality of sentence representations. Table I shows the datasets used to evaluate the performance of USE. These datasets are used by USE authors to evaluate the two USE models, Transformer and DAN.

- Results: The experimental results showed the performance of USE in many tasks, whether in the transformer model or the DAN model. In Table II, we show the results of the two models for different datasets.

We also observe that encoding by transformer generally works better than encoding by DAN. On the other hand, the USE authors point out that the complexity of the transformer-based model is $O(n^2)$, while the DAN model is $O(n)$. Therefore, the transformer model is slower than the DAN model; especially with increasing sentence length.

TABLE I. DATASETS USED TO EVALUATE THE PERFORMANCE OF USE

Dataset	Number of samples	Task
Text REtrieval Conference (TREC) [33]	6 000	Question and answering
Customer Reviews (CR) [34]	4 000	Product reviews
Subjectivity Summarization (SUBJ) [35]	10 000	Subjectivity/Objectivity
Multi-Perspective Question and Answering (MPQA) [36]	11 000	Opinion polarity
Movie Reviews (MR) [37]	11 000	Sentiment
Stanford Sentiment Analysis (SST) [38]	67 300	Sentiment

TABLE II. USE PERFORMANCE ACCORDING TO EACH ENCODING MODEL

Dataset	USE DAN	USE Transformer
TREC	91.19	92.51
CR	80.97	87.43
SUBJ	92.65	93.87
MPQA	85.38	86.98
MR	74.45	81.44
SST	77.62	85.38

IV. KP-USE APPROACH

In this section, we propose KP-USE, which is a new method for key-phrases extraction from documents.

A. KP-USE Process

The KP-USE process presented in Fig. 4 shows that our method consists of five main steps:

Step 1: Extraction of candidate key-phrases;

Step 2: Split the document into five main sections;

Step 3: Vector representation of main sections and candidate key-phrases by USE;

Step 4: Calculate the score of each candidate key-phrase;

Step 5: Sorting the candidate key-phrases according to their score and extracting the phrases with the best score as key-phrases.

B. Candidate Key-phrases

The process of candidate key-phrases extraction (Fig. 5) begins with removing non-text data from the document, as well as converting all text to lowercase, and translating foreign words from the text into the language being studied. Then, the tokenization technique is used to convert the cleaned text into an array of tokens, to determine the grammatical category of each word using the Part-Of-Speech Tagging (POST) technique.

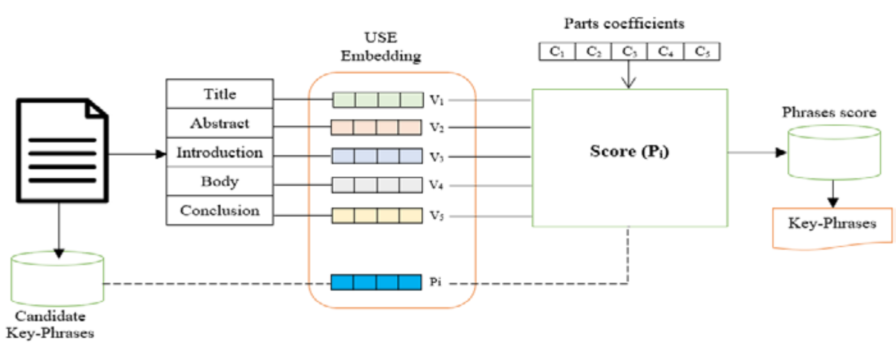


Fig. 4. The KP-USE Process.

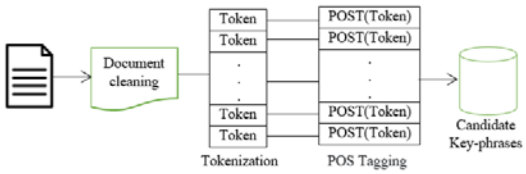


Fig. 5. Candidate Key-phrases Process.

Several AKE approaches have noted that key-phrases are noun phrases composed of one or more words such as [1], [39], and [40]. According to [19], the gerund form of a verb (VBG) can be used as a noun and the past participle form of a verb (VBN) can be considered as an adjective in the composition of key sentences. Therefore, (NN.*|JJ.*|VBN|VBG)*(NN.*|VBG) is the proposed pattern for candidate key-phrases extraction.

C. Document Fraction

Through our analysis of more than 50 scientific papers, we noticed that most of the key-phrases are semantically similar to the title of the article and are often mentioned in the abstract and to a lesser extent in the introduction and the conclusion. Based on this result (Table III), we propose to split the document into five parts, namely title, abstract, introduction, and conclusion, while the fifth part comprises the rest of the document. To favor the parts similar to the key phrases. In Table III, we propose for each part a proximity coefficient which will be used when calculating the similarity between the candidate phrases and the document.

TABLE III. THE PROXIMITY COEFFICIENTS FOR EACH PART

Section	Average Similarity	Coefficient
Title	0.61	3
Abstract	0.68	3
Introduction	0.42	2
Body	0.14	1
Conclusion	0.35	2

Some documents may not have one of the five parts. In this case, KP-USE considers the proximity coefficient of the non-existent part to be zero. Therefore, there will be no effect on the calculation of the similarity between the candidate phrases and the document. Thus, we will be able to apply KP-USE to any document.

D. Sentence Embedding

Our method proposes to extract key-phrases from the document based on the calculation of semantic similarity between the candidate key-phrases and the five parts of the document that have been identified in Table III. Before calculating the similarity, we must first represent the phrases candidate and the five parts by vectors. For this, we exploit USE as an embedding technique to obtain vectors of 512 dimensions. We test USE for encoding models, Transformer and DAN, to see which works best.

E. Candidate Key-phrase Score

The score of any candidate key-phrase is expressed by its semantic similarity with the document. This score is calculated based on the similarity of the candidate phrase to each of the five parts of the document, which is calculated by Formula 1. To exploit the proximity coefficient of each part of the candidate phrase, KP-USE uses Formula 3 to calculate the score for each candidate key-phrase.

$$Score(P_i) = \frac{1}{\sum_{j=1}^{5} C_j} \sum_{k=1}^{5} sim(V_k, U_i) \times C_k \tag{3}$$

P_i: The candidate key-phrase i.

U_i: the vector that represents phrase i.

V_k: The vector that represents part k.

C_k: Coefficient of part k.

C_j: Coefficient of part j.

F. Key-phrases Extraction

After calculating the score for each candidate key-phrase, KP-USE ranks these statements in descending order based on the score. KP-USE allows the user to choose the number of key-phrases to extract. The highest-rated phrases as key-phrases in the document.

V. EXPERIMENTAL EVALUATION

In this section, we will present the datasets that were used to evaluate KP-USE. Additionally, we will verify which of the two models exploited in USE performs better in AKE by evaluating their performance using the precision, recall, and F1-score metrics. Then, we compare the performance of our method with other AKE methods based on embedding techniques.

A. Datasets

To evaluate the performance of our method, we used three datasets that are considered the most widely used in evaluating key-phrase extraction approaches, which are.

- NUS [41], is a set of scientific data, consisting of 211 conference papers where each paper contains between 4 and 12 pages. Keywords were identified by the volunteer students where each student was asked to read three articles and extract the keywords.
- Krapivin2009 [42]: It is considered one of the largest datasets in terms of number documents, containing 2,304 research articles in the field of computer science. The keywords of the articles were identified by the authors and checked by the reviewers.
- Semeval-2010 [43]: This is one of the most widely used data sets for evaluating keyword extraction approaches. It consists of 244 scientific articles belonging to the field of computer science, where the number of article pages varies from 6 to 8 pages. The keywords for each article are defined by the authors and professional editors. Table IV shows the statistics for the three datasets.

The most important feature of the datasets used is that they contain scientific papers consisting of 4 and 15 pages, i.e. each paper contains the five parts we have discussed in Table III will confirm the credibility of the results we will obtain.

TABLE IV. DATASETS USED TO EVALUATE KP-USE

Dataset	Type of doc	Docs	Language	Tokens per Doc	KP per Doc	Annotation
NUS	Full paper	211	EN	8398.30	11.00	Reader
Krapivin2009	Full paper	2304	EN	8040.74	6.34	Author
SemEval2010	Full paper	244	EN	7961.20	14.70	Author/Reader

B. *Evaluation Metrics*

Several evaluation metrics are used in key-phrase extraction [44]. Researchers generally prefer three metrics, namely precision, recall, and F1.score because of their validity and ease of use. These are the same metrics we will use to evaluate KP-USE.

- Precision: To evaluate the precision of the approach, we use this metric. Its value is expressed as the proportion of correctly extracted key-phrases compared to the total number of extracted key-phrases. To calculate the precision, we use formula 4.

$$Precision = \frac{Correct\ KeyPhrases}{Extracted\ KeyPhrases} \quad (4)$$

- Recall: To evaluate the completeness of key-phrase extraction, we use the recall metric, which expresses the proportion of correctly extracted key-phrases among the author's or reader's selected key-phrases. To calculate the yield, we use formula 5.

$$Recall = \frac{Correctly\ Keyphrases}{Assigned\ Keyphrases} \quad (5)$$

- F1.Score: There is often an inverse interaction between precision and recall. When precision is high, recall is low. For this, the F1-score is used, to combine precision and recall. To calculate F1-score, we use formula 6.

$$F1.Score = \frac{2\times Precision\times Recall}{Precision+Recall} \quad (6)$$

C. *Comparison of USE Models*

To choose the appropriate encoding model for KP-USE, we experimented with extracting key phrases using both encoding models, Transformer and DAN. Table V shows the performance of KP-USE according to each model according to the F1-Score metric on the three datasets NUS, Krapivin, and Semeval2010 for top 5 (F1@5), top 10 (F1@10), and top 15 (F1@15).

From the results obtained, we find that KP-USE based on the Transformer model performs better than KP-USE based on the DAN model in all the datasets used. Although the complexity of the Transform model is higher than that of the DAN model, we prefer to use the Transformer-based model because it achieves promising results.

D. *Performance Comparison*

We compare KP-USE with three methods which are also based on embedding techniques namely EmbedRank [45] is an unsupervised method that uses the Sent2vec [46] embedding technique for the representation of phrases and documents. SIFRank [18] is an unsupervised key-phrase extraction method based on the SIF [21] and ELMo [47] embedding model. The third approach is MDERank [20], this is an unsupported approach. It implicitly embeds position and frequency offset information by encoding the document using BERT embedding technique [48].

In general, the results obtained by KP-USE remain acceptable compared to the performance of the other methods at the level of long documents. Table 6 present these results.

TABLE V. KP-USE PERFORMANCE ACCORDING TO TRANSFORMER AND DAN MODEL

Model	NUS			Krapivin			Semeval		
	F1@5	***F1@10***	***F1@15***	***F1@5***	***F1@10***	***F1@15***	***F1@5***	***F1@10***	***F1@15***
KP-USE Transformer	0.07	0.08	0.10	0.10	0.14	0.14	0.15	0.20	0.19
KP-USE DAN	0.03	0.06	0.10	0.08	0.11	0.11	0.12	0.16	0.17

TABLE VI. KP-USE PERFORMANCE COMPARED TO EMBEDRANK, SIFRANK, AND MDERANK

Approach	NUS			Krapivin			Semeval		
	F1@5	**F1@10**	**F1@15**	**F1@5**	**F1@10**	**F1@15**	**F1@5**	**F1@10**	**F1@15**
EmedRank	0.02	0.03	0.04	0.09	0.11	0.10	0.10	0.14	0.15
SIFRank	0.04	0.06	0.06	0.02	0.03	0.03	0.11	0.17	0.19
MDERank	**0.15**	**0.18**	**0.18**	**0.12**	0.13	0.13	0.13	0.17	**0.20**
KP-USE	0.07	0.08	0.10	0.10	**0.14**	**0.14**	**0.15**	**0.20**	0.19

E. Discussion

KP-USE is an unsupervised AKE method that takes advantage of the USE embedding technique to represent text vectorially. KP-USE relied on splitting the text to focus the search for key-phrases in the parts that might contain them, and this split also helps to improve the embedding of the document because instead of just being represented by a single vector, each part is represented by a vector. What also characterizes KP-USE is that the similarity calculation takes into account the parts of the document which often contain key-phrases, unlike other methods which consider all parts of the document to be of the same degree of importance, which makes KP-USE more efficient than these methods, especially in long documents. KP-USE can also be applied to short documents, where the importance coefficient for any part that does not exist in the document is zero, so the similarity calculation between candidate phrases and the document is not affected. KP-USE performance could be improved even more if we could develop it to be able to predict key-phrases that are not mentioned in the document.

VI. Conclusion

In this paper, we proposed KP-USE, an unsupervised method of extracting key-phrases from the document based on the USE embedding technique which uses the Transformer network model. Our method splits the document into five parts, namely title, abstract, introduction, body, and conclusion, to favor the parts likely to contain key-phrases while calculating the semantic similarity between the candidate key-phrases and the document. KP-USE was evaluated on three datasets continent of long documents, namely NUS, Krapivin, and Semeval 2010. F1-Score results showed that the performance of KP-USE was superior to the performance of unsupervised methods based on embedding techniques and on the calculation of semantic similarity to extract key-phrases. In the future, we will develop KP-USE to predict key-phrases not mentioned in the document.

Reference

[1] N. Nikzad-Khasmakhi, et al. "Phraseformer: Multimodal key-phrase extraction using transformer and graph embedding." arXiv preprint arXiv:2106.04939 (2021).

[2] E. Papagiannopoulou, and G. Tsoumakas. "A review of keyphrase extraction." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 10.2 (2020): e1339.

[3] H. Tsukagoshi, R. Sasano, and K. Takeda. "Comparison and Combination of Sentence Embeddings Derived from Different Supervision Signals." arXiv preprint arXiv:2202.02990 (2022).

[4] D. Buscaldi, et al. "Calcul de similarité entre phrases: quelles mesures et quels descripteurs?." 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Atelier DÉfi Fouille de Textes. ATALA; AFCP, 2020.

[5] D. Cer, et al. "Universal sentence encoder." arXiv preprint arXiv:1803.11175 (2018).

[6] S. Siddiqi, and A. Sharan. "Keyword and keyphrase extraction techniques: a literature review." International Journal of Computer Applications 109.2 (2015).

[7] S. R. El-Beltagy, and A. Rafea. "KP-Miner: A keyphrase extraction system for English and Arabic documents." Information systems 34.1 (2009): 132-144.

[8] R. Campos, et al. "Yake! collection-independent automatic keyword extractor." European Conference on Information Retrieval. Springer, Cham, 2018.

[9] M. Won, B. Martins, and F. Raimundo. "Automatic extraction of relevant keyphrases for the study of issue competition." Proceedings of the 20th international conference on computational linguistics and intelligent text processing. 2019.

[10] R. Mihalcea and P. Tarau. "Textrank: Bringing order into text." Proceedings of the 2004 conference on empirical methods in natural language processing. 2004.

[11] F.Boudin. "Unsupervised keyphrase extraction with multipartite graphs." arXiv preprint arXiv:1803.08721 (2018).

[12] Y. Yu, and V. Ng. "Wikirank: Improving keyphrase extraction based on background knowledge." arXiv preprint arXiv:1803.09000 (2018).

[13] C. Caragea, F. Bulgarov, A. Godea and S. D. Gollapalli. "Citation-enhanced keyphrase extraction from research papers: A supervised approach." Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014.

[14] S. McIlraith, and K. Q. Weinberger. "Learning feature representations for keyphrase extraction." Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, Louisiana, USA. 2018.

[15] R. Meng et al. "Deep keyphrase generation." arXiv preprint arXiv:1704.06879 (2017).

[16] J. Chen, X. Zhang, Y. Wu, Z. Yan, and Z. Li. "Keyphrase generation with correlation constraints." arXiv preprint arXiv:1808.07185 (2018).

[17] R. Alzaidy, C. Caragea, and C. L. Giles. "Bi-LSTM-CRF sequence labeling for keyphrase extraction from scholarly documents." The world wide web conference. 2019.

[18] Y. Sun, H. Qiu, Y. Zheng, Z. Wang, and C. Zhang. "SIFRank: a new baseline for unsupervised keyphrase extraction based on pre-trained language model." IEEE Access 8 (2020): 10896-10906.

[19] J. R. Asl, and J. M. Banda. "GLEAKE: Global and Local Embedding Automatic Keyphrase Extraction." arXiv preprint arXiv:2005.09740 (2020).

[20] L. Zhang, et al. "MDERank: A Masked Document Embedding Rank Approach for Unsupervised Keyphrase Extraction." arXiv preprint arXiv:2110.06651 (2021).

[21] S. Arora, Y. Liang, and T. Ma. "A simple but tough-to-beat baseline for sentence embeddings." International conference on learning representations. 2017.

[22] K. Ethayarajh. "Unsupervised random walk sentence embeddings: A strong but simple baseline." Proceedings of The Third Workshop on Representation Learning for NLP. 2018.

[23] Z. Yang, C. Zhu, and W. Chen. "Parameter-free sentence embedding via orthogonal basis." arXiv preprint arXiv:1810.00438 (2018).

[24] B. Wang, and C. C. J. Kuo. "Sbert-wk: A sentence embedding method by dissecting bert-based word models." IEEE/ACM Transactions on Audio, Speech, and Language Processing 28 (2020): 2146-2157.

[25] A. Conneau, D. Kiela, H. Schwenk, L. Barrault, and A. Bordes. "Supervised learning of universal sentence representations from natural language inference data." arXiv preprint arXiv:1705.02364 (2017).

[26] S. Subramanian, A. Trischler, Y. Bengio, and C. J. Pal. "Learning general purpose distributed sentence representations via large scale multi-task learning." arXiv preprint arXiv:1804.00079 (2018).

[27] N. Reimers, and I. Gurevych. "Sentence-bert: Sentence embeddings using siamese bert-networks." arXiv preprint arXiv:1908.10084 (2019).

[28] A. Vaswani, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).

[29] M. Iyyer, V. Manjunatha, J. Boyd-Graber, and H. Daumé III. "Deep unordered composition rivals syntactic methods for text classification." Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (volume 1: Long papers). 2015.

[30] M. Crawshaw. "Multi-task learning with deep neural networks: A survey." arXiv preprint arXiv:2009.09796 (2020).

[31] S. R. Bowman, G. Angeli, C. Potts, and C. D. Manning. "A large annotated corpus for learning natural language inference." arXiv preprint arXiv:1508.05326 (2015).

[32] A. Conneau, and D. Kiela. "Senteval: An evaluation toolkit for universal sentence representations." arXiv preprint arXiv:1803.05449 (2018).

[33] X. Li, and D. Roth. "Learning question classifiers." COLING 2002: The 19th International Conference on Computational Linguistics. 2002.

[34] M. Hu, and B. Liu. "Mining and summarizing customer reviews." Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. 2004.

[35] B. Pang, and L. Lee. "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts." arXiv preprint cs/0409058 (2004).

[36] J. Wiebe, T. Wilson, and C. Cardie. "Annotating expressions of opinions and emotions in language." Language resources and evaluation 39.2 (2005): 165-210.

[37] B. Pang, and L. Lee. "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales." arXiv preprint cs/0506075 (2005).

[38] R. Socher, et al. "Recursive deep models for semantic compositionality over a sentiment treebank." Proceedings of the 2013 conference on empirical methods in natural language processing. 2013.

[39] G. Rabby, S. Azad, M. Mahmud, K. Z. Zamli, and M. M. Rahman. "Teket: a tree-based unsupervised keyphrase extraction technique." Cognitive Computation 12.4 (2020): 811-833.

[40] X. Shen, Y. Wang, R. Meng, and J. Shang. "Unsupervised Deep Keyphrase Generation." arXiv preprint arXiv:2104.08729 (2021).

[41] T. D. Nguyen, and M. Y. Kan. "Keyphrase extraction in scientific publications." International conference on Asian digital libraries. Springer, Berlin, Heidelberg, 2007.

[42] M. Krapivin, A. Autaeu, and M. Marchese. "Large dataset for keyphrases extraction." (2009).

[43] S. N. Kim, O. Medelyan, M. Y. Kan, T. Baldwin, and L. P. Pingar. "SemEval-2010 Task 5: Automatic Keyphrase Extraction from Scientific." In Proceedings of the 5th International Workshop on Semantic Evaluation, pages 21–26. 2010.

[44] F. Liu, X. Huang, W. Huang, and S. X. Duan. "Performance evaluation of keyword extraction methods and visualization for student online comments." Symmetry 12.11 (2020).

[45] K. Bennani-Smires, C. Musat, A. Hossmann, M. Baeriswyl, and M. Jaggi. "Simple unsupervised keyphrase extraction using sentence embeddings." arXiv preprint arXiv:1801.04470 (2018).

[46] M. Pagliardini, P. Gupta, and M. Jaggi. "Unsupervised learning of sentence embeddings using compositional n-gram features." arXiv preprint arXiv:1703.02507 (2017).

[47] M. E. Peters, et al. "Deep contextualized word representations. arXiv 2018." arXiv preprint arXiv:1802.05365 12 (2018).

[48] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

Design of a Multiloop Controller for a Nonlinear Process

S Anbu[1]
Assistant Professor, Electronics and Instrumentation Engineering Department, Government College of Technology, Coimbatore, 641 013, Tamil Nadu, India
(On Deputation from Annamalai University, Department of Electronics and Instrumentation Engineering, Faculty of Engineering & Technology, Annamalai Nagar, Chidambaram, 608 002, Tamil Nadu, India)

M Senthilkumar[2], T S Murugesh[3]
Associate Professor, Department of Electronics and Communication Engineering, Government College of Engineering Srirangam, Tiruchirappalli, 620 012
Tamil Nadu, India
(On Deputation from Annamalai University, Department of Electronics and Instrumentation Engineering, Faculty of Engineering & Technology, Annamalai Nagar, Chidambaram, 608 002, Tamil Nadu, India)

***Abstract*—Among the category of nonlinear processes, the Continuous Stirred Tank Reactor (CSTR) is one popular unit that finds application in various verticals of chemical process industries. The process variables within the CSTR are highly interactive; hence developing control strategies become a laborious task as it can be viewed as a Multi Input Multi Output (MIMO) system. Often the CSTR is assumed as a Single Input Single Output (SISO) system and during the development of control strategies or algorithm, the main objective is on maintaining only a single process variable closer to its set point, even though many measured variables form part of it. On the contrary, when compared to a SISO system, the MIMO control includes sustaining different controlled variables at their appropriate set points concurrently; thereby achieving an improved efficiency. The components' concentration and the temperature inside the CSTR are highly interactive in nature and exhibit reasonably high non-linear steady state behaviour. Both the interaction and non-linear behaviours pose challenges to the overall system stability. A stabilizing Proportional + Integral (PI) controller employing Stability Boundary Locus (SBL) concept is designed for a CSTR which eventually encapsulates both the stability and closed loop performance in its design procedure and analysed through simulation in MATLAB with the results presented.**

Keywords—Nonlinear process; interaction; multi input multi output control; closed loop performance; stability

I. Introduction

A process industry consists of various process units coupled together to perform a process operation and hence they are typically Multi Input and Multi Output (MIMO) systems. However, to implement process control strategies, process units are treated as Single Input Single Output (SISO) systems [1,2]. The most important notion is in maintaining a particular process variable closer to the set point, although quite a few measured variables are part of it [3]. Generally, the interactions amongst those variables are not considered in the control system design. This results in the augmented use of energy and therefore upsurges the plant's operational costs. In contrast to SISO, the MIMO control objective performs to maintain quite a lot of controlled variables at their desired set points at the same time. For the control of MIMO systems, design technique of single loop tuned controller can't be implemented directly due to the heavy interaction between the loops that inflict intricacies in the control system design. For a MIMO system two kinds of control system design exists. The first type is the multivariable control method, where a solitary control algorithm oversees the control of all interacting loops within the process therefore the fail-safe design becomes more complicated [4]. The second type is the Multiloop control, a kind of multiloop control of individual loop that demonstrates a natural immunity to the loop failure thereby resulting in an easy and potent fail-safe design. The interaction within the CSTR is phenomenal with the Biggest Log modulus Tuning (BLT) [5,6] been implemented earlier [7] and in this present work, the performance of the same CSTR system is analyzed by employing Stability Boundary Locus (SBL) concept through simulation in MATLAB. The SBL concept graphically defines a boundary in the parametric design plane of the controller, to separate the stable and unstable regions of a feedback control system. The value addition of this paper lies in implementing the proposed stabilizing PI controller employing the Stability Boundary Locus (SBL) concept to specify the choice of controller parameters that results in the stable operation of the chosen highly non-linear CSTR process.

The paper is structured as below. In the subsequent sections, related work and then the equations related to the CSTR model and parameters are provided, followed by the multiloop control scheme. The next section details the design steps involved in the MIMO systems to plot the stability boundary loci to calculate stabilizing PI controllers for all the varied operating points of the CSTR. Simulation results to validate the control performance of the proposed method followed by discussion and conclusion are provided afterwards.

II. Related Work

Nusret Tan et al. (2006) have proposed a method to compute all the parameters of a PI controller which stabilize a control system [8]. Hanwate and Hote (2014) have designed a PID controller for cart inverted pendulum system based on the concept of stability boundary locus [9]. A mathematical model of the DC motor control system has been derived by Praboo and Bhaba (2014) based on the model fractional order PI{\lambda} controller using the stability boundary locus method to satisfy the required gain margin (GM) and phase margin (PM) of the system [10]. The work by Deniz et al. (2016) has introduced an integer order approximation method for the numerical implementation of fractional order derivative/integrator operators in control system based on fitting the stability boundary locus (SBL) of fractional order derivative/integrator operators and SBL of integer order transfer functions [11]. A generalized approach to identify all stabilizing PI controllers for processes with time delay that depends on modeling higher order plant transfer functions by a first order plus dead time model has been proposed by Kaya and Atıç (2016) from which the normalized form of the obtained model and controller transfer functions were used for plotting the stability boundary locus plane [12]. The computation of all stabilizing PI controllers for third order systems obtained using the Boundary locus and Kronecker summation method to guarantee the stability of a feedback system was proposed by Amarendra Reddy et al. (2017) [13]. The paper by Atic and Kaya. (2018) has proposed a method by which all PID controller tuning parameters, satisfying stability of any unstable time delay processes, can be calculated by forming the stability boundary loci [14]. A method based on stability region locus and the Mikhailov criterion for stability test has been proposed to determine the parameters of PI controllers to control a TITO (two-input two-output) NCS (networked control systems) with intrinsic and network induced time delays has been proposed by Mohamed-Vall (2021) [15].

III. CSTR Model and Parameters

An irreversible first order exothermic reaction (A→B) occurring in a CSTR as presented in Fig. 1 is considered. A cooling jacket that surrounds the reactor get rid of the heat produced during the reaction. It is assumed that perfect mixing occurs within the CSTR and also any changes in the volume owing to the reaction are considered negligible. As per the above assumptions, jacket water tends to be perfectly mixed with a continual water hold up happening within the jacket and the weight of the CSTR metal walls also being regarded negligible [16]. The fundamental model along with the resultant operating points of the CSTR is provided in Table I [17, 18] and the same has been taken up for the simulation studies.

The dynamics of process variables in the CSTR is given by;

$$\frac{dC_A}{dt} = \frac{F}{V}C_{A0} - \frac{F}{V}C_A - C_A K_0 e^{(-E/RT)} \tag{1}$$

$$\frac{dT}{dt} = \frac{F}{V}T_{in} - \frac{F}{V}T - \frac{H_r C_A K_0 e^{\left(-\frac{E}{RT}\right)}}{\rho C_p} - \frac{UA}{\rho C_p V}(T - T_c) \tag{2}$$

$$\frac{dT_c}{dt} = \frac{F_c}{V_c}(T_{cin} - T_c) + \frac{UA}{\rho_c V_c C_{p_c}}(T - T_c) \tag{3}$$

From the modelling equations of CSTR, it is evident that the process variables C_A, T and T_c remain to be a nonlinear function. Also due to their interactive nature, they cannot be determined independently. The Table I exhibit the CSTR's steady state operational considerations taken up herein.

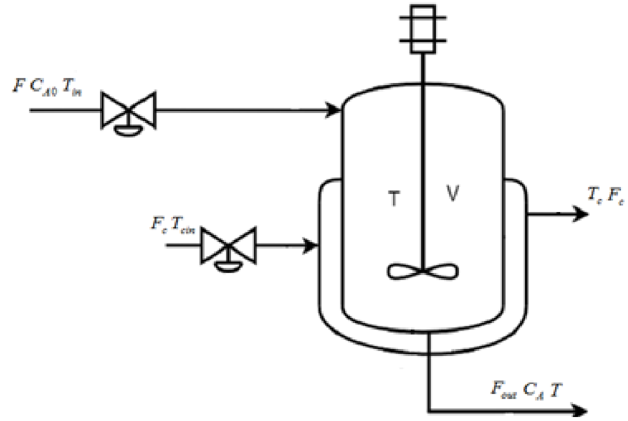


Fig. 1. Continuous Stirred Tank Reactor.

TABLE I. Nominal CSTR Parameters and Variables

Variable	**Description**	**Nominal Operating Values**
V	Reactor volume (l)	50
F_{in}	Reactor's inlet volumetric flow rate (l/min)	50
F_{out}	Reactor's outlet volumetric flow rate (l/min)	50
C_A	Component A concentration in the outlet stream (mole/l)	-
C_{A0}	Feed concentration of component A (mole/l)	1
K_0	Pre-exponential factor (l/min)	$7.8 * 10^{10}$
E	Activation energy in the Arrhenius equation (Cal/mole)	E/R=8567
R	Universal gas constant (Cal/mole. K)	
ρ	Density of inlet and outlet stream (g/l)	900
C_p	Inlet and Outlet streams - Heat capacity (Cal/g.K)	0.329
T	Reactants temperature in the reactor (K)	-
T_{in}	Temperature of the Inlet stream (K)	350
H_r	Reaction's Heat (Cal/mole)	$-5*10^4$
UA	Heat transfer term (Cal/min. K)	$5*10^4$
T_c	Jacket's coolant water Temperature (K)	-
ρ_c	Jacket's coolant water density (g/l)	1000

IV. Multiloop Control of CSTR

A CSTR is a specialized MIMO system that contains two controlled variables (CV), reactor concentration and temperature. These controlled variables have to be maintained at their nominal operating values. Relative Gain Array (RGA) analysis is carried out on the considered CSTR model to recommend the best pairing [19].

Based on the loop pairing, it is proposed that, to attain the best closed loop performance, the flowrate of the inlet coolant water Fc ought to be paired with the concentration of Component A in the outlet stream C_A and the inlet flowrate F need to be paired with the reactor temperature T. The multiloop control structure engaged in this effort is given in Fig. 2.

The RGA analysis is carried out for the proposed CSTR model and is calculated as given in (4).

$$\lambda = \begin{array}{c} Fc \\ F \end{array}\begin{bmatrix} 3.9880 & -2.9880 \\ -2.9880 & 3.9880 \end{bmatrix} \quad (4)$$

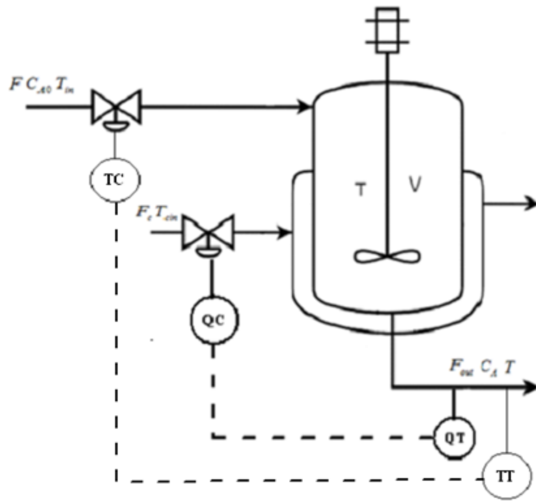


Fig. 2. Multiloop Control Scheme for CSTR.

The RGA analysis specifies that F_C needs to be paired with C_A, and F paired along with T to offer superior control. As λ_{ij} is seen as a small positive value, the gain in closed-loop is considerably greater than in open-loop. This phenomenon may be a source for decline in performance or instability when the loop is closed. This accentuates that, an input in open-loop that possess small influence on a specific output will hold a noteworthy effect in closed-loop owing to feedback as well as coupling.

V. Design of Multiloop Stabilizing PI Controller for CSTR

The robust performance and simplicity of the stabilizing PI (Proportional + Integral), PID (Proportional + Integral + Derivative) and lag/lead design find themselves being extensively used in the process industries. The Stability Boundary Locus (SBL) method performs the computation of the PI controller parameters which stabilizes the control system proposed by [20,21]. A stability boundary locus in the (K_C, K_i) plane is plotted and the stabilizing values of P+I controller parameters are evaluated. This method does not necessitate sweeping over the parameters and also does not require linear programming to solve a set of inequalities. The stabilizing PI controller evolved from the stability boundary locus not only guarantees the stability but also the desired closed-loop transients are obtained [8]. The following sections explain briefly the design steps involved in the MIMO systems.

A. *Stabilizing PI Controller for First Order Plus Dead Time (FOPDT) Process*

For a SISO system with,

$$G_p(s)=G(s)e^{-\theta s}=\frac{N(s)}{D(s)}e^{-\theta s} \quad (5)$$

and a P+I controller of the form

$$G_c(s) = K_c+\frac{K_i}{s} \quad (6)$$

where the problem is to compute the PI controller parameters that stabilizes the specified system [22,23]. The system's closed loop characteristic polynomial P(s) [24] is;

$$P(s) = sD(s) + (K_c s + K_i)N(s)e^{-\theta s}$$

$$= a_n s^n+a_{n-1}s^{n-1}+\ldots a_1 s+a_0 \quad (7)$$

It is to be noted that all coefficients or some of the coefficients a_i, i = 0,1,2….n are the function of K_c, K_i and $e^{-\theta s}$ depends on the order of polynomials N(s) and D(s). There exist 3 eventualities for a stable polynomial's root to cross over the imaginary axes, (i.e., the polynomial turns out to be unstable) in the parameter space approach. (i) Real Root Boundary: At s = 0, a real root crosses over the imaginary axis. As a result, the real root boundary can be attained from P(s) in (7) by substituting s = 0 which provides a_0 = 0. (ii) Infinite Root Boundary: At s = ∞, a real root crosses over the imaginary axis. Therefore, the infinite root boundary can be denoted by considering a_n = 0 from (7) (iii) Complex Root Boundary: The polynomial typified by (7) becomes unstable at s = jω while the roots cross the imaginary axis that signifies the real and the imaginary parts of (7) all becoming zero at the same time. Hence the complex root boundary can be attained as given below. The numerator and denominator polynomials of $G_P(s)$ can be decomposed into their equivalent odd and even parts by replacing s with jω which yields.

$$G(j\omega) = \frac{N_e(-\omega^2)+j\omega N_0(-\omega^2)}{D_e(-\omega^2)+j\omega\omega_0(-\omega^2)} \quad (8)$$

For the sake of simplicity ($-\omega^2$) will not be part of the ensuing equations. The closed loop characteristic polynomial of (7) can be written as;

$$\Delta(j\omega) = \left[(K_iN_e-K_c\omega^2N_0)\cos(\omega\theta)+\omega(K_iN_0+K_cN_e)\sin(\omega\theta)-\omega^2D_0\right]$$

$$+ j\,[\omega(K_iN_0+K_cN_e)\cos(\omega\theta)- (K_iN_e-\omega^2K_cN_0)\sin(\omega\theta)+ \omega D_e]$$

$$= R_\Delta+ jI_\Delta = 0 \quad (9)$$

Upon equating the real and imaginary parts of Δ(jω) to zero;

$$K_c[-\omega^2N_o\cos(\omega\theta)+ \omega N_e\sin(\omega\theta)] +$$

$$K_i[N_e\cos(\omega\theta)+\omega N_o\sin(\omega\theta)] = \omega^2D_o \quad (10)$$

$$K_c[\omega N_e\cos(\omega\theta)+\omega^2N_0\sin(\omega\theta)]+$$

$$K_i[\omega N_0\cos(\omega\theta)-N_e\sin(\omega\theta)] = -\omega D_e \quad (11)$$

Splitting into

$$Q(\omega) = \omega N_e\sin(\omega\theta)] - \omega^2N_o\cos(\omega\theta)$$

$$R(\omega) = \sin N_e\cos(\omega\theta)+\omega N_o\sin(\omega\theta)$$

$$X(\omega) = \omega^2D_o \quad (12)$$

$$S(\omega) = \omega N_e\cos(\omega\theta) +\omega^2N_0\sin(\omega\theta)$$

$$U(\omega) = \omega N_0\cos(\omega\theta) -N_e\sin(\omega\theta)$$

$$Y(\omega) = -\omega D_e \tag{13}$$

Then (10) and (11) can be denoted as

$$K_c Q(\omega) + K_i R(\omega) = X(\omega) \tag{14}$$

$$K_c S(\omega) + K_i U(\omega) = Y(\omega) \tag{15}$$

Solving (14) and (15)

$$K_C = \frac{X(\omega)U(\omega)\text{-}Y(\omega)R(\omega)}{Q(\omega)U(\omega)\text{-}R(\omega)S(\omega)} \tag{16}$$

and

$$K_i = \frac{Y(\omega)Q(\omega)\text{-}X(\omega)S(\omega)}{Q(\omega)U(\omega)\text{-}R(\omega)S(\omega)} \tag{17}$$

Upon substituting (12) and (13) into (16) and (17), it is found that,

$$K_C = \frac{(\omega^2 N_0 D_0 + N_e D_e)\cos(\omega\Theta) + \omega(N_0 D_e\text{-}N_e D_0)\sin(\omega\Theta)}{\text{-}(N_e^2 + \omega^2 N_0^2)} \tag{18}$$

$$K_i = \frac{\omega^2(N_0 D_e\text{-}N_e D_0)\cos(\omega\Theta)\text{-}\omega(N_e D_e + \omega^2 N_0 D_0)\sin(\omega\Theta)}{\text{-}(N_e^2 + \omega^2 N_0^2)} \tag{19}$$

It can be perceived that if the denominator of (18) and (19) $N_e(\text{-}\omega^2) + \omega^2 N_0(\text{-}\omega^2) \neq 0$, then the stability boundary locus, $l(K_c, K_i, \omega)$ can be constructed in the (K_c, K_i)-plane. Whereas if at any specific value of frequency, the denominator of equations (18) and (19) $N_e(\text{-}\omega^2) + \omega^2 N_0(\text{-}\omega^2) = 0$, then it implies that the frequency value should not be used. Here a discontinuous stability boundary locus will be attained which won't be problematic as far as the stabilizing controller's computation is concerned. When the stability boundary locus is attained, it becomes essential to assess if the stabilizing controllers are existent or not. This is because the stability boundary locus, the real root and infinite root boundary lines may perhaps split the parameter plane into stable and unstable regions. It can be seen that the line $K_i = 0$ can split the parameter plane (K_c, K_i) into two regions viz. stable and unstable. In this case the line $K_i = 0$ is the real root boundary line attained by substituting ω = 0 in (7) and then equating it to zero as a real root of Δ(s) of (7) may cross over the imaginary axis at s = 0.

Generally, for a transfer function, the order of D(s) is greater than the order of N(s) which guarantees no infinite root boundary line. The stability boundary locus is seen to be reliant on the frequency ω that varies from 0 to ∞ which signifies the importance of frequency gridding. The lessening of the range of frequencies that desires to be gridded can be effectively achieved by employing the Nyquist plot based method as provided in [25]. Here we need to find only the real values of ω that satisfy Im[G(s)] = 0 where s = jω. As the controller operates in this frequency range, it indicates that the frequency below the critical frequency ω_c or the ultimate frequency can be considered. For that reason, in order to get the stability boundary locus over a likely smaller range of frequency such as $\omega \, \varepsilon \, [0, \omega_C]$, the critical frequency can be employed. As the phase of $G_p(s)$ at $s = j\omega_c$ is equal to −180°, it can be written as;

$$\tan^{-1}\left(\frac{\omega N_O}{N_e}\right) - \tan^{-1}\left(\frac{\omega D_O}{D_e}\right) - \omega\theta = -\pi \tag{20}$$

$$\text{or } \tan(\omega\theta) = \frac{\omega(N_O D_e\text{-}N_e D_0)}{N_e D_e + \omega^2 N_O D_O} = f(\omega) \tag{21}$$

B. Implementation of SBL Algorithm for CSTR Control

From the models obtained in the resultant three operating regions of the CSTR, the stability boundary loci for the loops are obtained using (16) and (17). The boundary loci are depicted in Fig. 3. The (K_c, K_i) points are obtained by varying ω from 0 to ω_c in steps of 0.001 accordingly. The Multiloop controller parameters are obtained from stability boundary locus [26] using weighted geometrical centre with the aid of (22) and (23).

$$K_C = \frac{1}{n}\sum_{j=1}^{n} K_{cj} \tag{22}$$

$$K_i = \frac{1}{2n}\sum_{j=1}^{n} K_{ij} \tag{23}$$

The steady-state profile has yielded the notion of operating the CSTR at three diverse operating regions. In order to investigate the multiloop control of the CSTR, the operating points are prudently selected as per the steady-state input-output response as low (F = 70 l/min, F_C = 60 l/min), middle (F = 31 l/min, F_C = 99 l/min) and high (F = 25 l/min, F_C = 115 l/min).

The steady-state value of the three chosen operating regions is presented in Table II.

TABLE II. OPERATING POINTS OF CSTR-MIMO PROCESS

Operating Region	**C_A (mole/l)**	**T (K)**	**T_C (K)**
Low	0.05725	390.4	341.1
Middle	0.7963	318.3	306.1
High	0.8178	314.1	304.3

The MIMO model for the chosen three operating regions (low, middle and high) obtained using the process reaction curve (PRC) method are specified from (24) to (26) respectively.

$$G(s) = \begin{bmatrix} \frac{0.84e^{-0.1s}}{0.3s+1} & \frac{-0.46e^{-0.15s}}{0.75s+1} \\ \frac{-0.0017e^{-0.15s}}{0.45s+1} & \frac{0.0015e^{-0.25s}}{0.75s+1} \end{bmatrix} \tag{24}$$

$$G(s) = \begin{bmatrix} \frac{0.5e^{-0.1s}}{1.2s+1} & \frac{-0.2e^{-0.05s}}{1.35s+1} \\ \frac{-0.0025e^{-2.05s}}{1.35s+1} & \frac{0.0039e^{-1.35s}}{1.65s+1} \end{bmatrix} \tag{25}$$

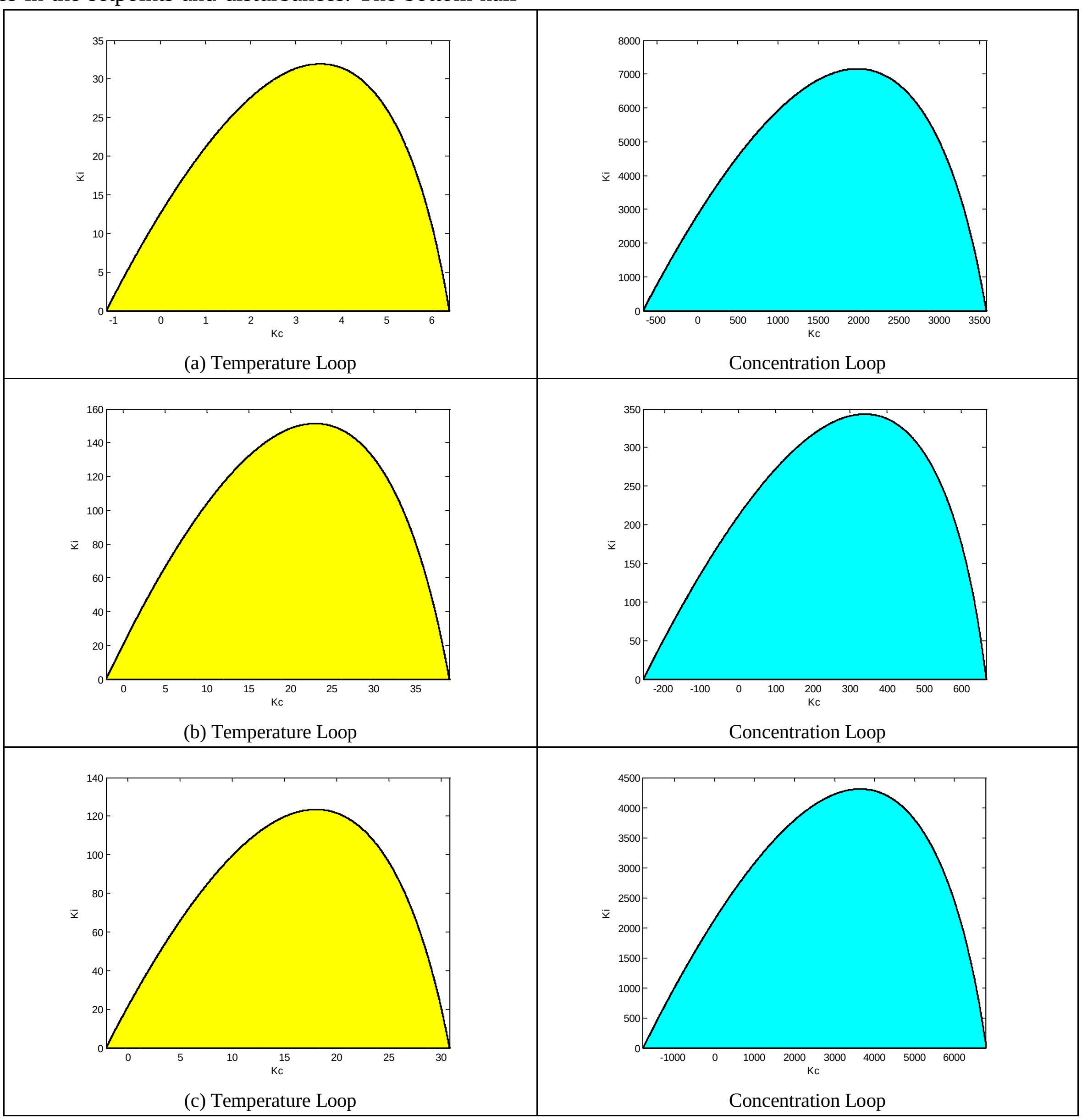
$$G(s) = \begin{bmatrix} \frac{0.48e^{-0.1s}}{0.9s+1} & \frac{-0.04e^{-0.4s}}{0.6s+1} \\ \frac{-0.00068e^{-3.3s}}{1.2s+1} & \frac{0.00056e^{-0.9s}}{1.8s+1} \end{bmatrix} \tag{26}$$

For the locally linearized models, the Multiloop controllers are designed at the chosen three operating points by employing the SBL technique. For the control scheme, the feed flow rate (F) and coolant flow rate (F_C) remain the manipulated variables (MV) whereas the concentration of component A (C_A) and reactor temperature (T) are the controlled variables. The design stage involves finding the PI controller parameters initially by means of the Ziegler Nichols method [27] for individual loop, with the multiloop controller parameters adjusted thereupon by making use of the detuning factor f. For this CSTR process, the detuning factor 'f' for the chosen three operating regions is estimated to be 1.195, 1.275 and 1.26 correspondingly.

C. Control Performance of Multiloop Stability Boundary Locus (SBL) method – Reactor Concentration Control

To investigate the closed loop system behaviour with the controller so designed, the concentration setpoints are changed over all the three operating regions. The Feed concentration C_{A0}, the temperature of the Inlet stream T_{in} and the inlet coolant water temperature in the jacket T_{cin} act as the disturbances for this system, which are intentionally changed from the nominal values at various sampling instants. All the figures in this and subsequent sections consist of two halves. The top half shows the trends of the process variables (PV) for the changes in the setpoints and disturbances. The bottom half displays the corresponding changes in the manipulated variable which drives the PV to the setpoints. Fig. 4, 6 and 8 show the servo and regulatory responses for the operated low, middle and high regions.

It is found that the reactor concentration tracks the setpoint changes without any offset and at the same time the disturbance rejection is also phenomenal. The three different disturbances in the CSTR mentioned earlier are introduced intentionally at the sampling instances of 40, 80 and 190 respectively. The rejection trend shows how effectively the disturbances are removed. Fig. 5, 7 and 9 show the interaction effect of concentration on the reactor temperature. In all the above-mentioned figures it is found that the process being MIMO in nature, remains susceptible to the interaction effect. However, the degree of interaction is very much reduced due to the effective SBL design methodology.



(a) Temperature Loop | Concentration Loop

(b) Temperature Loop | Concentration Loop

(c) Temperature Loop | Concentration Loop

Fig. 3. Stability Regions of the CSTR Process (a) Low (b) Middle (c) High.

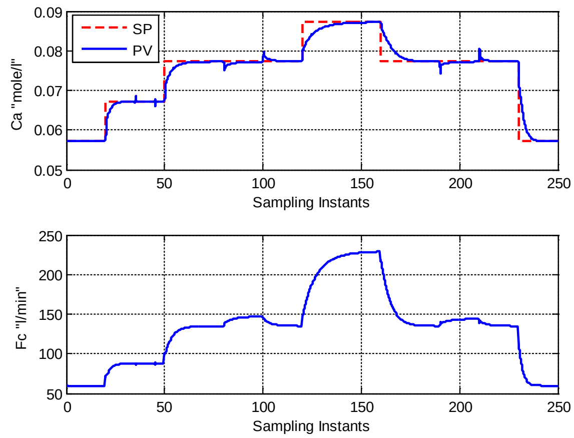


Fig. 4. Servo and Regulatory Responses of Multiloop SBL Design for Concentration Control at Low Region.

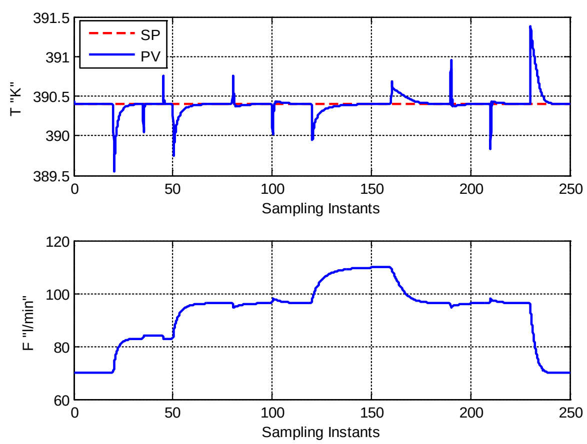


Fig. 5. Interaction Behaviour of Temperature with SBL Design for Concentration Control at Low Region.

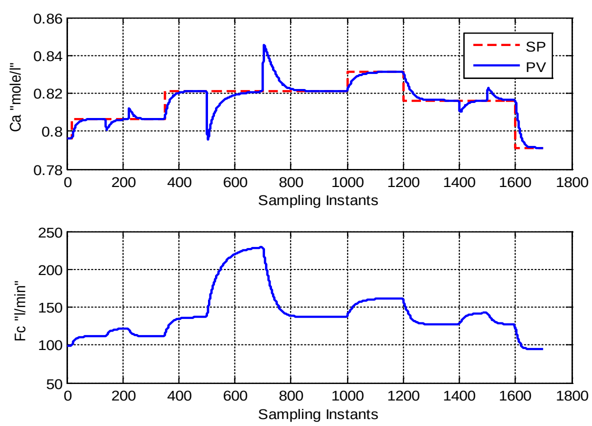


Fig. 6. Servo and Regulatory Responses of Multiloop SBL Design for Concentration Control at Middle Region.

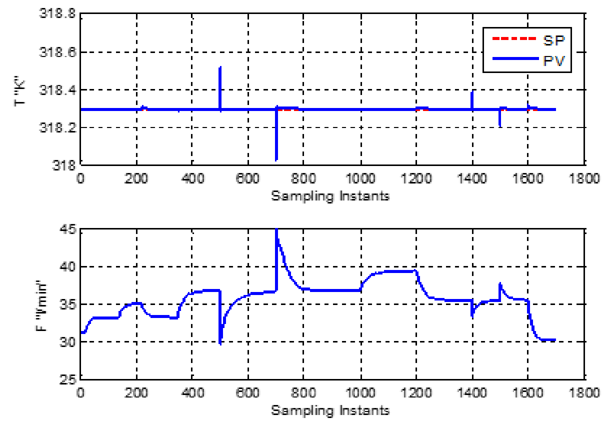


Fig. 7. Interaction Behaviour of Temperature with SBL Design for Concentration Control at Middle Region.

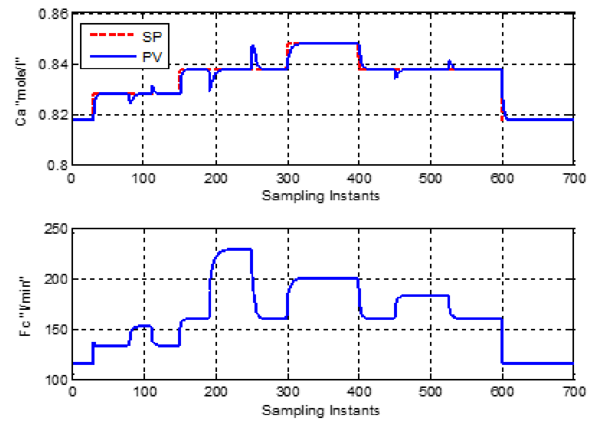


Fig. 8. Servo and Regulatory Responses of Multiloop SBL Design for Concentration Control at High Region.

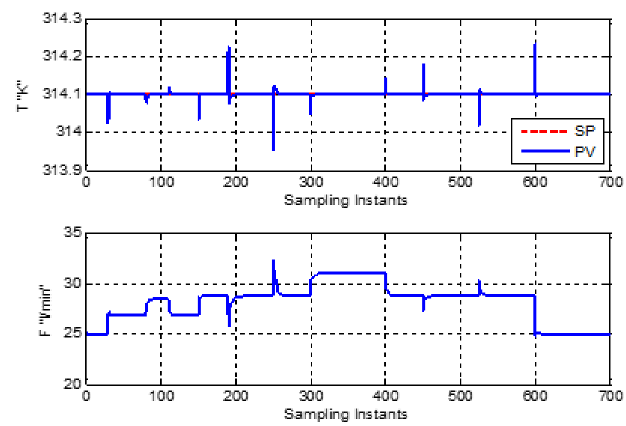


Fig. 9. Interaction Behaviour of Temperature with SBL Design for Concentration Control at High Region.

D. Control Performance of Multiloop Stability Boundary Locus (SBL) Method - Reactor Temperature Control

The closed-loop performances of the designed controller are studied in all three chosen operating regions of the CSTR. The servo and regulatory responses are obtained for the analogous setpoint and disturbance patterns employed already for concentration control. Fig. 10, 12 and 14 show the servo and regulatory responses. The step change in the setpoints is made at various sampling instances in both directions to check the setpoint tracking feature of the designed controller. It is found that the designed controller behaves well.

The bottom halves of the figure show the trend of manipulated variable (the feed flow rate). It is also observed that the SBL method smoothly rejects the interaction effect on concentration due to changes in temperature setpoints and disturbances which are recorded in Fig. 11, 13 and 15.

The spikes in the trends on the top halves of the figures show the deviation of concentration from its constant nominal values at various sampling instances. The servo and regulatory responses are obtained for both the concentration and temperature control as per the controller tuning constants obtained using multiloop controllers based on the Stability Boundary Locus method and the same is provided in Table III.

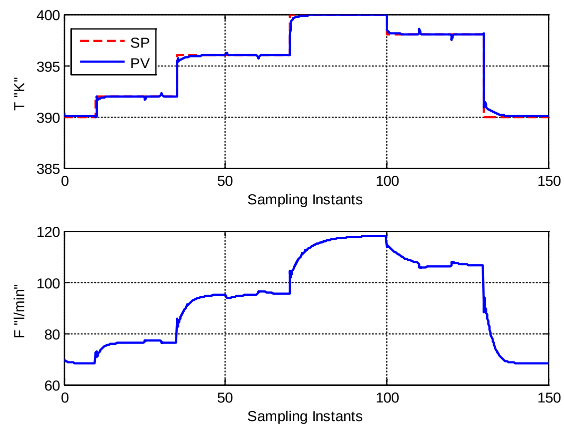


Fig. 10. Servo and Regulatory Responses of Multiloop SBL Design for Temperature Control at Low Region.

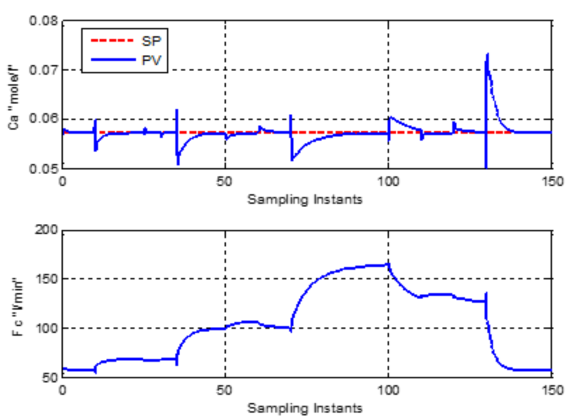


Fig. 11. Interaction Behavior of Concentration with SBL Design for Temperature Control at Low Region.

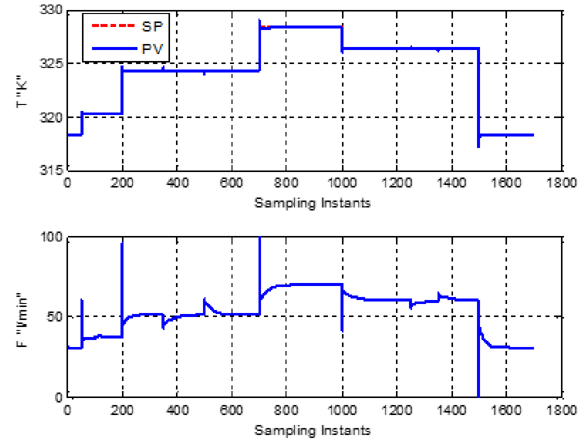


Fig. 12. Servo and Regulatory Responses of Multiloop SBL Design for Temperature Control at Middle Region.

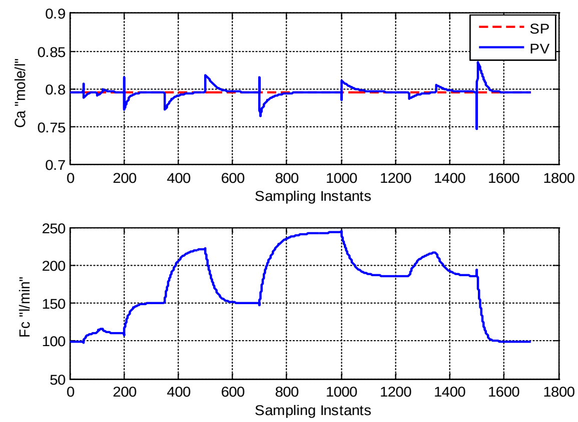


Fig. 13. Interaction Behavior of Concentration with SBL Design for Temperature Control at Middle Region.

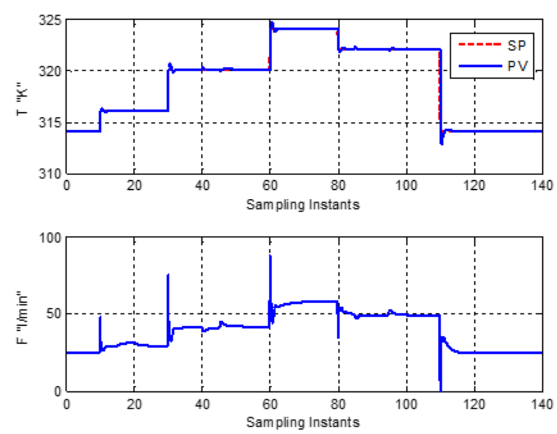


Fig. 14. Servo and Regulatory Responses of Multiloop SBL Design for Temperature Control at High Region.

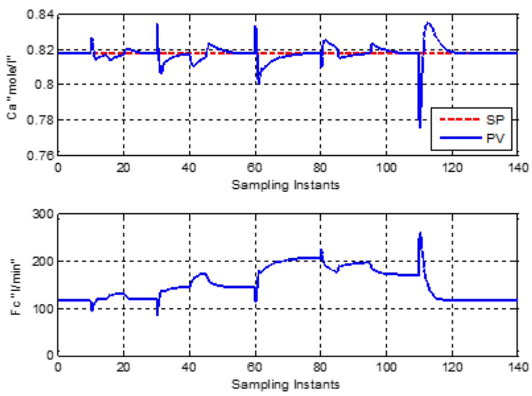


Fig. 15. Interaction Behavior of Concentration with SBL Design for Temperature Control at the High Region.

TABLE III. MULTILOOP CONTROLLER PARAMETERS FOR CSTR PROCESS USING SBL DESIGN METHOD

Operating Region	K_C		K_i	
	Temperature Loop	Concentration Loop	Temperature Loop	Concentration Loop
Low	1.9942	1117.5	8.6737	1.942
Middle	14.7728	146.9798	40.9988	93.4175
High	11.4149	1875.0	33.4506	1170.5

TABLE IV. PERFORMANCE INDICES OF MULTILOOP SBL DESIGN FOR SETPOINT CHANGE DURING CONCENTRATION CONTROL

Operating Regions	Setpoint Change	ISE		IAE	
		T	C_A	T	C_A
Low	0.0572 – 0.0673	6.028386	0.000579	15.0567	0.149571
	0.0673 - 0.0772	4.29804	0.000921	15.41896	0.247092
	0.0772 - 0.0872	2.623592	0.001821	15.6639	0.487632
	0.0872 - 0.0772	2.050618	0.00241	15.66487	0.487642
	0.0772 - 0.0572	1.18E-07	0.0004	0.000303	0.020007
Middle	0.7963 - 0.8063	0.001325	0.006266	0.497038	1.306073
	0.8063 - 0.8163	0.002623	0.018643	0.842773	2.70376
	0.8163 - 0.8263	0.00294	0.041567	1.484041	5.97387
	0.8263 - 0.8163	0.002003	0.029539	0.941316	3.686016
	0.8163 - 0.7963	0.007016	0.048059	1.272217	3.524355
High	0.8178 - 0.8278	0.034768	0.000932	0.67775	0.163779
	0.8278 - 0.8378	0.025621	0.001183	0.598351	0.228391
	0.8378 - 0.8478	0.017796	0.001645	0.674321	0.343457
	0.8478 - 0.8378	0.014713	0.001831	0.674334	0.343457
	0.8378 - 0.8178	0.114312	0.004482	1.228779	0.39217

The servo and regulatory performances of the designed controller are analyzed by computing the ISE (Integral square error) and IAE (Integral Absolute Error) [28] at various operating points within the three operating regions of the CSTR. The performance indices are computed during concentration and temperature control and tabulated. Table IV and Table VII display the performance measures calculated for each setpoint change during concentration and temperature control respectively. The performance measures computed during disturbance changes are given in Table V and Table VI for concentration and temperature control, respectively.

TABLE V. PERFORMANCE INDICES OF MULTILOOP SBL DESIGN FOR DISTURBANCE CHANGE DURING CONCENTRATION CONTROL

Operating Regions	Disturbance Change	ISE		IAE	
		T	C_A	T	C_A
Low	C_{A0}	0.261004882	3.38187E-06	1.145382264	0.00461
	T_{cin}	0.427337781	5.19373E-05	3.433905973	0.061376
	T_{in}	0.560756951	3.78241E-05	2.727808278	0.045502
Middle	C_{A0}	0.000944948	0.003304249	0.437309074	1.082304
	T_{cin}	0.261543941	0.121135311	3.214269128	9.889035
	T_{in}	0.639564818	0.309675259	8.537920774	26.11381
High	C_{A0}	0.006169874	0.000339233	0.49342688	0.157271
	T_{cin}	0.086943188	0.003011574	1.613168071	0.580593
	T_{in}	0.023119094	0.000446222	0.70756471	0.201456

TABLE VI. PERFORMANCE INDICES OF MULTILOOP SBL DESIGN FOR DISTURBANCE CHANGE DURING TEMPERATURE CONTROL

Operating Regions	Disturbance Change	ISE		IAE	
		T	C_A	T	C_A
Low	C_{A0}	0.205901436	1.98753E-06	1.013312227	0.003348628
	T_{cin}	0.275668499	2.09492E-05	2.854061744	0.034021004
	T_{in}	0.654673618	7.08659E-06	2.251541109	0.013654934
Middle	C_{A0}	0.001294486	0.00268816	0.372627445	0.694312277
	T_{cin}	0.399054716	0.087630653	3.711265015	7.675565711
	T_{in}	0.135731116	0.017026752	1.75823604	3.348912205
High	C_{A0}	0.009394668	0.000210094	0.487764198	0.094506423
	T_{cin}	0.102501726	0.001058962	1.434894951	0.211575404
	T_{in}	0.1104586	0.000212938	1.209813883	0.123608686

TABLE VII. PERFORMANCE INDICES OF MULTILOOP SBL DESIGN FOR SETPOINT CHANGE DURING TEMPERATURE CONTROL

Operating Regions	Setpoint Change	ISE		IAE	
		T	C_A	*T*	C_A
Low	390 - 392	7.278633732	9.19617E-05	10.30603635	0.059160341
	392 - 396	29.30308893	0.00048227	23.56951958	0.170717264
	396 - 400	29.51332384	0.000925221	28.19125551	0.332755641
	400 - 398	7.363435376	0.000355801	13.07605805	0.169436652
	398 - 390	64.08801105	8.71647E-09	8.005466429	9.31349E-05
Middle	318.3 - 320.3	4.748005943	0.006064203	4.636015839	1.262702397
	320.3 - 324.3	19.66531355	0.050494916	10.62003549	4.39818696
	324.3 - 328.3	21.24057865	0.138593034	13.02562158	10.09649517
	328.3 - 326.3	5.194835414	0.050938797	6.096147516	6.234963292
	326.3 - 318.3	64.00049012	1.13771E-08	8.000030629	0.000106654
High	314.1 - 316.1	5.098656774	0.000610959	5.200968794	0.150966628
	316.1 - 320.1	21.27278662	0.002822406	11.40449762	0.379213217
	320.1 - 324.1	22.31137176	0.005225345	12.51482279	0.641269532
	324.1 - 322.1	5.4637475	0.001563806	5.576953456	0.265346417
	322.1 - 314.1	64.00416577	1.07172E-08	8.000260355	0.000103522

VI. DISCUSSION

One of the important traits of a Multiloop controller is the smooth removal of interaction among the loops thereby providing the operator with the freedom to operate near the limits. The SISO control design conveniently ignores the interaction effect and thereby made the design easy at the cost of performance degradation. As the CSTR is operated as a MIMO system, the interaction of change in the temperature on the concentration and vice versa are dominant. The performance of the designed Multiloop PI controller using the SBL design method is analysed by intentionally perturbing the CSTR with setpoint and disturbances at three different operating conditions. The performance is quantified for the control of both the concentration and temperature with ISE and IAE. The effect of interaction is captured and presented alongside the servo and regulatory response plots at various operating conditions. They show that interaction plays a meagre role under SBL control and hence only a slight deviation from the nominal values are recorded. By a scrutiny of the response plots along with the performance metrics it is found that the suggested Multiloop control based on SBL design method performed indubitably well under various test conditions as presented in the paper.

VII. CONCLUSION AND FUTURE WORK

A Multiloop control philosophy is tried on the CSTR utilizing a powerful SBL design technique. One interesting feature of the SBL method lies in its simple design procedure, which involves considering only the diagonal elements, hence the interaction dynamics are discarded during its design phase. As an extension of this work, Multivariable control strategies can be implemented to analyze the interaction effects.

REFERENCES

[1] W. L. Luyben. (1986): A simple method for tuning SISO controllers in multivariable systems. Industrial & Engineering Chemistry Process Design and Development. 25(3), pp. 654-660.

[2] T. J. Monica, C. C. Yu and W. L. Luyben. (1988): Improved multiloop single-input/single-output (SISO) controllers for multivariable processes. Industrial & engineering chemistry research. 27(6), pp. 969-973.

[3] S. Cha, D. Chun and J. Lee. (2002): Two-step IMC-PID method for multiloop control system design. Industrial & engineering chemistry research. 41(12), pp. 3037-3041.

[4] N. Kamala, T. Thyagarajan and S. Renganathan. (2010): Multivariable Control of Nonlinear Process. Int. J. of Recent Trends in Engineering and Technology. 4(4), pp. 106-110.

[5] Q. G. Wang, T. H. Lee and Y. Zhang. (1998): Multiloop Version of the Modified Ziegler-Nichols Method for Two Input Two Output Processes. Industrial & engineering chemistry research. 37(12), pp. 4725-4733.

[6] S. Mohanraj, T. S. Murugesh and M. Senthilkumar. (2021): Design of decentralized controller for coupled tank system using BLT method, Elsevier, Materials Today: Proceedings. 46(20), pp. 11198-11201.

[7] S. Anbu. (2016): Multiloop Control of Continuous Stirred Tank Reactor Using Biggest Log Modulus Method, Asian Journal of Electrical Sciences. 5(2), pp. 54-61.

[8] N. Tan, I. Kaya, C. Yeroglu and D. P. Atherton. (2006): Computation of stabilizing PI and PID controllers using the stability boundary locus, Energy Conversion and management. 47(18-19), pp. 3045-3058.

[9] S. D. Hanwate and Y. V. Hote. (2014): Design of PID controller for inverted pendulum using stability boundary locus, 2014 Annual IEEE India Conference (INDICON), pp. 1-6, doi: 10.1109/INDICON.2014.7030563.

[10] Praboo, N. N., & Bhaba, P. K. (2014): Simulation work on Fractional Order PI {\lambda} Control Strategy for speed control of DC motor based on stability boundary locus method. arXiv preprint arXiv:1402.1347.

[11] Deniz, F. N., Alagoz, B. B., Tan, N., & Atherton, D. P. (2016): An integer order approximation method based on stability boundary locus for fractional order derivative/integrator operators. ISA transactions, 62, 154-163.

[12] İ. Kaya and S. Atıç. (2016): PI controller design based on generalized stability boundary locus. 20th International Conference on System Theory, Control and Computing (ICSTCC), pp. 24-28, doi: 10.1109/ICSTCC.2016.7790635.

[13] B. Amarendra Reddy, S. Muaz and M. Ramesh. (2017): Computation of Stability Regions using Boundary Locus and Kronecker Summation Methods," 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC), pp. 986-991, doi: 10.1109/CTCEEC.2017.8455200.

[14] S. Atic and I. Kaya. (2018): PID Controller Design Based on Generalized Stability Boundary Locus to Control Unstable Processes with Dead Time. 26th Mediterranean Conference on Control and Automation (MED), pp. 1-6, doi: 10.1109/MED.2018.8442568.

[15] Mohamed-Vall, O. M. (2021): Design of decoupled pi controllers for two-input two-output networked control systems with intrinsic and network-induced time delays. acta mechanica et automatica, 15(4).

[16] S. Anbu and N. Jaya. (2014): Design of gain scheduling adaptive control for continuous stirred tank reactor, International Journal of Automation and Control. 8(2), pp. 141-157.

[17] N. F. Thornhill, S. C. Patwardhan and S. L. Shah. (2008): A continuous stirred tank heater simulation model with applications, Journal of Process Control. 18(3), pp. 347-360.

[18] P. Albertos and A. Sala. (2006). Multivariable control systems: An engineering approach, Springer Science & amp; Business Media.

[19] E. H. Bristol. (1966): On a new measure of interaction for multivariable process control, IEEE Transactions on Automatic Control. 11(1), pp. 133-134.

[20] N. Tan. (2005): Computation of stabilizing PI and PID controllers for processes with time delay, ISA transactions. 44(2), pp. 213-223.

[21] R. Matušů. (2011): Calculation of all stabilizing PI and PID controllers. International Journal of Mathematics and Computers in Simulation. 5(3) pp. 224-231.

[22] M. Senthilkumar and S. A. Lincon. (2012): Design of stabilizing PI controller for coupled tank MIMO process, International Journal of Engineering Research and Development. 3(10), pp. 47-55.

[23] M. Senthilkumar and S. A. Lincon. (2015): Multiloop PI Controller for achieving simultaneous Time and Frequency domain specifications, Journal of Engineering Science and Technology. 10(8), pp. 1103-1115.

[24] S. E. Hamamci and N. Tan. (2006): Design of PI controllers for achieving time and frequency domain specifications simultaneously, ISA transactions. 45(4), pp. 529-543.

[25] N. Munro and M. T. Soylemez. (2000): Fast calculation of stabilizing PID controllers for uncertain parameter systems, IFAC Proceedings. 33(14), pp. 549-554.

[26] C. Onat. (2013): A new concept on PI design for time delay systems: weighted geometrical center, International Journal of Innovative Computing, information and control. 9(4), pp. 1539-1556.

[27] J. C. Basilio and S. R. Matos. (2002): Design of PI and PID controllers with transient performance specification, IEEE Transactions on education. 45(4), pp. 364-370.

[28] B. W. Bequette. (2003): Process control: Modeling, design, and simulation. Prentice-Hall Professional.

Appendix A

MIMO	Multi Input Multi Output
SISO	Single Input Single Output
CSTR	Continuous Stirred Tank Reactor
BLT	Biggest Log modulus Tuning
PI	Proportional + Integral
PID	Proportional + Integral + Derivative
RGA	Relative Gain Array
FOPDT	First Order Plus Dead Time
SBL	Stability Boundary Locus
ISE	Integral square error
IAE	Integral Absolute Error
K_C	Proportional Gain
K_i	Integral Gain

License Plates Detection and Recognition with Multi-Exposure Images

Seong-O Shim[1]*
Department of Computer & Network Engineering, College of Computer Science and Engineering
University of Jeddah
Jeddah, Saudi Arabia

Romil Imtiaz[2], Asif Siddiq[3]
Electrical Engineering Department
Pakistan Institute of Engineering & Technology
Multan, Pakistan

Ishtiaq Rasool Khan[4]
Department of Computer Science and AI, College of Computer Science and Engineering
University of Jeddah
Jeddah, Saudi Arabia

***Abstract*—Automatic License Plate Recognition (ALPR) has been an important research topic for many years in the intelligent transportation system and image recognition fields. License Plate (LP) detection and recognition has always been a challenging issue due to several factors, including different weather and lighting, unavoidable data acquisition noise, and requirement for real-time performance in state-of-the-art Intelligent Transportation Systems (ITS) applications. Different techniques have been proposed based on machine learning, deep learning, and image processing for the detection and recognition of LPs. This paper proposes a method that performs vehicle LP detection and character recognition with high accuracy by using artificially generated multi-exposure images of the LP. First, one under-exposed and three over-exposed images are generated from a reference image taken from the camera. Then, LP detection and character recognition algorithms are applied on these five images – one real image and four synthesized images. At each character location in LP, the character detected with the highest confidence level among these images is selected as the final predicted character. The system is fully automated, and no pre-processing, calibration, or configuration procedures are needed. Experimental results show that the proposed method achieves high accuracy and works robustly even in challenging conditions. The proposed method can be used in any existing ALPR system and the results on three recent ALPR techniques show that the accuracies are further improved when they are combined with the proposed method.**

Keywords*—License plate detection; character recognition; multi-exposure images; convolutional neural networks

I. Introduction

Automatic License Plate Recognition (ALPR) has become a popular topic of study and development in recent years. In modern urban societies, there is a constant demand for powerful intelligent systems for transport security and monitoring, as well as to assist in automating toll collection [1], traffic law enforcement [2], border control [3], private space access control [4], and road traffic monitoring [5]. Intelligent Transportation Systems (ITS) became a necessity in the quest to turn cities into smart cities. ALPR is an important part of ITS because it includes complex computer vision operations of object segmentation, and recognition [6, 7, 8]. As a result, there is a need to make the license plate (LP) recognition systems robust in a wide variety of backdrops and other environmental elements such as illumination, camera angle, and noise and distortion level in images, which make ALPR a challenging task. To recognize car licenses plates in restricted backdrops, a variety of advanced computer vision technologies and artificial intelligence algorithms have been developed [5, 7, 9], which may not work very well in dynamic conditions. In addition, it is impossible to cover entire range of roads, highways, or motorways with stationary/fixed ALPR cameras on ground, bridges, and utility poles [10]. Therefore, it is important to solve the relevant technical issues and develop ALPR systems for dynamic backgrounds, which can be mounted on mobile platform to patrol the area.

The main contribution of this paper is to increase the accuracy of detection and recognition of licenses plates with the help of multiple images taken at different exposure times. However, to avoid the use of complex hardware, and to keep the technique compatible with the existing ALPR systems, we capture only one image at a fixed exposure setting and synthesise others variants using a camera model. This model, for each image captured by a mobile camera, generates four additional images for different exposure times 0.5, 1.5, 2, and 2.5. Thus, we have five images of a scene at different levels of brightness. Our hypothesis is that in dark and very bright conditions when the original image is under/over exposed, some of the synthesisesd images should show the LP and its characters better than the original image captured by the camera. ALPR can be done on each of the five images and results can be combined to detect and recognize LP at high accuracy.

To validate the performance of the proposed technique, we develop a deep learning based baseline ALPR system, and observe the improvement in its accuracy as result of applying the proposed method. Our ALPR system uses You Only Looks Once (YOLO) algorithm [11] to detect licenses plates. After the detection process, the images are fed into a Convolutional Neural Network (CNN) model for the recognition of characters and digits on the plate. We train both modules using a custom dataset of LPs extracted from real traffic videos captured in Saudi Arabia. Using multiple exposures as suggested above, the accuracy of the plate detection reaches 100%. The recognition phase is more challenging, as we have included some very challenging LP images, having faded digits, rusted plates, and varying size and lighting. However, we see a very significant improvement in recognition too, when images at multiple exposures are used. Our CNN model returns the confidence level with which a character is recognized. We pick the individual characters predicted with the highest confidence level, if there are mismatches found in the results obtained from multiple plates.

As we capture only one real image of the vehicle at a fixed exposure setting, and generate the rest artificially, the proposed technique works with existing ALPR systems without requiring any changes in their capturing modules. It can be used as an add-on to their existing software. We experiment with several existing ALPR systems and observe significant increase in their accuracy, especially in the challenging scenarios.

The rest of work is organized as follows. In Section II, existing works related to each ALPR stage are discussed. This is followed by presenting the fundamental idea of the proposed ALPR method in Section III. Experiment results are presented in Section IV. Concluding remarks and ideas for future work are given in Section V.

II. Related Work

For LP detection, segmentation, and identification, several ALPR systems have been suggested. Various object detection approaches have been used by researchers to handle the LP detection step. Chen et al. [12] have reduced the YOLO model layers from 27 to 13, consisting of 7 CNN layers and 6 fully connected layers. They call this a tiny model which is used to detect only one single class. A total of 36 classes are used in which 10 classes are used for numeric data, 25 for alphabets, and 1 for plate recognition. A Taiwan License plates dataset was used and results were reported with 98.22% detection accuracy and 78% recognition accuracy.

Tourani et al. [13] divide LPR into two parts: plate detection and character recognition. A YOLOv3 model is used to detect the LP which is cropped from the image and resized to 224 x 224 pixels. Both color and grayscale images are used and the last layer of the first YOLO model is modified. For character recognition, a second YOLOv3 model is used. The method got 97.77% accuracy in detection and 95.05% in character recognition on Iranian license plates dataset.

Laroca et al. [1] used two CNN arranged in a cascaded manner to detect car frontal or back-views and LPs, having the lowest false positive rate. Rashtehroudi et al. [14] have combined the Optical Character Recognition (OCR) and segmentation method which is used for digit recognition with the help of the YOLO algorithm. Using some preprocessing steps in the first stage of the technique, they remove camera angle issues by applying a Hough Line transform and rotational filters based method on the license plate. In the second stage, they apply the Bradley method to minimize the effect of light angle on the image. After this, the noise level is reduced, and a YOLO model is trained with 1000 images. Ozbay et al. [2] used a smearing method which is a morphological process on preprocessed images. Gou et al. [15] used top-hat transformation, vertical edge detection, and morphological procedures to detect the license plate.

Yonetsu et al. [16] used two-stage YOLOv2 for precise license plate detection to decrease false detection, and thus boost the detection rate. Jain et al. [17] used edge information and geometric characteristics to extract license plate candidates, which are then fed to a CNN classifier for license plate detection. Liu et al. [18] used a YOLOv3 detector. In their method, the picture is divided into T × T (where T is a natural integer) rectangles, and bounding boxes and class probability are predicted in each rectangle. The rectangle with the highest confidence level is detected in the proposed technique. Wang et al. [19] utilize a vertical projection approach in which they scan license frames from left to right until a projection region with a width bigger than a preset threshold is discovered. Izidio et al. [20] created a CNN model based on synthetic imagery. Khan et al. [21] purposed a framework for LP recognition addressing the issues of low image resolution, light conditions, blurriness effect due to motion of vehicle, height, and noise. Zhang et al. [22] created three networks in their proposed method. The first network is a CNN cascade model for fast and accurate plate detection. The second network is Recurrent Neural Network (RNN) for a segmented free recognition. The third network is Generative Adversarial Networks (GAN) to enhance ALPR accuracy.

III. Proposed Method

In this section, we explain the method to synthesise images at multiple exposure from a given image captured at a fixed exposure. Then we explain several steps in designing an ALPR system that leverges the information contained in multiple images to achieve high detection and recognition rates.

A. Synthesising Multiple Exposures

The quality of images taken by a camera is affected by ambient light, and in very dark or bright images, it becomes difficult to discern the characters. Therefore, it is necessary to adjust the exposure time of the camera according to surrounding lighting conditions. However, it is not always possible to find a suitable setting automatically, and therefore an easier way could be to acquire multiple images at different exposure times and read the LP from the best image to increase the recognition accuracy. Capturing multiple images of a scene at different exposure settings requires more sophisticated hardware to achieve higher frame rate, especially when the vehicle and the camera are moving. In this article, we use a camera model to artificially create various images corresponding to different camera exposure times (multi-exposure images) from a single image. The proposed technique does not require changes in the existing hardware used for LP recognition and is a post-processing module that can be added to the existing systems for enhanced accuracy.

The accumulated light at a sensor location p for Δt units of time defines a sensor exposure $E(p) \cdot \Delta t$, where $E(p)$ is sensor irradiation at p . Then, the pixel intensity value $L(p)$ at sensor location p is defined as a function of sensor exposure as.

$$L(p) = f(E(p) \cdot \Delta t) \tag{1}$$

where, f is called the camera response function. Among various models to estimate the camera response function [23, 24], many camera manufacturers adopt a gamma curve to design the response function [25, 26]. In general, a camera response function is assumed to be a gamma curve, in which case the intensity value at pixel location p is represented as.

$$L(p) = f(E(p) \cdot \Delta t) \approx (E(p) \cdot \Delta t)^{\gamma} \quad (2)$$

If multiple images $L_1, L_2, \ldots, L_n$ are acquired at different camera exposure times $\Delta t_1, \Delta t_2, \ldots, \Delta t_n$, we can obtain a gamma corrected exposure time ratio $\alpha_i^j(p)$ [26] between $L_i(p)$ and $L_j(p)$ as

$$\alpha_i^j(p) = \frac{L_j(p)}{L_i(p)} \approx \left(\frac{E(p)\cdot\Delta t_j}{E(p)\cdot\Delta t_i}\right)^{\gamma} = \left(\frac{\Delta t_j}{\Delta t_i}\right)^{\gamma} \quad (3)$$

Since $\Delta t_i, \Delta t_j, \gamma$ are known constants and the value $\alpha_i^j(p)$ is the same for all the pixels in an image, the image L_k corresponding to camera exposure time Δt_k can be artificially generated from the reference image L_r by multiplying α_r^k at each pixel in L_r. In this article, four images are generated: one under-exposed image (α = 0.5) and three over-exposed images (α = 1.5, 2, 2.5) with respect to the reference image L_r with the expectation that one of these five images (including the reference image) will provide better quality for discerning LP numbers under given ambient lighting conditions.

We apply our LP detection and recognition algorithm proposed next in this section to each of these five images. The algorithm returns a value of confidence level for each detected alphanumeric. Among the characters detected multiple times in these five images, we take the one detected with the highest confidence level.

B. Data Collection

To improve the detection model's accuracy, an adequate amount of labeled data is required. In this part, we explain the data collection procedure to train and test the LP detection and recognition networks. First, the videos of vehicles were recorded using a camera positioned on a moving car. The data is collected from the real traffic in Saudi Arabia and includes different LPs of different color and size according to vehicle type and size. In our study, we only consider cars' LPs which are 32cm by 16cm in size. We created a framing algorithm that extracts frames from the recorded videos. A total of 950 frames containing cars were extracted and stored in the Portable Network Graphics (PNG) format at 1920x1080 resolution. The dataset comprises frames containing LPs with easily readable text, as well as several challenging LPs in which text is faded, covered with mud, or distorted in several other ways. Some samples are shown in Fig. 1.

After extracting these frames, we applied different exposures (α = 0.5, 1.5, 2, 2.5) to them using the method described above and thus a total of 3800 frames were obtained. A typical set of five frames is shown in Fig. 2, where the frame marked as 1x is captured by the camera while the rest are generated using our model stated above. To produce more realistic scenarios, we applied augmentation techniques such as gray scaling, brightness variation, rotation of LPs, and flipping and rotation of digits and characters. This way, in total, we got 5000 unique images of LPs in our dataset.

Each frame in the dataset is labeled using LabelImg tool [27]. The software gives XML files that contain coordinates of bounding boxes.

C. Detection and Recognition

We train YOLOv5 network [28] for detection of LPs, on a computer equipped with Rx 570 GPU. YOLOv5 is considered a better model for object detection compared to its competitors YOLO [18], YOLOv3 [20], and Faster-RCNN [29] in terms of storage, processing speed, and auto anchoring. The training images are fed into the object detection model YOLOv5, along with the XML files extracted in the previous step mentioned above. Using transfer learning, we extract the weights of YOLOv5. As shown in Table I, we split the dataset: 70% data for training the model, 10% for the validation process, and 20% for testing. At convergence, we got 98% accuracy in LP detection on the testing dataset. After detecting LP, we move to the next step which is the segmentation of the dataset and distributing the LP characters among 27 classes of digits and alphabets. LPs in Saudi Arabia use only 17 characters and 10 digits, as shown in Fig. 3, and there are three letters and four digits in an LP. For the recognition process, a convolutional neural network with convolution layers at the beginning and fully linked layers at the end is used, as shown in Fig. 4, where the layers are labeled as follows:

Conv<number of filters><size of filters>

Fig. 1. Some Samples of License Plates in our Dataset.

Fig. 2. Five Images of the Same Scene at different Camera Exposures. The Second Image from Left is Captured with Camera while the Rest are Synthesies using a Camera Model.

TABLE I. DATA SPLIT RATIO OF TRAINING DATASET FOR LP DETECTION

Dataset	**Split Percentages**	**Number of Frames**
Training	70%	3500
Testing	20%	1000
Validation	10%	500

Fig. 3. LP Dataset.

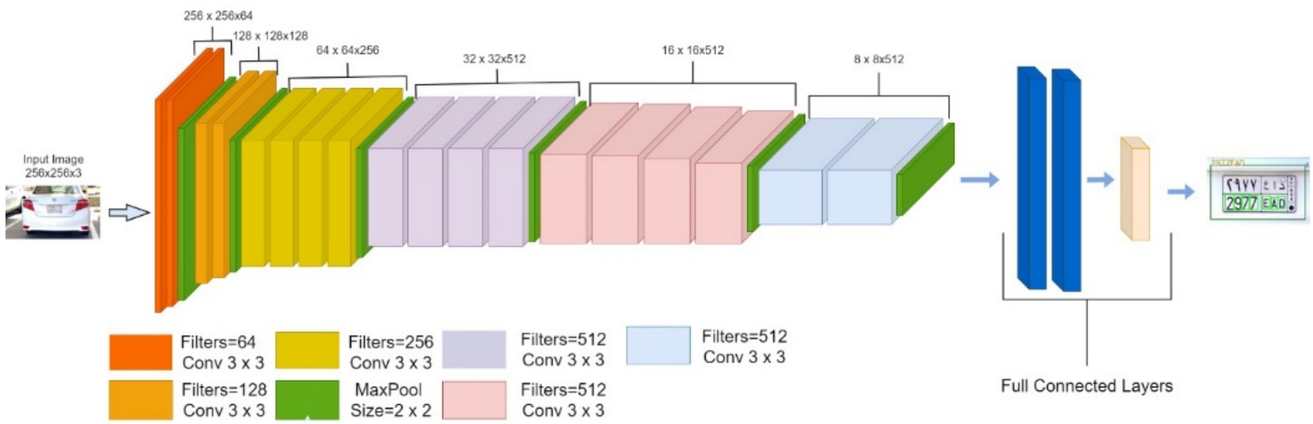


Fig. 4. Proposed CNN Architecture.

The size of the kernel is 3 x 3 and input image size is 416 x 416 x 3. To extract rich information from a given image, we used several alternating structures of multiple convolutional layers and nonlinear activation layers. We used activation function ReLU with stride size 1 and Maxpool layer with size 2x2. The output is passed to another completely linked layer with 4096 neurons and then to another fully connected layer with 1000 neurons. At the end of the network, SoftMax is used to make the final decision. At the input stage, the ReLU [30] function is applied while the Sigmoid [31] function is used at the output stage. For optimization, we employ the Adam optimizer with an initial learning rate of $\alpha_0 = 1E - 4$ and we use categorical cross-entropy as the loss function which is commonly used in multi-class classification problems. This function deals with problems in which the output of the model belongs to a specific class. To protect our model from overfitting, we drop some values of our neurons and use early stopping which saves us from the long execution time taken by the model for training. For the training of this model, we split the dataset into training, validation, and testing sets with the same ratio as shown above in Table I.

Several methods have been proposed for the detection and recognition of LP. In the proposed method, we extract a frame from a video containing vehicle LP and apply the proposed multi-exposure method to that frame. YOLOv5 is used to detect the LP from frames. Once the LP is detected, we crop the LP part from the images and apply a sliding window approach [32] with our custom CNN model on that cropped image to detect and recognize LP digits. The output of the CNN contains 27 confidence levels, one for each class. After going through the whole frames, we get coordinate of each character with its class label and confidence level. Then, at each character coordinate in LP, the confidence levels of class labels derived from images with different exposure times are compared. Finally, the class label with the highest confidence level is determined as the final predicted character.

In the proposed technique, only the LPs with Intersection over Union (IoU) values greater than 0.69 are detected, and the character class label is selected based on the highest confidence level obtained from the CNN model. IoU is a measure of the overlapping of the annotated and the predicted bounding boxes. We come up with that number after conducting a series of experiments on different LPs. A number of experiments are conducted on the dataset for the optimum value of IoU. If we chose a value less than 0.69, then it detects some objects containing features like LPs.

IV. Result and Discussion

In this section, we conducted experiments to verify the effectiveness of our proposed methodology. The PyTorch framework was used for all the experiments. To calculate the accuracy of detection and recognition we use the following equations:

$$\text{Detection accuracy} = \frac{\text{total number correct detections}}{\text{total number of images}} \tag{4}$$

Recognition accuracy

$$= \frac{\text{total number of correct recognitions of alphanumeric characters}}{\text{total number of alphanumeric characters}} \tag{5}$$

We test the trained modules of detection and recognition of LP using our testing dataset. Results are shown in Table II. A 95.34% detection accuracy is achieved on 1x exposure time images and 95.75% on multi-exposure images. For recognition of each digit and alphabet, augmentation and multi-language repudiate techniques [33] proved helpful. We get the overall recognition accuracy of 68.2% for 1x exposure time images and 75.965% recognition accuracy for different exposure time images. One example demonstrating the effectiveness of using multi-exposure images is shown in Fig. 5. The alphabet G and T are not recognised in the single exposure time image whereas they are recognized by the proposed technique of using multi-exposure images.

The proposed multi-exposure techniques can work with other existing techniques and improve their accuracy. We demonstrate this using three recent techniques [11, 33, 34]. In [11], the authors used YOLOv5 for the detection of LPs and digits, and proposed a custom CNN architecture for characters recognition. The authors of [33] used YOLOv5 with 19 CNN layers for detection of each class on Saudi Arabia LPs. In [34], thresholding and OCR techniques were used to recognize the digits. We applied the proposed multi-thresholding technique to all these three methods [11, 33, 34] and observed significant improvement in their accuracy, as shown in Table II.

TABLE II. Comparison of Detection and Recognition Accuracy with Single and Multi-exposure Images

Methods	**Detection Single Exposure**	**Detection Proposed multi-exposures**	**Recognition Single Exposure**	**Recognition Proposed multi-exposures**
[11]	94.39%	95.15%	66.58%	73.45%
[33]	95.34%	95.75%	66.34%	74.585%
[34]	83.3%	83.49%	69.975%	71.8%
Proposed Method	95.34%	95.75%	68.2%	75.965%

(a) Single exposure time (b) Multi-exposure times

Fig. 5. Detection and Recognition Result on Single and Multi-exposure Images.

V. Conclusion

In this paper, we proposed a method utilizing multi-exposure images to improve both detection and recognition accuracies in ALPR. The proposed approach aims to present an efficient system for automatically detecting and recognizing license plates (LP) in unconstrained real-time settings with the help of different exposure time images. The proposed ALPR system is trained and evaluated on a custom dataset containing 5000 images. Experimental results show that the accuracies of the existing ALPR methods can be improved by using multi-exposure images. In this work, the multi-exposure images are generated artificially. Therefore, for the pixels having very dark or bright shades, the synthesised values may not be very accurate, however, still a significant improvement in accuracy of recognition is observed. In future work, we plan to further improve the accuracy of ALPR by using real images captured at multiple exposure times.

Acknowledgment

The authors extend their appreciation to the Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia for funding this research work through the project number MoE-IF-G-20-12.

References

[1] 1. R. Laroca, E. Severo, L. A. Za nlorensi, L. S. Oliveira, G. R. Gonçalves, W. R. Schwartz, and D. Menotti, "A robust real-time automatic license plate recognition based on the YOLO detector," International Joint Conference on Neural Networks, pp. 1–10, 2018.

[2] S. Ozbay and E. Ercelebi, "Automatic vehicle identification by plate recognition," World Academy of Science, Engineering and Technology, vol. 9, pp. 222–225, 2005.

[3] A. Roy and D. P. Ghoshal, "Number Plate Recognition for use in different countries using an improved segmentation," 2nd National Conference on Emerging Trends and Applications in Computer Science, pp. 1–5, 2011.

[4] I. Kilic and G. Aydin, "Turkish vehicle license plate recognition using deep learning," International Conference on Artificial Intelligence and Data Processing (IDAP), pp. 1–5, 2018.

[5] H. Xiang, Y. Zhao, Y. Yuan, G. Zhang, and X. Hu, "Lightweight fully convolutional network for license plate detection," Optik, vol. 178, pp. 1185–1194, 2019.

[6] Z.-X. Chen, C.-Y. Liu, F.-L. Chang, and G.-Y. Wang, "Automatic license-plate location and recognition based on feature salience," IEEE Trans. Veh. Technol., vol. 58, pp. 3781–3785, 2009.

[7] O. Bulan, V. Kozitsky, P. Ramesh, and M. Shreve, "Segmentation-and annotation-free license plate recognition with deep localization and failure identification," IEEE Trans. Intell. Transp. Syst., vol. 18, pp. 2351–2363, 2017.

[8] A. Safaei, H. L. Tang, and S. Sanei, "Real-time search-free multiple license plate recognition via likelihood estimation of saliency," Comput. Electr. Eng., vol. 56, pp. 15–29, 2016.

[9] P. Marzuki, A. Syafeeza, Y. Wong, N. Hamid, A. N. Alisa, and M. Ibrahim, "A design of license plate recognition system using convolutional neural network," Int. J. Electr. Comput. Eng., vol. 9, pp. 2088–8708, 2019.

[10] Y. Wen, Y. Lu, J. Yan, Z. Zhou, K. M. von Deneen, and P. Shi, "An algorithm for license plate recognition applied to intelligent transportation system," IEEE Trans. Intell. Transp. Syst., vol. 12, pp. 830–845, 2011.

[11] R. Brüngel, J. Rückert, and C. M. Friedrich, "DFUC 2020: An Approach on Diabetic Foot Ulcer Detection via YOLOv5."

[12] R. C. Chen, "Automatic License Plate Recognition via sliding-window darknet-YOLO deep learning," Image Vis. Comput., pp. 47–56, 2019.

[13] A. Tourani, A. Shahbahrami, S. Soroori, S. Khazaee, and C. Y. Suen, "A robust deep learning approach for automatic iranian vehicle license plate detection and recognition for surveillance systems," IEEE Access, vol. 8, pp. 201317–201330, 2020.

[14] A. R. Rashtehroudi, A. Shahbahrami, and A. Akoushideh. "Iranian license plate recognition using deep learning," International Conference on Machine Vision and Image Processing (MVIP), pp. 1–5, 2020.

[15] C. Gou, K. Wang, Y. Yao, and Z. Li, "Vehicle license plate recognition based on extremal regions and restricted Boltzmann machines," IEEE Trans. Intell. Transp. Syst., vol. 17, pp. 1096–1107, 2015.

[16] S. Yonetsu, Y. Iwamoto, and Y. W. Chen, "Two-stage YOLOv2 for accurate license-plate detection in complex scenes," IEEE International Conference on Consumer Electronics (ICCE), pp. 1–4, 2019.

[17] V. Jain, Z. Sasindran, A. Rajagopal, S. Biswas, H. S. Bharadwaj, and K. Ramakrishnan, "Deep automatic license plate recognition system," Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing, pp. 1–8, 2016.

[18] Z. Liu, Z. Wang, and Y. Xing, "Wagon number recognition based on the YOLOv3 detector," IEEE 2nd International Conference on Computer and Communication Engineering Technology, pp. 159–163, 2019.

[19] Q. Wang, "License plate recognition via convolutional neural networks," IEEE International Conference on Software Engineering and Service Science (ICSESS), pp. 926–929, 2017.

[20] D. M. Izidio, A. P. Ferreira, H. R. Medeiros, and E. N. D. S. Barros, "An embedded automatic license plate recognition system using deep learning," Des. Autom. Embed. Syst., pp. 1–21, 2019.

[21] K. Khan and M. R. Choi. "Automatic license plate detection and recognition framework to enhance security applications," J. Electron. Imaging, vol. 28, no. 1, pp. 013036, 2019.

[22] X. Zhang, N. Gu, H. Ye, and C. Lin,"Vehicle license plate detection and recognition using deep neural networks and generative adversarial networks," J. Electron. Imaging, vol. 27, no. 4, pp. 043056, 2018.

[23] J. Qiu and H. Xu, “Camera response prediction for various capture settings using the spectral sensitivity and crosstalk model,” Appl. Opt., vol. vol. 55, pp. 6989–6999, 2016.

[24] M. D. Grossber and S. K. Nayar, “What is the Space of Camera Response Functions,” Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. II–602, 2003.

[25] S. -O. Shim, S. Alharbi, I.R. Khan, and W. Aziz, “De-ghosting in High Dynamic Range Imaging Based on Intensity Scaling Cue”, Adv. Electr. Comput. Eng., vol. 20, no. 3, pp. 3–10, 2020.

[26] S. -O. Shim, “Estimation of gamma-corrected exposure time ratio in multi-exposure images for removal of moving objects”, Appl. Opt., vol. 59, no. 13, pp. 4076–4080, 2020.

[27] C. -W. Yu, Y. -L. Chen, K. -F. Lee, C. -H. Chen, and C. -Y. Hsiao, "Efficient Intelligent Automatic Image Annotation Method based on Machine Learning Techniques," IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), pp. 1–2, 2019.

[28] https://github.com/ultralytics/yolov5.

[29] M. A. Rafique, W. Pedrycz, and M. Jeon, "Vehicle license plate detection using region-based convolutional neural networks," Soft Comput., vol. 22, pp. 6429–6440, 2018.

[30] A. F. Agarap, "Deep learning using rectified linear units (relu)." arXiv preprint arXiv:1803.08375, 2018.

[31] D. Costarelli and R. Spigler, "Approximation results for neural network operators activated by sigmoidal functions," Neural Netw., vol. 44, pp. 101–106, 2013.

[32] N. I. Glumov, E. I. Kolomiyetz, and V. V. Sergeyev, "Detection of objects on the image using a sliding window mode," Opt. Laser Technol., vol. 27, no. 4, pp. 241–249, 1995.

[33] S. T. A. Ali, A. H. Usama, I. R. Khan, M. M. Khan and A. Siddiq, "Mobile Registration Number Plate Recognition Using Artificial Intelligence," IEEE International Conference on Image Processing (ICIP), pp. 944–948, 2021.

[34] http://34.https//platerecognizer.com.

An Urban Water Infrastructure Management System Design with Storm Water Intervention for Smart Cities

Edward B. Panganiban[1], Rafael J. Padre[2], Melanie A. Baguio[3], Oliver B. Francisco[4], Orlando F. Balderama[5]
Isabela State University, Echague, Isabela, Philippines[1, 2, 3, 5]
Cauayan City Local Government, Cauayan City, Isabela, Philippines[4]

***Abstract*—Cauayan City is one of the hubs of economic development and activities in the northern part of the Philippines. Since this is an urban area, there is a tendency for people and businesses to converge, which results in higher water demand. At present, the combined distribution efficiency of its water infrastructures under the management and supervision of the Cauayan City Water District (CCWD) is only 87% or with a combined distribution loss of 10%, which is 1282.50m3 losses per day. This study suggests the necessity to introduce new and innovative water management technologies and systems adapted to climate change to address the city's needs. Problems that need to be addressed include a low-efficiency performance of the existing water infrastructure systems, lack of management tools for more efficient delivery of water services, limited service coverage of the water district due to limited water resources, and depletion and contamination of aquifers and other water sources since shallow aquifer is mainly utilized. Hence, a decision-support application based on geographic information systems (GIS) for managing urban water infrastructure with Storm Water intervention is a designed solution to address the needs of the city. The combination of decision support systems (DSS) and geographic information systems (GIS) was presented in this paper to maximize and properly utilize water infrastructure. One of the tools used as DSS is MIKE Operation. This is a complete GIS-integrated modeling tool that enables users to make sound decisions and create future concepts for urban storm water systems - cost-effective and resilient to change. A conceptual framework and relevant methodologies were presented as a guide for the success of the designed new technology.**

Keywords—Water infrastructure; geographic information systems; storm water; decision support systems; water management

I. Introduction

Cauayan City in Isabela is one of the country's first smart cities and one of the Philippines' premier agro-industrial hubs [1]. This advancement in science and technology is being acknowledged by the Department of Science and Technology (DOST). Since urban areas, such as Cauayan City, are the hub of economic development and activity, there is a tendency for people and businesses to converge, which results in higher water demand.

At present, Cauayan has a total of fifteen (15) water pumping stations and eight (8) elevated water tanks serving only 12,067 households (29%) out of 30,207 households (71%). The combined distribution efficiency of these water infrastructures under the management and supervision of the Cauayan City Water District (CCWD) is only 87% which is accounted for the 9,229.63 cubic meters of metered water out of the distributed water of 10,518.13 cubic meters daily or with a combined distribution loss of 10% which is 1282.50 cubic meter losses per day [2]. This study suggests the necessity to introduce new and innovative water management technologies and systems adapted to climate change to address the need of the city like improving the performance of the existing water infrastructure systems, lack of management tools for more efficient delivery of water services, limited service coverage of the water district due to limited water resource and depletion and contamination of aquifers and other water sources since shallow aquifer is mainly utilized. Hence, a GIS-based decision support tool in managing urban water infrastructure with the Storm water invention is a designed solution to address the need of the city.

To maximize and properly utilize water infrastructure, the merging of decision support systems (DSS) and geographic information systems (GIS) is required to provide a visualization tool and the necessary spatial database for turning a basic geographical query into a sophisticated spatially and analytical distributed modeling tool [3], [4]. Reitsma [5] defines a DSS for water resource applications as computer-based systems which integrate state information, dynamic or process information, and plan evaluation tools into a single software implementation. One of the tools used as DSS is called MIKE Operation. It is a complete GIS-integrated modeling tool that enables users to make sound decisions and create future concepts for urban storm water systems - cost-effective and resilient to change [6], [7].

The general goal of this paper is to establish a management tool for urban water infrastructure to help decision-makers, policymakers, and managers in Cauayan City, Isabela. Its specific goals are to assess and analyze the risk of urban water infrastructure systems in Cauayan City, establish a decision support system for optimal water infrastructure management, and design, build and test an urban Storm water storage system integrated with monitoring sensors.

II. Literature Review

A. Rain/Storm Water Harvesting

In the municipality of Erisos, on the island of Kefalonia in northern Greece, rainwater is collected in twenty-three ferroconcrete rainwater storage tanks. This was developed in the 1970s and positioned at the bottom of cement-paved hill slopes acting as catchment areas ranging from 600 cubic meters to 3,000 cubic meters [8].

This Storm water harvesting infrastructure system was constructed at Brimbank Park, Melbourne, Australia, to irrigate 1.47 hectares of park public open space, utilizing a 1.7 hectares carpark as the principal storm water catchment. The system design featured a 2,000 cubic meter storm water storage tank, a 450 mm storm water collecting pipe, a 600 m^2 bioretention/filtration system, a 150 mm pressure main with a 35 kW pump, and a sprinkler system with a 125 mm manifold and 50 mm laterals [9].

Rainwater harvesting [10] supplies water for drinking, sanitation, and increasing the productivity of agro-ecosystems in Kenya's semi-arid savannah towns of Kaijado and Lare. Roof water harvesting for residential reasons (drinking and sanitation), runoff collecting in ponds for small gardens, trenches for groundwater recharge, and afforestation were among the technologies implemented [11].

The rainwater harvesting system built at Sapilang, Bacnotan, La Union, which consists of roofs, gutters, downspouts, a filter, and a storage tank, can collect rainfall to feed and maintain the drinking water requirements of 8-12 family members [12].

B. Storm Water Infrastructure

To augment the existing water supply, G.C. Dandy et al. (2019) emphasized the significance of having alternate water sources in many cities, such as storm water collection and processed wastewater. Harvesting schemes for Storm water that complement current water systems, according to them, are complicated because they are made up of integrated components set that execute the duties of collecting, storing, treating, distributing, and discharging. Furthermore, while selecting and evaluating Storm water harvesting solutions, factors must be considered. These are environmental, economic, environmental, technical viability, and social variables [13].

According to the International Water Association, IWA (2019), population increase and urbanization are adding strain on urban water supplies. According to UN predictions, by 2025, fifty percent of the world's demographic would be residing in water-stressed regions. Furthermore, the current worldwide health catastrophe caused by COVID-19 emphasizes the availability of clean water to ensure a healthy population [14].

K.A. Feehan et al. Rainwater collecting has been performed for generations and was formerly the principal technique of acquiring water for home usage, according to (2012). Increasing water needs, water consumption limitations, new storm water management rules, and the expansion of low-impact development (LID) and "green construction" methods have rekindled interest in rainwater collection [15].

C. SWH Management

Stormwater harvesting (SWH) is one alternative water resource that might replace typical urban water supply while also providing a variety of social and environmental advantages. This addresses water scarcity by increasing resilience to the effects of climate change [16].

Urban planning for Water Sensitive Urban Design (WSUD) has been inconsistent, lacking direction and yielding sub-optimal results. Green storm water management infrastructure is being used more widely throughout the world to mitigate the negative effects of climate change and urbanization while also delivering several ecosystem services. WSUD is also used in Australia, Low Impact Development (LID) in the United States, Sponge City systems in China, and Nature-Based Solutions (NBS) in Europe [17].

D. GIS-Based Decision Support Tool for SWH

In the research of Shao, H. et al. (2017) said that an open-source geographic information system (GIS) based decision support system (DSS) for constructing modeling economic costs, water quality/quantity advantages, and BMP scenarios at the farm field, watershed scales and subbasin can be established. This DSS included a Soil and Water Assessment Tool (SWAT), a farm economic model, and an optimization model in an open-source GIS program called Whitebox Geospatial Analysis Tools, The DSS was applied to 14.3 square kilometers of Gully Creek watershed in southern Ontario, Canada, a coastal watershed that flows directly into Lake Huron [18].

Rainwater harvesting systems (RWHSs) have been recognized as a simple and effective technique to alleviating the deterioration of urban water stress. A recent study addressed the temporal-spatial complex of rainfall and proposed a GIS-simulation-based decision support system (DSS). Sensitivity study confirmed that, as compared to the previous generalized technique, this DSS increased the information value. As a result, the DSS, which provides more holistic and complete support, has been highlighted as a valuable instrument for mitigating the problem of urban water stress. This is a comprehensive method for promoting domestic RWHSs in a large-scale water-saving scheme that incorporates water consumption reducing equipment (WCRE) and gray water reuse (GWR) [19].

B. Mbilinyi and colleagues introduced a decision support system (DSS) based on a geographic information system (GIS) that employs remote sensing (RS) and a short field survey to identify possible areas for RWH technologies. The DSS receives soil texture, maps of rainfall, soil depth, slope, drainage, and land use as inputs. This produces maps of suitable locations for bench terraces, water storage systems, borders, and stone terraces as outputs. The findings of their study, which included testing and validation of the generated DSS, suggested that the tool may be used to anticipate probable locations for RWH technologies in semi-arid environments with high accuracy. The majority of anticipated RWH innovations were identified through testing in extremely high and highly appropriate places (41.4 percent and 40 percent, respectively). During the assessment, 36.9 percent of RWH techniques were located in moderately appropriate sites, with 23.6 percent discovered in very highly suitable and extremely suitable places [20].

MIKE OPERATION consists of MIKE 11 and MIKE Basin modeling tool, which has GIS features, and an easy-to-use online modeling framework for water prediction and operational control. It is intended for model-based forecasting services as well as online operational management of river systems, water collecting systems, and water transportation systems. This is also a simple real-time system for real-time modeling, optimizes water distribution collection systems, and controls the groundwater (CTNN Report, 2017) [21].

The MIKE 11 program was used to assess the water resources in the catchment of the Strymonas River. It also simulated unstable water movement in surface water bodies and catchment runoff [22].

MIKE basin software is used in the Pinho river's hydrographic basin. This is helped by a geographic information system to represent the basin both spatially and temporally. Thus, allowing for easier visualization and interpretation of findings. According to the acquired data, it was confirmed that the hydrographical basin now satisfies the considered water demands. The DSS has proven to be a valuable instrument in assisting decision-making in the examined river basin [23].

III. Materials and Methods

A. Conceptual Framework

Fig. 1 shows the conceptual framework of this paper. For the inventory of existing water infrastructure, the list of existing water infrastructure will be obtained from the City Planning and Development Council (CPDC) and from the Cauayan City Water District (CCWD). These infrastructures will be validated using a handheld Geographic Positioning System (GPS) and will be assessed. In Data Processing, Design, Construction, and Evaluation of the Storm water Storage System and Water Distribution Control System, all the data collected will be processed using MIKE Operations with GIS software.

Risk analysis of urban water infrastructure systems in Cauayan City will be done using Fuzzy Fault Tree Analysis (FFTA). Data about natural hazards, human-made hazards, and operational hazards will be obtained from internet-based information and survey form. At the same time, the development of urban digital model/system for water infrastructure involves the development of an urban digital model/system for water infrastructure will be implemented based on two steps (Fig. 2), such as the establishment of the urban digital model and utilizing a sensing layer for data collection. The Geographic Information System will play a role in these steps.

B. Construction of the Urban Digital Model

The first stage is to create an urban digital model that emphasizes the contents of the urban constructed and natural surroundings. The digital model gives geolocation and characteristics for each urban component (attributes). GIS will be used to produce a digital model of urban 'horizontal components' including the natural environment, urban networks, and transit facilities while establishing information modeling will be used to represent vertical components like structures. The combination of BIM and GIS results in a powerful tool for constructing an urban digital model utilizing georeferenced data and presenting this data in an intuitive manner. The combination of which will be integrated with MIKE Operations to come up with real-time simulation-based optimization. MIKE Operations will be used to optimize water distribution and collection systems. It will also be used to establish real-time data management and forecasting and monitor leaking water in the distribution system, minimizing the losses within the water infrastructure.

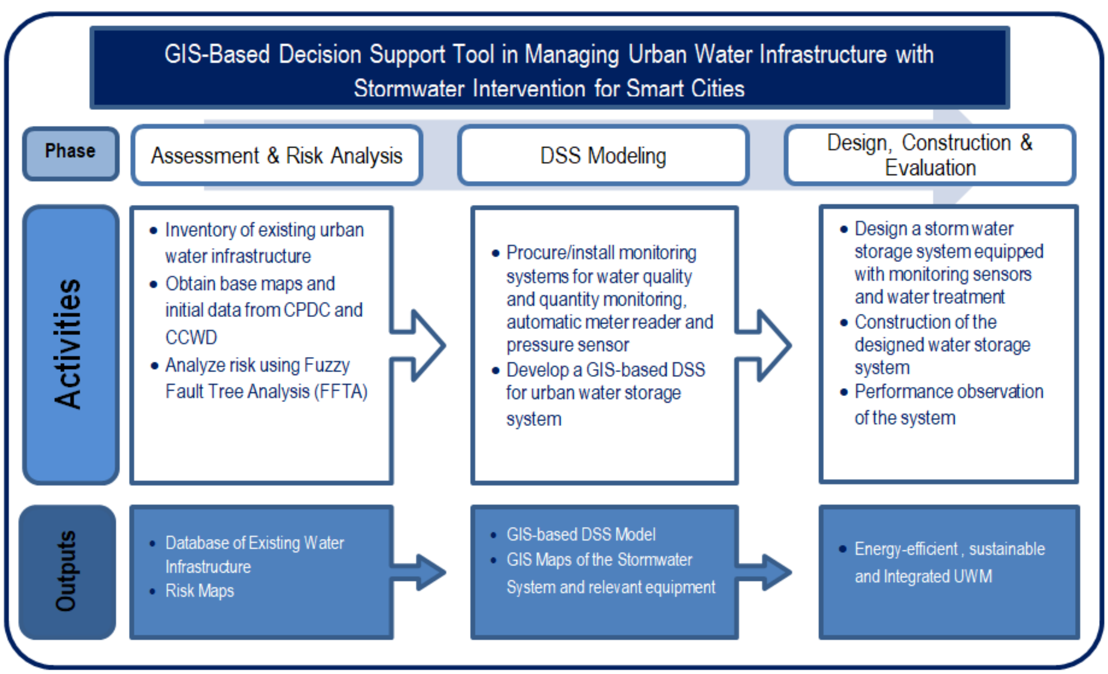


Fig. 1. Conceptual Framework.

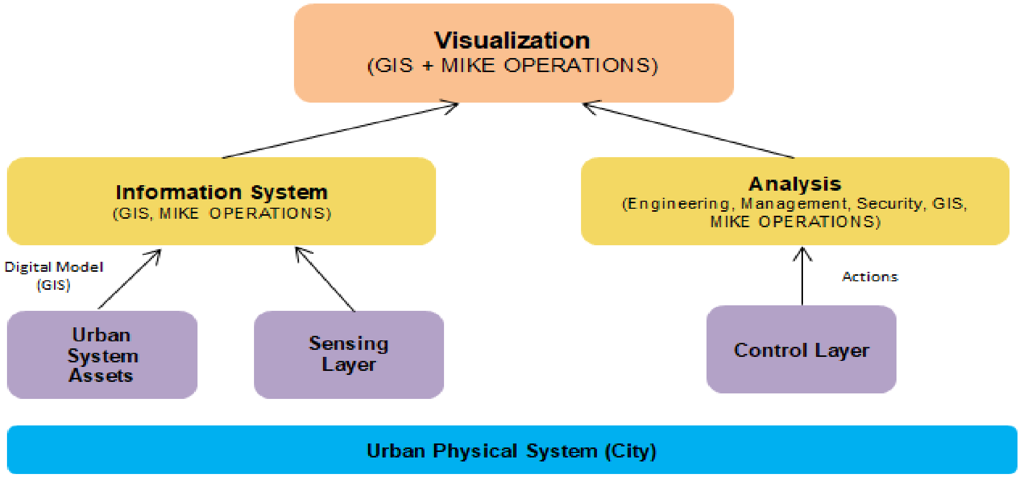


Fig. 2. Steps for the Implementation of Smart City Programs.

C. Sensing Layer

The second phase is to build the sensing layer, which will convey urban operational data to the smart city information system. This layer's sensors will be employed to monitor urban networks and infrastructure. Data can also be complemented with audio recordings, videos, and images, resulting in urban big data. Some of the sensors that will be utilized to monitor energy and water utilities are seen in Fig. 3. Water usage will be tracked using automatic meter readers (AMRs), water pressure will be tracked using pressure sensors, and water quality will be tracked using water quality devices (conductivity, turbidity, chlorine, pH). In the drainage system, sensors will be employed to monitor water level, flow, and quality (pH, temperature, turbidity, etc.). It enables the early identification of floods. Sensors will be installed in the district heating system to record fluid temperature, pressure, flow, and the status of the valve. It enables early defect identification and system performance enhancement. GIS + Mike Operations allows you to view the monitoring system and the characteristics and state of the sensors. It also allows you to see real-time and historical data on GIS maps.

1) Input Data: Since the rainfall and design method are site-specific, the rainfall data should be gathered for the design site. The capacity of nearby sewer systems should be analyzed.

2) Design Data: The tank will be designed according to the choice of controllable measures. The pumping rate and operation rules will be specified. The infiltration capacity will be determined based on site conditions and soil characteristics.

D. Design and Construction of Intervention Storm Water Infrastructure Systems

The storm water intervention facility has a total catchment area of three hectares with a storage capacity of 1750 cubic meters. Fig. 4 depicts a schematic diagram of the planned storm water infrastructure system, including the various components (examples include a storage tank, a flushing treatment device, groundwater recharge tank, UV treatment, elevated tank, suction pumps, valves, piping, and a water quality monitoring system). The recommendation to consider runoff water from the surface/roads aside from rainwater and Storm water falling from the roof is not feasible because the design storage capacity of the storm water is limited only to one (1) hour rainfall duration, which is enough to fill the storage tank. On the other hand, elevating the water storage tank in the new design and system is much appreciated. However, the said recommendation is not feasible considering that the building structures in the proposed site were not uniform regarding the number of the storey (i.e., BJMP and PNP buildings are four storey buildings while others are two-storey and one-storey buildings). Elevating the storage tank with this physical condition still requires motor pumps and heavy foundation structures. In addition, the space of the proposed site had limited space. Hence, when covered, the space occupied by the proposed subsurface storage tank will be useful for other purposes. Nevertheless, the recommendations will still be considered a possible option when scaled up to other areas.

E. Water Distribution Control System

The water distribution control system illustrates in Fig. 5. The monitoring and control system will measure values for pipe steel analysis through online and offline systems. The pipe steel analysis can distribute pressure control to the appropriate distribution amount and set values for the monitoring and control system. The demand assignment, parameter setting, and database update are also under pipe steel analysis.

F. Statistical Data Analysis

The statistical analysis to be employed is the following Rainfall Frequency Analysis for rainfall data analysis, Regression analysis for the performance of the storm water infrastructure. Secondary statistical data on the projection of population, water demand, and a groundwater abstraction rate from 2020 to 2050 was obtained from the study of Alvarez et al. (2020) on Benchmarking of the water supply and wastewater management system of a Smart City: the case of Cauayan City.

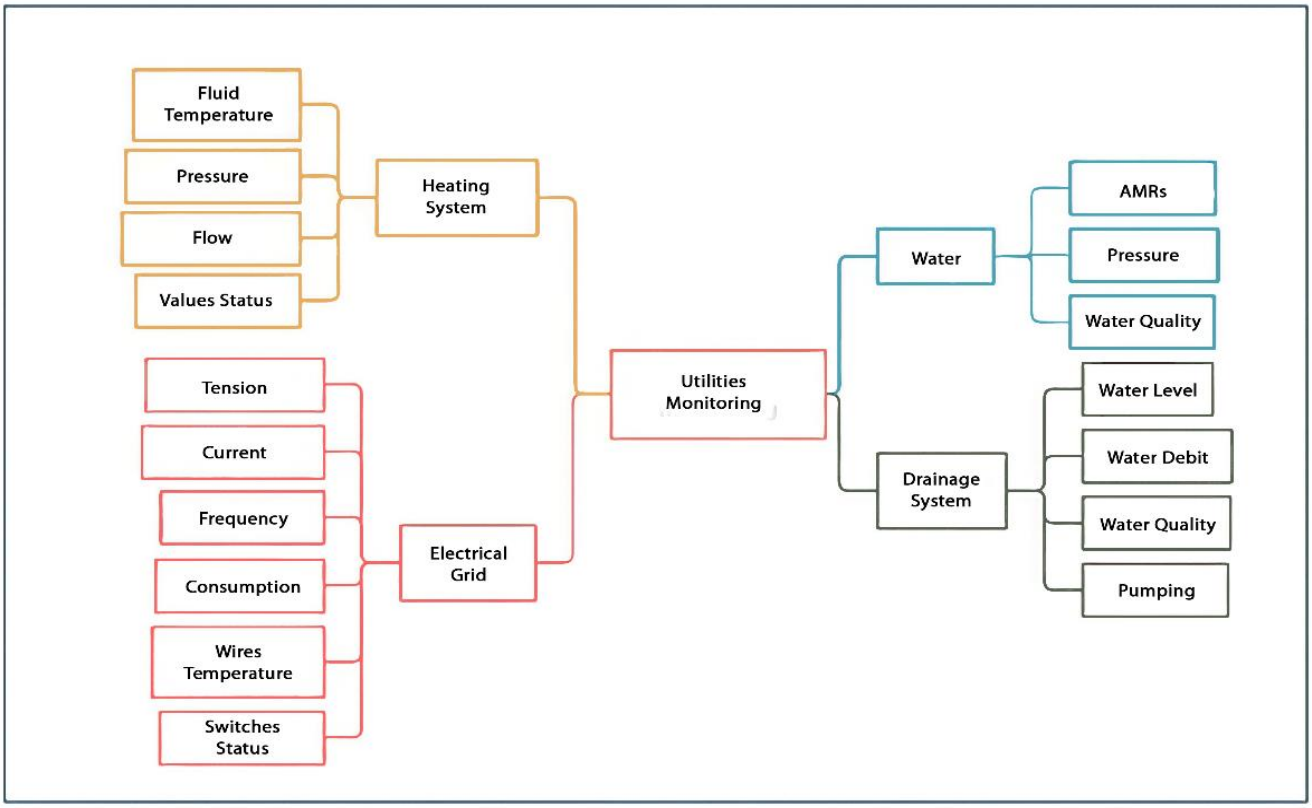


Fig. 3. Sensors in Water Monitoring.

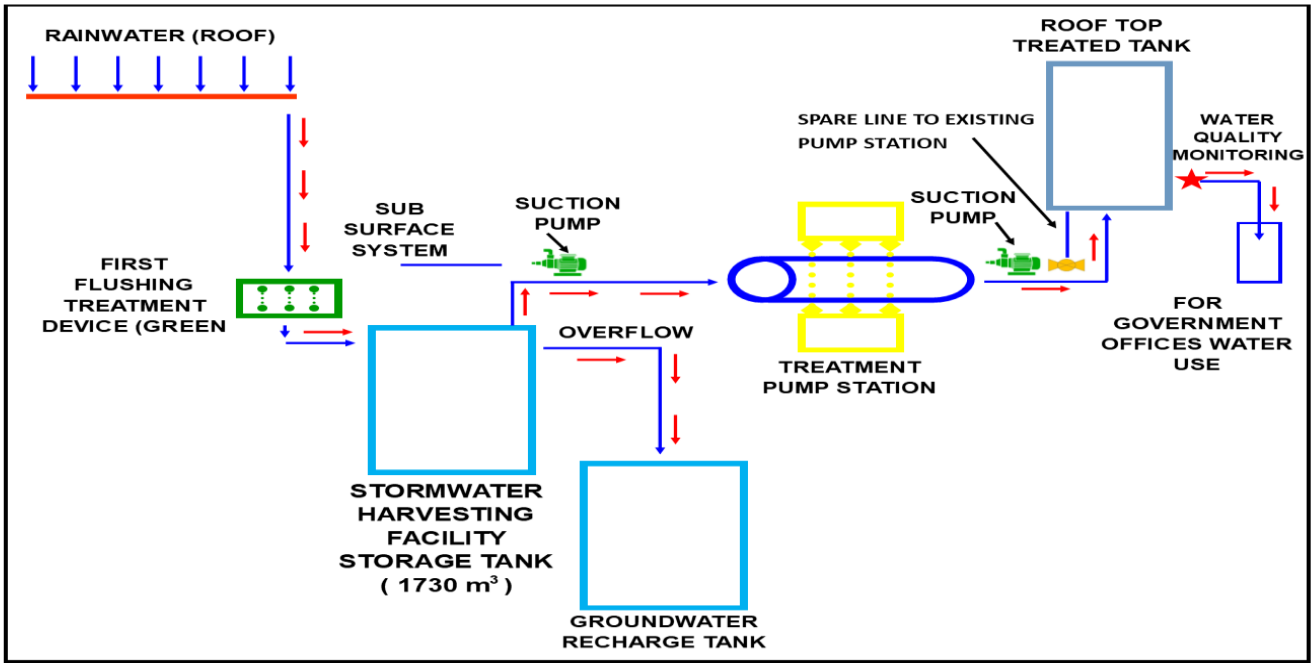


Fig. 4. Schematic Diagram of Subsurface Storm-Water Storage System.

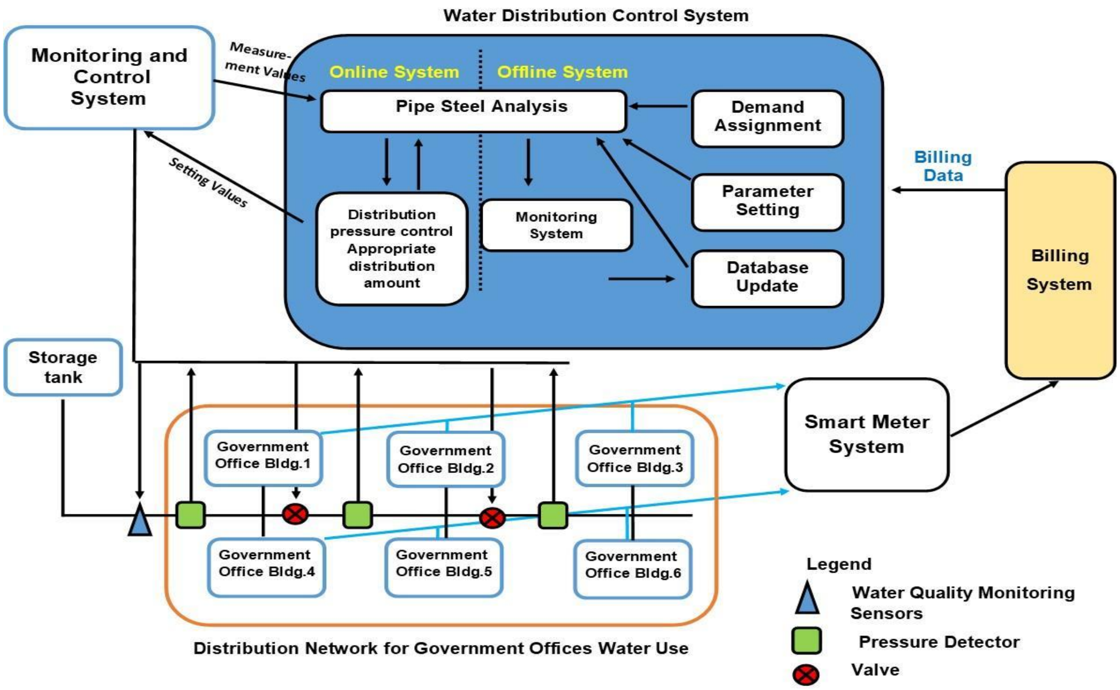


Fig. 5. Water Distribution Control System.

IV. Result and Discussion

A. Differentiation (GIS-Based Decision support Tool in Managing Water Infrastructure)

The authors compared the designed methods versus the existing practices locally and abroad regarding the various GIS-based decision support tools in Managing water Infrastructure. This is presented and summarized in Table I.

B. Differentiation (Storm Water Harvesting Infrastructure)

In terms of comparison between the designed methods over existing Storm water Harvesting Infrastructure, Table II presents the difference in various parameters indicated.

C. Risk Management Plan

Table III and Table IV show the risk management solution or a proposed contingency plan in case there will be challenges that will happen during the duration of the designed method.

TABLE I. Differentiation (GIS based Decision Support Tool in Managing water Infrastructure)

Parameters	**Designed Method**	**Existing Practices (Local)**	**Abroad**
Management Practices	GIS-based decision support tool in managing urban water infrastructure with storm water intervention.	Conventional - Non ICT tool management without water intervention (CCWD) as a practice	GIS-Based Management Tool
Performance Distribution Efficiency	Low distribution losses (2%)	Distribution Losses is 13 %	Distribution Losses 3-4%
Equipment tools	- Instruments with digital output (meters and sensors), such as water quality monitoring, flow meters, rain gauges, and other environmental parameters. - Real-time acoustic leak detecting devices - video camera-based asset management - Smart water measuring devices to monitor its use - Pressure monitoring to identify leaks and optimize pumps, UV treatment - GIS software	Manual monitoring pressures, flow meters, and leakages	- Flowmeters, rain gauges, monitoring of water quality as well as other environmental parameters. - Acoustic leak detecting devices in real-time - Asset management using video devices - Smart water systems to monitor their use - Pressure monitoring to detect leaks and optimize pumps - GIS/SCADA software

TABLE II. DIFFERENTIATION (STORM WATER HARVESTING INFRASTRUCTURE)

Parameters	Designed Method	Existing Practices (ISWM-Cebu))	Abroad
Design function of storm water infrastructure	Storm water collection and groundwater recharge facility equipped with water monitoring (quality) and treatment system. For flooding control, groundwater recharge and water supply for government offices will be implemented.	Rain and storm water collection with wastewater treatment Water supply for cleaning and gardening	SWH was designed to harvest rainwater with an infiltration facility and redirect rainwater to sewer systems. Other designs were used for irrigation and toilet flushing.
Design Capacity	-1750 cubic meter + groundwater reservoir	- 60 cubic meter	4 -40 cubic meter
Efficiency	90%	75%	30% to 90%
Cost of Investment	Php 3Million Or Php 1714.30/ cubic meter	Php 2Million Or 33,330.33 per cubic meter	Php 30,000,00/cubic meter
Water Dependence	-Reduced water dependence to 90%	-Reduced water dependence to 75% during month	
Component	-Primary storm water treatment facility (Green filter) - Storage Tank - pipe and fittings *Groundwater recharged Tank *Elevated water Tank *UV treatment *Water quality assessment facility	-Primary storm water treatment facility (Green filter) - Filtration system based on the micro membrane (MF System) - Rainwater Remote Administration System Integrated - Rainwater Drainage in Case of Emergencies - Water Treatment System for Reuse	-System of its collection -An inlet filter -Diverter for the first flush -Tank for storing items -Overflow -Treatment methods -Pump -Flow metering -An indicator of water level

TABLE III. RISK MANAGEMENT

Cause of delay	Solution
The foreseen major risk of the paper may be caused by a change of Leadership in the LGU and the procurement of needed materials.	There must be a signed Memorandum of Agreement (MOA); involve the university finance department during the inception activity, and perform intensive monitoring during the procurement process.

TABLE IV. OTHER POSSIBLE RISKS

Item	Risk	Designed Contingency Plans
Conduct Training and meetings	Restriction in the conduct of trainings and meetings	Activities will be undertaken using a combination of a face-to-face meeting in small groups and video conference with those affected by travel restrictions. A fully-equipped virtual conference venue will be set up at ISU.
Acquisition of software/ and equipment	Difficulty in the purchase and delivery of equipment due to travel restrictions	Sets of software will be acquired online; All equipment for purchase is locally available; hence delivery mode would be made by couriers or using ISU official vehicles under special permits
Data gathering	Restriction to travel due to community quarantine protocols	Since travels are in the provincial/ regional jurisdictions, special permits will be secured, and gathering in the field will be made using local assets.
Coordination with stakeholders	Travel difficulty for a face-to-face coordination meeting	Coordination will be made by online calls and videoconferencing.
Reports presentation and submission to PCIEERD	Travel restrictions to Manila for meetings and submission with the presentation of reports	It will be done by video conferencing and online mode.

V. CONCLUSION AND RECOMMENDATIONS

The authors determined the potential outcomes, socio-economic impact, and environmental impact through various analyses of the designed methods and materials identified. It improved the efficiency and quality of urban water services of Cauayan City and developed an integrated, energy-efficient, and sustainable urban storm water management. The potential socio-economic impacts of this paper are; at least 90% of distribution losses will be reduced (from 13% to 2% distribution losses), at least 10% of the energy used from water supplies of government offices situated in the area will be reduced, reduces the cost to build storm drain Infrastructure and at least 10% increase of income by laundry businesses. In terms of environmental impact, the construction of the GIS-based decision support tool in managing urban water infrastructure includes reduced standing water and localized flooding, reduced storm water runoff into the combined sewer system, promote infiltration and groundwater recharge, removal or reduction of contamination of hazardous elements to the surface water or groundwater resources, protection or rehabilitation of natural and artificial waterways and conserved water by increasing distribution efficiency.

For its sustainability plan and scaling-up, pilot intervention and its feasibility and acceptability are demonstrated in this pilot scale, the steps that will be undertaken for scaling-up are a policy recommendation for adoption. This is through city legislation and a roll-out of technology in clustered urban watersheds will be conducted. The technology set-up will also be incorporated in the designed Cauayan City Drainage Master Plan study in 2022 since LGU-Cauayan City and Cauayan City Water District (CCWD) are committed to accept and adopt the paper and they are willing to scale up to other areas within the city.

Finally, in terms of the coverage of the paper, the urban water infrastructure with the storm water infrastructure will be piloted at Cauayan City, Isabela. The storm water infrastructure has a catchment area of 3 hectares only. The design specification of the said storm water infrastructure was based on hydrologic data obtained from the site (i.e., from a roof to a storm water harvester).

ACKNOWLEDGMENT

We want to extend our sincerest gratitude to the Department of Science and Technology – Philippine Council for Industry, Energy, and Emerging Technology Research and Development (DOST-PCIEERD) for funding this paper. Likewise to the Local Government Unit of Cauayan City and Cauayan City Water District for being the cooperating agencies of this paper. Finally, Isabela State University as the implementing agency.

REFERENCES

[1] Visaya, "Drones deployed as Cauayan City reports 1st Covid-19," 2020. [Online]. Available: https://www.pna.gov.ph/articles/1099744#:~:text =Cauayan City%2C as one of,to the people more efficiently.

[2] M. Teresa, S. Alvarez, A. M. Ramos, E. U. Cammayo, J. Lloyd, and R. Bareng, "Benchmarking of the Water Supply and Wastewater Management System of a Smart City: A Case Study of Cauayan City, Philippines," TEST Eng. Manag., vol. 83, no. July, 2020.

[3] R. Buffat, L. Schmid, N. Heeren, A. Froemelt, M. Raubal, and S. Hellweg, "GIS-based Decision Support System for Building Retrofit," Energy Procedia, vol. 122, pp. 403–408, 2017, doi: 10.1016/j.egypro.2017.07.433.

[4] R. E. Mohamed, E. Kosba, K. Mahar, and S. Mesbah, "A Framework for Emergency-Evacuation Planning Using GIS and DSS," Lect. Notes Geoinf. Cartogr. B. Ser., no. September 2021, pp. 213–226, 2021, doi: 10.1007/978-3-319-59539-9_16.

[5] R. F. Reitsma, "Structure and support of water-resources management and decision-making," J. Hydrol., vol. 177, no. 3–4, pp. 253–268, 1996, doi: 10.1016/0022-1694(95)02962-1.

[6] Mike Powered by DHI, "Storm water and wastewater," https://www.mikepoweredbydhi.com/products/mike-urban/collection-systems, 2022.

[7] R. E. Mohamed, E. Kosba, K. Mahar, and S. Mesbah, "A Framework for Emergency-Evacuation Planning Using GIS and DSS," pp. 213–226, 2018, doi: 10.1007/978-3-319-59539-9_16.

[8] S. M. Hamdan, "A literature based study of stormwater harvesting as a new water resource," Water Sci. Technol., vol. 60, no. 5, pp. 1327–1339, 2009, doi: 10.2166/wst.2009.396.

[9] S. Park and S. Passmore, "Stormwater Management Report," 2020.

[10] M. Semaan, S. D. Day, M. Garvin, N. Ramakrishnan, and A. Pearce, "Optimal sizing of rainwater harvesting systems for domestic water usages: A systematic literature review," Resour. Conserv. Recycl. X, vol. 6, no. February, p. 100033, 2020, doi: 10.1016/j.rcrx.2020.100033.

[11] N. Aroka, "Rainwater Harvesting in Rural Kenya: Reliability in a Variable and Changing Climate," 2010.

[12] A. T. Esguerra, A. E. Madrid, and R. G. Nillo, "Rainwater harvesting, quality assessment and utilization in region I," E-International Sci. Res. J., vol. 8, no. 2, pp. 145–155, 2011, doi: 10.2166/wh.2009.037.

[13] M. K. Wittholz, B. K. O'Neill, C. B. Colby, and D. Lewis, "Estimating the cost of desalination plants using a cost database," Desalination, vol. 229, no. 1–3, pp. 10–20, 2008, doi: 10.1016/j.desal.2007.07.023.

[14] S. of Green, "Groundwater-based water supply: How groundwater can aid in securing water supply amidst population increases and climate change," White Pap. Green Transit., 2021.

[15] K. A. Feehan et al., "Stormwater Management : Rainwater Harvesting in Residential-Scale Landscapes," Water Manag. – Water Qual. Lawn Gard. – Water Manag., vol. G2148, no. May, 2012, [Online]. Available: http://extensionpublications.unl.edu/assets/html/g2148/build/g2148.htm.

[16] L. Fisher-Jeffer, K. Carden, N. P. Armitage, and K. Winter, "Stormwater harvesting: Improving water security in South Africa's urban areas," Res. Lett., vol. 113, no. 1, pp. 2–5, 2017.

[17] M. Maurer, "Department Urban Water Management," https://www.eawag.ch/en/department/sww/, 2022.

[18] H. Shao, W. Yang, J. Lindsay, Y. Liu, Z. Yu, and A. Oginskyy, "An Open Source GIS-Based Decision Support System for Watershed Evaluation of Best Management Practices," J. Am. Water Resour. Assoc., vol. 53, no. 3, pp. 521–531, 2017, doi: 10.1111/1752-1688.12521.

[19] Y. R. Chiu, K. Aghaloo, and B. Mohammadi, "Incorporating rainwater harvesting systems in iran's potable water-saving scheme by using a GIS-simulation based decision support system," Water (Switzerland), vol. 12, no. 752, 2020, doi: 10.3390/w12030752.

[20] B. P. Mbilinyi, S. D. Tumbo, H. F. Mahoo, and F. O. Mkiramwinyi, "GIS-based decision support system for identifying potential sites for rainwater harvesting," Phys. Chem. Earth, vol. 32, no. 15–18, pp. 1074–1081, 2007, doi: 10.1016/j.pce.2007.07.014.

[21] DHI, "Mike operations An online modelling framework designed for water forecasting and operational control," https://www.mikepower edbydhi.com/products/mike-operations, 2017.

[22] C. Doulgeris, P. Georgiou, D. Papadimos, and D. Papamichail, "Ecosystem approach to water resources management using the MIKE 11 modeling system in the Strymonas River and Lake Kerkini," J. Environ. Manage., vol. 94, no. 1, pp. 132–143, 2012, doi: 10.1016/j.jenvman.2011.06.023.

[23] L. F. S. Fernandes, M. J. Marques, P. C. Oliveira, and J. P. Moura, "Decision support systems in water resources in the demarcated region of Douro - case study in Pinhão river basin, Portugal," Water Environ. J., vol. 28, no. 3, pp. 350–357, 2014, doi: 10.1111/wej.12042.

Sentiment Analysis on Amazon Product Reviews using the Recurrent Neural Network (RNN)

Roobaea Alroobaea
Department of Computer Science
College of Computers and Information Technology
Taif University, P. O. Box 11099, Taif 21944, Saudi Arabia

***Abstract*—In this paper, the problem of sentiment analysis on Amazon products is tackled. In fact, sentiment analysis systems are applied in all business and social fields. This is because the opinions are at the center of all human activities, and they are key influencers of our behaviors. In this study, the recurrent neural network (RNN) model is used to classify the reviews. Three Amazon review datasets were applied to predict the sentiments of the authors. In conclusion, our results are comparable to the best state of the art models with an accuracy of 85%, 70% and 70% for three datasets.**

Keywords—Sentiment analysis; natural language processing; deep learning; RNN

I. Introduction

Sentiment analysis is the area of study that investigates people's opinions and feelings, attitudes, and emotions. It is one of the most active research fields in natural language processing. It is widely studied in data mining filed. Therefore, this research has been extended outside science to cover management and social sciences. The increasing importance of the sentiment analysis overlaps with the evolution of social media, such as chat rooms, blogs, micro-blogs, Twitter [1]. For the first time in human history, there are huge amount of opinions in a digital form for analysis. The sentiment analysis systems are applied in all the business and social fields because of that the opinions are at the center of all human activities. Thus, they are significant influencers of our behaviors. Our beliefs and perceptions of reality as well as the selections we do are conditioned by how others see and value the world? For this reason, when a decision should be made, the opinions of others are often sought. This is true not only for individuals, but also for organizations. On Amazon, Internet users give their opinions on products. However, these reviews vary from a product to another and are used to improve and alert companies that have negative reviews of their products [1]. The aim of this paper is to use the recurrent neural network (RNN) model to classify the reviews of three Amazon review datasets[1] to predict the sentiments of the authors. The rest of the paper is organized as follows. Section two presents a definition of the sentiment analysis. Section three deals with the state-of-the-art. Section four introduces the methodology. Section five gives the results and experiments performed on the Amazon datasets. Finally, the last section concludes the paper.

[1] https://data.world/datafiniti/consumer-reviews-of-amazon-products

II. Contributions

The contributions in this study can be summarized as; 1) Arabic data set are used in this research because few research have been done in this area [12]. 2) We present a method using RNN to predict sentiment of authors. 3) We compare our proposed model to Long short-term memory (LSTM), Gated Recurrent Unit (GRU), and Convolutional Neural Network (CNN).

III. Literature Review

A. Sentiment Analysis

Sentiment analysis, also called the opinion mining, is a sub-field of computer science besides. It is considered as a part of an automatic natural language processing and aims to classify feelings expressed in texts. There are other related names slightly referring to different tasks, such as opinion mining, sentiment mining, subjectivity analysis, effect analysis, and emotion analysis [2].

Furthermore, sentiment analysis is made using one of two fundamental methods, which are rule-based classifiers and machine learning classifiers. The former are rules derived from the linguistic study of a language that are utilized to sentiment analysis. The latter is statistical machine learning algorithms that are used to automatically learn sentiment signals [1][2]. Therefore, there are two terms that should be defined which are sentiment and opinion. In the literature review, there is confusion between these two words. In the Oxford Dictionary, feeling is defined as a point of view or opinion that is held or expressed as an emotion. For the word opinion, it is a belief, or a judgment formed about something, which is not necessarily based on facts or knowledge but on the beliefs or opinions of a group or of the majority of people. Consequently, it can be said that the term of feeling is a person's emotion about something, while the term of opinion represents or shows a person's point of view. According to [3], sentiment analysis, also well-known as opinion mining, is the field of study that analyzes people's opinions, feelings, evaluations, attitudes, and emotions. There are two methods that are used to automatically learn sentiment signals which are supervised and unsupervised learning methods. The former involves the presence of two sets of data which are a training set and a test set. The method is called supervised since the system is trained on a training sub-set, which contains models that have already been processed. The latter (unsupervised), which recommends only one data set, requires the system to autonomously restructure the information so that the most similar data will be placed in the same group [4].

Furthermore, there are three analytical approaches which are lexicon-based approach, corpus-based approach, and a Hybrid approach (mixed approach) [4]. Regarding the Lexicon-based approach, it identifies the polarity of a text using two sets of words, the first expresses a positive feeling and the second expresses a negative feeling. Then, the model counts in the text the number of positive words and that of negative words while the sum gives an overall evaluation of the text feeling. Therefore, if the number of positive words outweighs that of the negative ones, the text is considered positive, conversely, the text is considered negative or neutral if the numbers are equal [4].

In terms of Corpus-based approach, the automatic sentiment analysis based on corpora requires the creation of two manually annotated corpora. The first is the learning corpus, which is used to train an automatic system. It includes notes added by human annotators. Thus, a system can perform an analysis on its own while the second uses the test corpus, which is trained to verify the performance of the automatic system. In an ideal scenario, the results of the analysis performed by the automatic system should fully correspond to those of the learning corpus. Therefore, in order to maximize the automatic system's performance, it is important that the learning corpus is representative for the test corpus. [4] gives an example of algorithm like neural network.

In addition, Hybrid approach takes advantage of the previous two methods there are three ways to do it. The first is to use linguistic tools to develop the corpus and then classify the texts using a supervised learning tool. The second way is to use machine learning to build the body of opinion needed for the lexicon-based approach. The third way is the combination of the two previous approaches and the collection of their results [5]. Furthermore, the next section will show the related works on the sentiment analysis.

B. Related Work on Sentiment Analysis

Although the field of sentiment analysis is an emerging field in the natural language processing community, the work done for the Arabic language is still extremely limited [6]. Indeed, most of the research studies have focused on the polarity classification of documents to avoid the excessive cost of sentence annotation. Also, to adopt the machine learning-based approach to escape the excessive cost of creating the lexicon of opinion with good coverage [7]. Thus, the obstacle that slows the advancement of the field of the sentiment analysis for Arabic is the unavailability of resources in terms of annotated corpora and lexicons of opinion.

In this context, [6] are among the first researchers who were interested in constructing an opinion lexicon for the Arabic language. In fact, their work was part of the development of a tool that enables to classify the stakeholders' opinions in the business field. Therefore, the starting number of words to be set were more than 600 positive words, 900 negative words and 100 neutral words. The evaluation tests performed showed good accuracy but a low recall. In addition, [7] proposed the combination approach based on three steps for the classification of Arabic documents according to their polarities. They used an opinion lexicon constructed from the English Lexicon SentiStrength after having done the translation and consulted online dictionaries to enrich the lexicon with the synonyms of the already existing words. However, the size of the lexicon was not mentioned and the accuracy of the classification tool for the lexicon-based step is 48.7%. Also, [8] presented the lexicon "Sifaat", a manually constructed lexicon containing 3325 adjectives categorized in three classes: positive, negative, and neutral. The results of the assessment showed an improvement of 6% in the classification of subjectivity and 40% in the classification of polarity. The authors also proposed to extend the lexicon by translating three English opinion lexicons, namely the SentiWordNet, the Youtube lexicon and the General Inquirer. However, this method focused only on adjectives that have not been evaluated. Furthermore, [9] proposed to create a lexicon by translating 300 words of SentiStrength. Then, enriching it with synonyms and emoticons. The last version of the lexicon included 3479 entries 1262 of which are positive and 2217 are negative. The results of the experiment on the collected corpus reached 59.6% in terms of accuracy. As for [10], they proposed to build an Arabic version of SentiWordNet, which is an opinion lexicon derived from the WordNet database, by going through two stages: updating the base of the WordNet Arabic 2.0 by mapping the WordNet 3.0 for English, and also the base obtained from the SentiWordNet 3.0 for English. In fact, the lexicon coverage assessment reports that 5% of the words in the annotated corpus are not in the lexicon.

In this study the Arabic data set and RNN model will be used to predict sentiments of authors. Next sections will explain the methods uses and demonstrate that our results surpass those obtained by the best deep learning models like LSTM, CNN and GRU.

IV. Methodology

In 2005, things started to change. In fact, the outlook on the field of artificial intelligence has dramatically changed with the machine learning while the emergence of deep learning, which accounts for the bulk of research conducted by specialists, especially as it intervenes in many areas, such as natural language processing. Deep learning (DL) is a branch of machine learning, where the latter is a branch of artificial intelligence, in which machines can learn by knowledge and gain skills without the human contribution. Therefore, based on artificial neural networks, the algorithms encouraged by the human brain, learn from large amounts of data. Consequently, DL enables models composed of multiple processing layers to understand descriptions of data with multiple levels of abstraction. [11].

A. Deep Learning vs. Machine Learning

Machine learning (ML) is a sub-field of artificial intelligence (AI), which focuses on designing systems that learn or improve performance based on the data. These systems are intended to train a safe set of algorithms for larger amounts of data that can help classify future data. ML is divided into Supervised Machine Learning and Unsupervised Machine Learning. The former is the most generic form of machine learning. It aims to make the algorithm "learn" by comparing its actual output to the "taught" outputs in order to check for errors and modify the model accordingly. The latter is also known as learning from observations. The learning algorithm in this method finds common points on its own among its input data. On the other hand, in a machine learning process, the algorithm must be instructed on how to make an accurate prediction using more information (for example, by manually performing an extraction of the relevant features) [7]. In fact, Deep Learning (DL) algorithms can learn to make an accurate prediction through their own data processing, thanks to the structure of the artificial neural network. For example, they ignore manual steps and feature extraction, but the modeling process is automatically performed. Another major difference is the fact that deep learning algorithms evolve with the data. Therefore, to succeed in the deep learning application, a large volume of data and large dimension are needed to train the model by adding one or more GPUs (graphics processor) to quickly process the data. This implies that, if these elements are not needed, it is better to use machine learning rather than deep learning.

Indeed, feature extraction is one of the greatest challenges of traditional machine learning models, which has been automated by deep learning models to enable them to achieve a particularly high accuracy rate for computer vision tasks. Therefore, the ability to manage a large number of features makes deep learning powerful when it comes to unstructured data [13]. Nevertheless, DL algorithms can be overkilled for less complex problems because they need access to a large amount of data to be effective. On the other hand, the deep learning algorithms success depends on the availability of more training data. In fact, Google, Facebook, and Amazon have already started using it to analyze their massive amounts of data. In practice, all deep learning algorithms are neural networks, also called ANN. ANNs are information processing models that simulate the functioning of a biological nervous system. They are similar to the way the brain manipulates information at the functional level. Actually, all the neural networks are made up of interconnected neurons that are organized in layers. Then, the details of the neuron and the description of its functioning will be explained in the next part. In fact, what forms the neural networks are the artificial neurons inspired by the real neuron that exists in our brain [13].

On the other hand, the activation function is an essential component of the neural network. This function consists in checking whether the neuron is activated or not by calculating the weighted sum of the inputs and adding the bias. This is a nonlinear transformation of the input value. Moreover, the nonlinearity is particularly important in neural networks however, without the activation function, a neural network simply becomes a linear model [14]. Actually, there are many types of these functions, such as sigmoid function, ReLu function, and Softmax function. The distinct types of neural networks available for use, such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) [14] [15] [16].

B. Steps of our Approach

Deep Learning has been proven to be effective in many complex problems using artificial neural networks to learn and extract patterns and information significant from the data [15]. Therefore, we find many contributions that attempt to adapt this approach as a solution to the problem of sentiment analysis. In this research, the proposed model (RNN) will be used. Also, other types of neural networks will be applied as comparison experiments to the proposed model. Now we are going to present the tasks necessary to conduct this work as showed in Fig. 1.

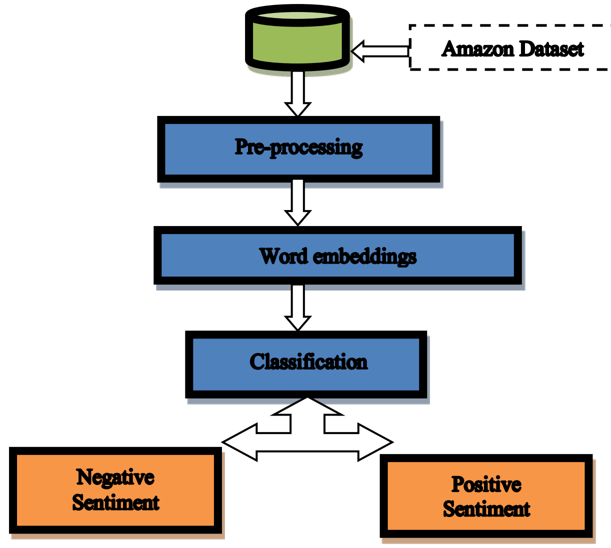


Fig. 1. Architecture of our Method.

- Pre-processing is preparing the dataset to have an excellent quality. This is important tasks that must take place before using a dataset for model training. In order for the models to be professionally trained and to provide the expected results, the data used must be representative, clean, precise, complete, and well labeled. For example, for the prediction the feelings of comments from social networks, the corpus must contain the same type of documents. The preparation of these data is therefore a crucial step [4].
- Word embeddings are the mapping of words to number vectors real in a reduced dimensional space. Word embedding vectors represent words and their contexts; thus, the words with similar meanings (synonyms) or with close semantic relations will have more similar embeddings. In addition, word embeddings should reflect how the words are related to each other. For example, the inscriptions for "man" should be "King" like "woman" is "queen". Since learning word embeddings takes time and the power of calculation, the method started with pre-trained incorporations.
- In the classification step, we apply the RNN model in order to predict the sentiment of the author. We have two output classes: positive sentiment (Happy, in love, smile, etc.) and negative (afraid, not happy, angry).

V. Result and Discussion on Amazon Datasets

For this system to function, the method has to go through the following experiments. The proposed method is validated in two steps, in the first amazon dataset, the corpus is divided into 4,149 test comments, which (which represents 20% of the dataset) to assess this model and have its precision, and 16,593 training comments, and obtain as precision 75% for the training and 70% for the test. Fig. 2 shows the results.

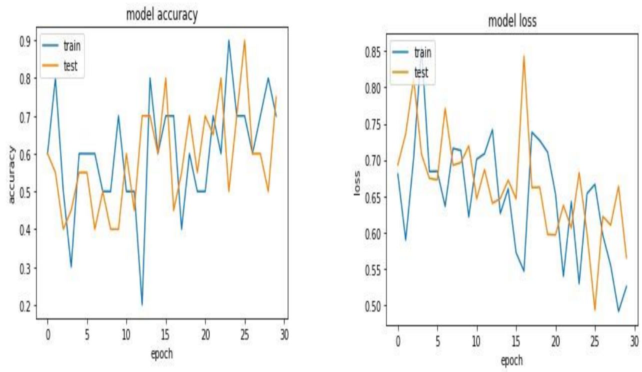


Fig. 2. Precision and Error of this Model for the First Test.

In the second experiment, the data is trained using binary sentiment classification model on a dataset 2 of 66666 items, 33333 positives and 33333 negatives, as shown in the Fig. 4. The best accuracy is obtained with 25 epochs for training and also for test. For the loss, the minimum value is also obtained with 25 epochs for test and 27 for training. Additionally, the corpus is divided into 6667 test comments (which represent 20% of the dataset) to test the model and have its accuracy, and 59999 training comments. Then, we obtained 90% as precision for training and 85% for the test. Fig. 3 shows the results plotted in graphs. For the loss, the best value is obtained with 27 epochs.

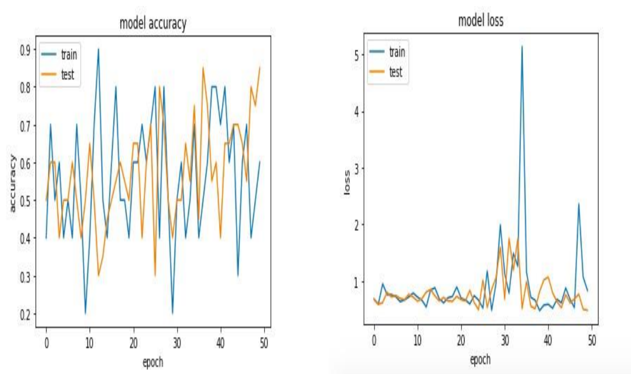


Fig. 3. Precision and Error of the Model for the Second Test.

In the third experiment, we trained our binary sentiment classification model on a dataset 3 of 49,870 items, more than 24,900 positives and 24,900 negatives. Moreover, the corpus is divided into 9973 test comments (which represent 20% of the dataset) to evaluate our model and have its accuracy, then 39897 training comments, where results are 90% for training and 70% for accuracy. For the loss, the best results are obtained with 17 epochs for training and 30 epochs for test. Fig. 4 shows the results plotted in the graphs.

In the first validation, the model was applied to the three data sets mentioned above and the evaluation was made in terms of the accuracy and recall metrics, as summarized in the Table I. Accuracy is the number of relevant documents found compared to the total number of documents proposed for a given query. The recall is defined by the number of relevant documents found with regard to the number of relevant documents that the database has [17].

Our decision to choose different datasets based on their size was to see how a DL model works in two situations where we have a large and small amount of data. For the final tests on the datasets, we trained the 30-epoch model. From what we have as results, the effectiveness of the proposed model and the possibility of its use in this area. For second validation, three models were applied on the same data sets as shown in Table II. Our results obtained using RNN are compared with three most known deep learning Models. In fact, our results surpass those obtained by CNN. For LSTM and GRU we did not obtain encouraging result.

In the third validation, as shown in Table III, we compared our results on the dataset amazon reviews2 with Palash experiments. It can be noted that Palash participated on the kaggle3 competition on Amazons' reviews datasets. Palash presents a method using bidirectional LSTM, for the document representation he based on n-grams (a chunk from 2 to 6 grams) to classify sentiments (positive, negative). The best result for his experimentation is obtained using 5-grams. In Comparison with the last study, we obtained the best result for the accuracy with 96%, but we did not surpass those of Palash for the recall.

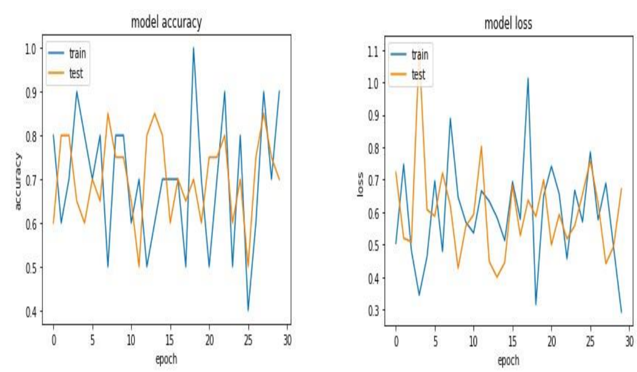


Fig. 4. Accuracy and Loss of our Model for the Third Test.

TABLE I. RESULT ON THREE AMAZON DATASETS

Amazon data	**Training**	**Test**	**Accuracy**
20742	16593	4149	70%
66666	59999	6667	85%
49870	39897	9973	70%

TABLE II. COMPARISON BETWEEN 4 DEEP LEARNING MODELS IN TERMS OF ACCURACY

Deep Learning Model	**LSTM**	**GRU**	**RNN**	**CNN**
DATA1	76	80	85%	83
DATA2	68	65	70%	70
DATA3	68	64	70%	68

[2] https://jmcauley.ucsd.edu/data/amazon/
[3] www.kaggle.com

TABLE III. COMPARISON RESULTS WITH PALASH4 STUDY USING BIDIRECTIONAL LSTM AMAZON REVIEWS DATASET CONTAINING 400000 REVIEWS

	Bidirectional LSTM (Palash)	Our Model
Accuracy	94%	96%
Recall	90%	85%

VI. DISCUSSION

Experiments have shown that our system is efficient compared to other models (GRU, LSTM, etc.). Indeed, for the three datasets our system has obtained the best precision. Nevertheless, for recall we found a worse result than LSTM. Indeed, Lstm used by Spalash gave 90% against our model 85% for the Amazons corpora, our results were less good for data–one and three because the size is minimal, while the results for corpus two were encouraging since we had 66,000 documents.

VII. CONCLUSION

In this paper, we aimed to contribute to the sentiment analysis field by presenting the tools and Arabic datasets, as well as the steps we have followed to obtain the appropriate results for our model. In fact, our sentiment classification model helped us to correctly categorize more than 85% of all the 7480 test items, which means that our system has correctly categorized more than 5984 items. Moreover, the experiments proved that as the size of the training data increases, the accuracy rate increases, which implies that the percentage of the test dataset is as important as it was in the first experiments. Therefore, our model performs well if we use it for the sentiment analysis problem, where the amount of learning data plays a significant role. The proposed system succeeded in obtaining an accuracy of 85%. As Future work, we will depend on multilingual corpora to predict sentiments. Also, we will use reviews about Covid-19 in KSA twitter to predict emotions of people in period of virus vaccination.

REFERENCES

[1] Abbasi, H. Chen, A. Salem, "Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums," ACM Transactions on Information Systems. 008 Jun 20;26(3):1-34.

[2] B. Liu, "Sentiment analysis and subjectivity," In: Handbook of Natural Language Processing, 12. 2010 Feb;2(2010):627-66.

[3] J. Waters, J. Lester, "The Everything Guide to Social Media: All you need to know about participating in today's most popular online communities". Simon and Schuster; 2010 Oct 18.

[4] V. Hee Cynthia," L'analyse des sentiments appliquée sur des tweets politiques : une étude de corpus", Faculté associée de linguistique appliquée Université Bruxelles Belgique, 2013.

[5] D. Poirier, F. Fessant, C. D.e. Bothorel, E.G, Neef, M. Boullé, "Approches statistique et linguistique pour la classification de textes d'opinion portant sur les films". Revue des Nouvelles Technologies de l'Information. 2009 Nov 1: Pages-147.

[6] M. Elhawary and M. Elfeky. "Mining Arabic Business Reviews". In Proceedings of International Conference on Data Mining Workshops (ICDMW), pages 1108–1113. IEEE, 2010.

[7] A. El-Halees. "Arabic Opinion Mining Using Combined Classification Approach". In Proceedings of the International Arab Conference on Information Technology (ACIT), 2011.

[8] M. Abdul-Mageed and M. Diab. "Toward building a large-scale Arabic sentiment lexicon". In Proceedings of the 6th International Global WordNet Conference, Matsue, Japan, 2012.

[9] A. Nawaf, , N. Abdulla, A. Ahmed, A. Mohammed and M. Al-Ayyoub: "Arabic Sentiment Analysis: Lexicon-based and Corpus-based", 2013 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), 2013.

[10] S. Alhazmi, W. Black, and J. McNaught, "Arabic SentiWordNet in Relation to SentiWordNet 3.0", International Journal of Computational Linguistics. 2013 et 4(1):1-11.

[11] Patterson J, Gibson A. Deep learning: A practitioner's approach. " O'Reilly Media, Inc."; 2017 Jul 28.

[12] H. Alamro, M. Alshehri, B. Alharbi, Z. Khayyat, M. Kalkatawi, II. Jaber, X. Zhang , "OVERVIEW OF THE ARABIC SENTIMENT ANALYSIS 2021 COMPETITION AT KAUST", 2021.

[13] I. Vasilev, D. Slater, G. Spacagna, P. Roelants, V. Zocca, " Python Deep Learning: Exploring deep learning techniques and neural network architectures with Pytorch" , Keras, and TensorFlow. Packt Publishing Ltd; 2019 Jan 16.

[14] I. Goodfellow, Y. Bengio, A. Courville "Convolutional networks". InDeep learning 2016 Nov (Vol. 2016, pp. 330-372). Cambridge, MA, USA: MIT Press.

[15] N. Buduma N. Locascio," Fundamentals of deep learning: Designing next-generation machine intelligence algorithms. " O'Reilly Media, Inc."; 2017 May 25.

[16] A. Almulihi, F. Alharithi, S. Mechti, R. Alroobaea, S. Rubaiee, "A Software for Thorax Images Analysis Based on Deep Learning". International Journal of Open Source Software and Processes (IJOSSP). 2021 Jan 1;12(1):60-71.

[17] D. Kumar, U. Bansal, R. Alroobaea, A. Baqasah, M. Hedabou. An Artificial Intelligence Approach for Expurgating Edible and Non-Edible Items. Frontiers in Public Health. 2021;9.

[4]Graduate Researcher at Pukyong National University
Busan, Busan, South Korea

Arabic Sign Language Recognition using Lightweight CNN-based Architecture

Batool Yahya AlKhuraym, Mohamed Maher Ben Ismail, Ouiem Bchir
College of Computer and Information Sciences
King Saud University, Riyadh, KSA

***Abstract*—Communication is a critical skill for humans. People who have been deprived from communicating through words like the rest of humans, usually use sign language. For sign language, the main signs features are the handshape, the location, the movement, the orientation and the non-manual component. The vast spread of mobile phones presents an opportunity for hearing-disabled people to engage more into their communities. Designing and implementing a novel Arabic Sign Language (ArSL) recognition system would significantly affect their quality of life. Deep learning models are usually heavy for mobile phones. The more layers a neural network has, the heavier it is. However, typical deep neural network necessitates a large number of layers to attain adequate classification performance. This project aims at addressing the Arabic Sign Language recognition problem and ensuring a trade-off between optimizing the classification performance and scaling down the architecture of the deep network to reduce the computational cost. Specifically, we adapted Efficient Network (EfficientNet) models and generated lightweight deep learning models to classify Arabic Sign Language gestures. Furthermore, a real dataset collected by many different signers to perform hand gestures for thirty different Arabic alphabets. Then, an appropriate performance metrics used in order to assess the classification outcomes obtained by the proposed lightweight models. Besides, preprocessing and data augmentation techniques were investigated to enhance the models generalization. The best results were obtained using the EfficientNet-Lite 0 architecture and the Label smooth as loss function. Our model achieved 94% and proved to be effective against background variations.**

Keywords—Arabic sign language recognition; supervised learning; deep learning; efficient lightweight network based convolutional neural network

I. Introduction

Communicating with others is an important skill for humans to interact with their environment community. In its absence, exchanging experience and expressing its opinion and feelings become a very challenging task. In fact, communication allows the discussion of different types of issues that concern humans in order to address them and come up with appropriate solutions to facilitate their daily life. Besides, communication is a censorious factor for individual’s mental health [1]. Actually, communication types can be categorized into four groups: (i) verbal, (ii) non-verbal, (iii) visual and (iv) written communication.

Hearing Impaired (HI) persons are actually affected by their disability, and cannot ensure regular communication with others. They typically use non-verbal gestures as a visual communication type that relies on hand movements and facial expressions [2]. On the other hand, deaf people use sign language as their primary communication technique, where sign language is based on visual-motion codes. This codified system is defined using standard positions and hand movements supported by facial expressions [3]. Typically, hearing impaired people use hand gestures, called signs, to interact with others. Commonly, for existing sign languages, the main signs features are the handshape, the location, the movement, the orientation and the non-manual component. Thus, sign language recognition system assists people with hearing disabilities to communicate with healthy persons. Particularly, such recognition systems are meant to deliver the semantic corresponding to hand gestures of sign language to the interlocutor.

Sign language is considered as a descriptive language which it composed of hand gestures and facial expression. Arabic Sign Language (ArSL) used in approximately 22 Arab countries with different gestures. The discrepancy noted for some word gestures can be attributed to the cultural diversity between these countries. Despite this lack of uniformity of ArSL, all of the 22 countries use the same gestures for the Arabic letters and numbers [4]. Fig. 1 shows the standard Arabic sign language that expresses the Arabic alphabets.

Next sections are mention and discover the main types of ArSLs recognition systems, first, image-based ArSL recognition, can be categorized into three main groups: (i) continuous, (ii) isolated word, and (iii) alphabet recognition systems. Typically, those systems contain five major stages, where each single stage is meant to achieve a particular task. Namely, those five stages are: acquisition of images, preprocessing the images, then extraction the features from that images, after that the images segmentation, and lastly classification process.

Second, alphabet recognition, signers usually perform letters signs independently in the context of alphabet recognition scenario. To represent letters, static poses are used, and the size of vocabulary is minimal. Although the Arabic alphabet includes 28 letters, ArSL considers 39 signs [5]. Actually, the 11 added signs reflect simple signs that combine two letters. Third, word sign recognition, its techniques are relying on further analysis of the images. These techniques are more practical than alphabet recognition, but the alphabet recognition is less complex than word recognition. The major purpose of using word sign recognition is process various images in a sequence.

Fig. 1. Unified Alphabets for Arabic Sign Language.

Deep learning is introduced in Artificial Intelligence (AI) as a sub-group of machine learning [6]. For image recognition issues, deep learning algorithms have provided efficient solutions. In deep learning, Convolutional neural network (CNN) is considered as a class of deep neural networks commonly exploited in computer vision. The optimization of CNN networks can be introduced as a set of algorithms or methods used to alter the neural network attributes like weights, learning rate for reducing the loss costs, and provide an accurate result. In particular, Adam optimizer is an optimization algorithm for stochastic gradient descent used to train the models of deep learning [7]. In 2019, Google introduced new family of CNNs named EfficientNet [8]. EfficientNet proved to have lower number of parameters and Floating-Point Operations Per Second (FLOPS) and yield competitive accuracy [8].

In this project, we propose an image-based ArSL recognition system that relies on lightweight Efficient Network (EfficientNet) [8] to enhance the overall ArSL recognition performance and reduce the number of parameters and hence the time complexity. Standard performance measures and real datasets will be considered to evaluate the performance of the proposed system.

II. Related Work

ArSL recognition is even more challenging than ASL due to the considerable similarity between some of its letters signs. Recently, several approaches have emerged to address issues relevant to ArSl recognition. Particularly, researchers have investigated glove-based and glove-free approaches for ArSL classification. Both approaches rely on machine learning techniques but to different extent. In this situation, some models employ deep learning techniques, while others deploy shallow classification models. This literature surveys aims at giving an overview on state-of-the-art deep-learning and shallow-model based ArSl recognition approaches.

A. *Shallow Model based Approaches*

An ArSL recognition system was proposed in [9]. Specifically, a framework based Scale-Invariant Features Transform (SIFT) features was designed to extract the visual properties of ArSL gestures. In fact, SIFT extraction algorithm encloses five main steps. First, it requires rolling together with the Gaussian filter of different widths with the image. It is essential as it can help demonstrate the Gaussian function pyramid's difference. The extrema of the Gaussian pyramids are identified by comparing each point to its neighbors. The removal of main points at the extremes that are thought to be susceptible to noise are situated on edge. The next important thing is to determine the orientation. Next, in order to achieve that, it creates a histogram. It can be created from the sample points' gradient orientations. Finally, for the purpose of a local image region, a descriptor is produced. Afterward, a Linear Discriminant Analysis (LDA) was conducted to achieve dimensionality reduction. The researchers fed the resulting feature to classifiers: K-Nearest Neighbor (KNN) and Support Vector Machine (SVM). The test results revealed that SVM outperformed KNN with a 98.9% accuracy. In [10], the authors introduced a similar approach to the one outlined in [9]. Although, their approach used LDA alongside Gaussian Mixture Model to fit the dataset. The experiments relied on a dataset collected using two dual Leap Motion Controllers (LMC). Two native signers were involved in the collection of 100 Arabic signs. The proposed framework achieved an accuracy of 92%, which is relatively higher than for similar sensor-based techniques.

In [11], a system for automated ArSL recognition using visual descriptors was introduced. The system was designed to generate a large number of visual descriptors in order to provide an effective ArSL alphabet recognizer. In particular, a one-versus-All SVM was used to classify the generated visual descriptors. According to their findings, the Histograms of the Oriented Gradients (HOG) descriptor significantly outperform the other descriptors. In addition to this, as far as the ArSL recognition is concerned, the method that was brought forward attained 90.55% of recognition accuracy.

Other researches were conducted to develop ArSL recognition schemes based on the nearest neighbour classification. Specially, in [12], a lexicon of 80 words was considered and a sensor-based dataset was collected. This dataset was then used to assess the performance of the glove-based ArSL recognition system. Following the data labeling stage, a low-complexity preprocessing and feature extraction techniques were deployed in order to capture the data temporal dependency. Additionally, a Modified K-Nearest Neighbor (MKNN) was utilized to classify the outputs. The system achieved a sentence recognition rate of approximately 98.9%. The approach was considered superior to vision-based approaches pertinent to classification rates. Similarly, in [13], a KNN algorithm was also used in the design of an ArSL algorithm. With regards to the classification rates, it was witnessed that the method outperforms the methods that were vision-based. In an ArSL algorithm design, the use of the KNN algorithm was done additionally. There were various things that were included in the system. These things included histograms and their comparisons and generation, image masking, narrowing and clipping of images, and recording and questioning of images. A test for the glove recognition and hits was done. The test was done on the established algorithm, and ultimately, around 90% of the hit rate was attained by most of the characters. As the algorithm did not possess an engine, which is a self-learning one, it was regarded as imperfect and faulty. The researchers in [14] suggested a method for detecting hands and faces based on skin profiles using input images translated to YCbCr color space. As image transformation, morphological dilation operation was also used. The Prewitt operator was used to detect hand form edges, and the Principal Component Analysis (PCA) was applied to achieve dimensionality reduction and obtain the final feature. The experiment showed a 97% accuracy on KNN based on 150 signs and gestures.

A Gray Level Co-occurrence Matrix (GLCM) feature was used to implement an ArSL transformation system suggested in [15]. Despite the fact that the system was not deep learning-based, it had four phases: processing, feature extraction, matching technique, and display translation. A mixture of 15 GLCM and histogram features is used to remove features. The system correctly recognized nineteen Arabic alphabets with a 73% accuracy rate, according to the test results. Existing systems used Hidden Markov Models (HMMs) to identify the ArSL in various ways and techniques. HMMs help developers create signer-independent systems that do not need gloves or attached sensors. The authors in [16] described an HMM-based automated ArSL recognition system. The main three steps of keeping the tabs on and detecting the hand involved tracking the fingerprints, detecting the edge, and identifying the skin. In order to reduce error rate and improve the detection of edge significantly, a certain type of algorithm was utilized. That algorithm is referred to as a Canny algorithm [17] where this algorithm acts like an optimal edge detector. For all of the frames, it was scrutinized that whether the outline of the image was rounded or not. Along with that, for the purpose of assessing observed regions of skin, an integrated component analysis was also deployed. Although the experiment yielded 82.22% detection rate, the device only used eight features per frame. As a result, the system's ease of use relies heavily on fewer features. In [18], several spatiotemporal feature extraction techniques that could be used to recognize isolated ArSL both offline and online were introduced. The researchers used forward, backward, and bi-directional projections to extract video-based movements. Now, to make certain that in order to delineate the video series, little amount of coefficients was used, and to terminate the terrestrial reliance, all of the errors were gathered and assembled into one particular image. The model that was proposed encompassed filtering such as a low pass filtering, Radon transformation, and Zonal coding. Along with these, it also includes a 2-D transformation. The experimental findings showed that in the identification of ArSL, output accuracy ranged from 97% to 100%. Different techniques involved various pattern recognitions and feature extraction techniques such as PCA, Local Binary Patterns (LBP), and the HMM were used in [19]. Hands and heads were detected, attached components were labelled, and features were extracted accordingly. The experiments were based on 23 isolated Arabic words performed by three different signers. If we talk about the experiment done initially, the rate of the system recognition was computed in the absence of skin segmentation. Moreover, the rate of recognition with the features such as PCA-only, accompanied by the 30 eigenvectors, reached approximately 99.8%. The combination of LBP and PCA features with HMM led to the achievement of 99.9% recognition rate with 20 eigenvectors. LBP was applied in the description of the shape and texture of images, while PCA was used in the reduction of dimensionality. In [20], the researchers presented a method for creating a smart glove that can recognize ArSL. They did not report the accuracy of the simulation device, even though it was found to be cost-effective. In [21], the implementation of a mobile-based system was outlines. Since it contains several levels, the system tends to be compatible with a self-supervising system for deep learning. The results include details about the system's compatibility but no indication of its accuracy. The method described in [22] can also be thought of as a non-deep learning approach. Because the data was not reliable, the system relied on blob tracking using Euclidean distance. According to the report, the system had a 97% recognition rate based on a dataset of 30 different words.

B. Deep Learning based Approaches

In [23], a vision-based system for recognition for the Arabic sign language was proposed. It relies on CNN architecture to translate gesture images into Arabic speech in a supervised manner. This system detects hand sign alphabet and speaks out the corresponding Arabic letter using deep learning model. The feature extraction, considered as the first essential component of the system was followed by the classification component. During the feature extraction phase, the system transfers the input images into a three-dimensional (3D) matrix for depth, width and height specification. Then, the pooling layer reduces the size by decreasing the parameters number. After that, the classification is achieved through the fully connected layer. The reported performance attained 90% for the accuracy. Another Arabic sign language recognition system based on fine-tuning deep CNN was proposed in [24]. The system was evaluated using a set of 32 hand gestures categories. The system adapted Residual Network (ResNet-152) [25] and Visual Geometry Group (VGG-16) [26] as typical deep learning architectures, but with soft-max layer classification after the fully connected layer to improve the prediction performance. The input images fed into the resulting networks are two-dimensional (2D) images unlike in [23]. The reported accuracy was nearly to 99%, which considered as a very high accuracy. In [27], the researchers proposed an ArSL recognition system based on LeCun Network (LeNet-5) [28]. The considered network is composed of seven layers. The first four layers are responsible for the feature extraction process from the image. The last three layers are responsible for the classification task to categorize the input images into their corresponding meaning. The dataset used in that work contains 7869 Arabic sign language letters and numbers. The testing phase of the system has been done using 80% of dataset as a training set. An accuracy of 90.0% was achieved by their model. In [29], a different ArSL recognition system based on deep CNN was depicted. The system aims at reducing the number of parameters in order to extract and detect the hand gestures. A collection of 50000 images including signs performed by a population of signers from various age groups. The system achieved an accuracy of 97% as best performance reported in their experiments. In [30], the authors investigated different transformation techniques for features extraction. The extracted features were then conveyed to the classification model for the prediction the Arabic sign. This study has compared three different classification methods. Namely, they used SVM, KNN and Multilayer Perceptron (MLP) for classification. The association of MLP with Hartley [30] transformation technique achieved nearly to 99% accuracy as recognition rate. Different approaches, such as image-based and sensor-based, were introduced in [31]. The proposed system for Arabic sign language recognition deepened on using a very common technique called Leap Motion Controller (LMC) [32], used as a backbone for this system. The LMC device aimed to detect the hand in order to provide the hand location and motion to facilitate the feature extraction process to classify the Arabic letters correctly. The system achieved a higher performance rate compared to Naive Bayes classifier with 98%, and 99% compared to the MLP neural networks. These results were obtained using 2800 images of Arabic sign language letters. Coupled with the findings in [30], the results show that MLP can be used successfully with LMC in the development of ArSL systems. Likewise, the researchers in [33] proposed an ArSL recognition system based on MLP neural networks for digital-sensor. The system achieved approximately 88% accuracy rate by recognizing 50 different dynamic markers represented by two signers. In [34], the authors developed an Arabic Sign Language Alphabet Translator (ArSLAT) system. Their system includes five stages: pre-processing, frame determination, category, features extraction, and finally the classification stage; for features extraction using translation scale and rotation invariant to achieve the flexibility of system; applied comparison between accuracy of system by using MLP and Minimum Distance Classifier (MDC) [35]. In addition, the result illustrated that system with MDC has an accuracy with 91%, while the MLP with system has 83%. In [36], a model was proposed to recognize the hand gestures of ArSL letters using supervised learning techniques and natural user interface libraries with the LMC skeleton [37] and Kinect [38]. In order to predict the hand captured from an input images that fed to the system in 3D manner as the system introduced in [30]. The results showed that their system cab recognize 22 out of the 28 alphabets with 100% accuracy.

In [39], the researchers developed an ArSL recognition system based on Recurrent Neural Networks (RNN). This system relies on four different phases: acquisition, processing, feature extraction, and recognition of image. The processing of the image requires a colour system called Hue, Saturation, and Intensity (HSI) to extract the features associating with colour layers. As result, this system accomplished 95% accuracy in the recognition of 28 ArSL letters. In [40], an Adaptive Neuro-Fuzzy Inference System (ANFIS) [41] was used for ArSL recognition. The system is based on gestures that represent the Arabic alphabets. Designing the system of many ANFIS networks in order to training on one gesture per network. Pre-processing the hand gesture to extract the features for classification process. The result of this system attained 93.55% accuracy in detecting 30 Arabic letters. On the other hand, a deep learning framework for isolated Arabic Sign Language gestures recognition was proposed in [42]. The purpose of that framework is to encounter three different challenges faced by different ArSL recognition systems. The challenges are: hand segmentation, extract features, and recognition of gesture sequence. To handle these challenges, the authors proposed three different networks. Namely, they used DeepLabv3+ for segmentation process, Convolutional self-organizing map for features extraction process, and lastly the Bi-directional long short-term memory for classification process. Finally, for testing, they used 3450 images that expressed 23 isolated words from three signers and accomplished approximately 89.5% accuracy.

In conclusion of this section, one can notice that existing related works can be categorized into two main different groups: (i) Approaches for ArSL recognition systems based on shallow machine learning models, (ii) deep learning based approaches. Both categories include different classification techniques and methods to recognize alphabet, numbers, words and sentences of Arabic Sign Language. In this project, we achieved our aim by proposing a lightweight based CNN-architecture to enhance the computational and classification efficiency of the ArSL recognition system.

III. Proposed Method

We proposed an Arabic sign language (ArSL) recognition system based on lightweight EfficientNet CNN [43]. The proposed system is intended to translate automatically hand gesture images into the corresponding Arabic sign language letter. ArSL recognition problem is tackled as a supervised deep learning using a lightweight EfficientNet network. This choice of deep architecture aims at achieving a good trade-off between the ArSL recognition rate and the model complexity. In other words, the proposed research seeks a lightweight deep learning model to ensure high ArSL recognition accuracy. In particular, we adapted two EfficientNet models learned from the baseline model, EfficientNet-B0 [44]. Specifically, we scaled it down in terms of number of channels, depth and resolution. In other words, the CNN-based EfficientNet is the backbone structure of our proposed architecture.

The classification task is performed using fully connected (FC) layers where an FC layer hosts neurons which manage comprehensive connections. These connections determine a previous layer's activation. Also, when specific input and output are present, their representational mapping is aided by the FC layer. This layer borrows the regular neural network principles to execute its assigned functionalities. However, this FC layer works with one-dimensional data. The three-dimensional to one-dimensional data transformation would requires the use of a flatten function for the proposed system to be implemented successfully. Subsequently the output will be the Arabic letter that associated with the input image of Arabic sign language.

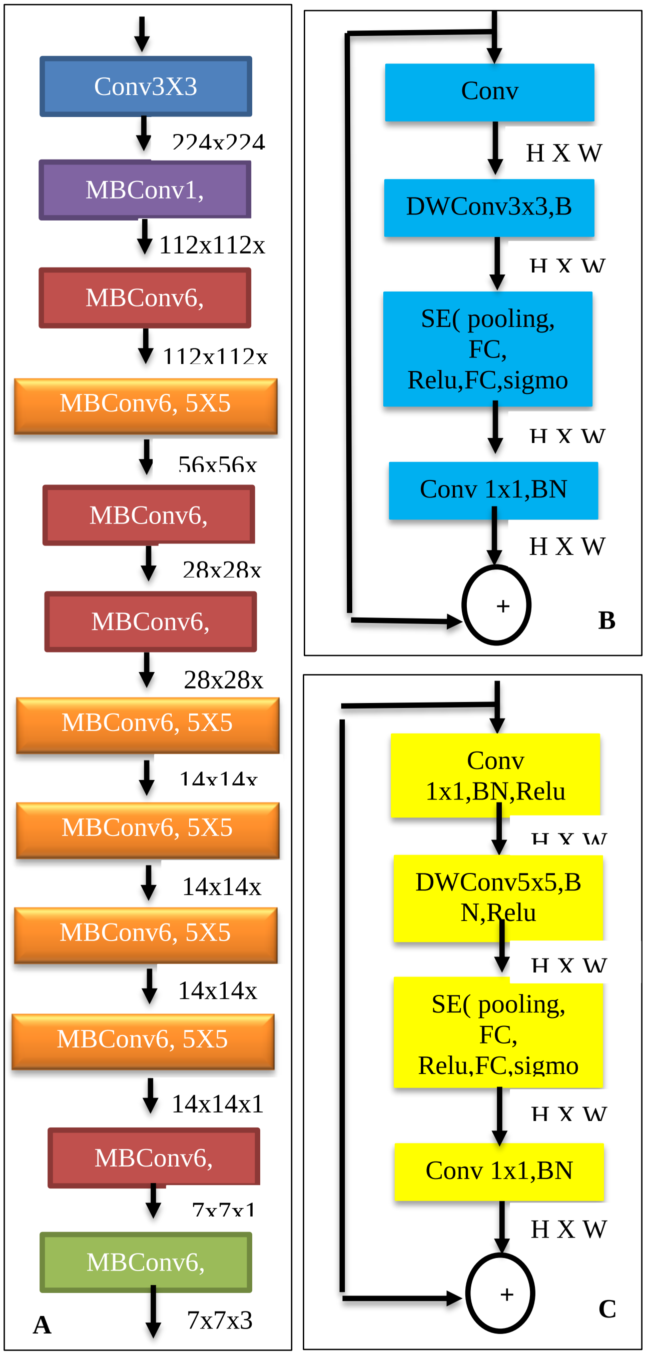


Fig. 2. Model Architecture (A) A Lightweight EfficientNet, (B) MBConv6 (k 3x3) and (C) MBConv6 (k 5x5).

One should note that since the input size tends to be bigger than the feature map size, the implementation of padding becomes the ideal solution to overrule the feature map shrinkage. It helps match the input with stride and kernel size. The resulting system state would yield an improved performance. On the other hand, the natural position of the pooling layer is between convolution layers. It serves the objective of reducing dimensionality and lessening the computations involved. The successful implementation of the pooling layer is due to the fewer number of parameters in play.

The proposed pooling is max pooling. It scans all windows and takes the highest value resulting in a reduced feature map size.

The proposed model uses inverted residual block called (MBConv) as the main building block of the architecture. In fact, MBConv is a mobile inverted bottleneck. It encloses a squeeze-and-excitation optimization [45] element. This mobile inverted bottleneck can be perceived as an inverted residual block that includes a 1×1 convolution layer with batch norm and Rectified Linear Unit (Relu). It is also followed by a 3×3 or 5×5 depth-wise convolution with batch norm and Relu. Besides, a pooling-based Squeeze-and-Excitation (SE) optimization block is added. Next, FC layer and a Relu followed by FC layer and a sigmoid are stacked. It is worth noting that the scaling (i.e., Multiplication Operator (MUL)) aims at multiplying each channel by the input feature to obtain the final output of the SE Block. Finally, a 1×1 convolution layer with batch norm is inserted. Note that the inverted residual block exhibit skips connection between the layers. Fig. 2 details the proposed network architecture. Besides, the loss function is the one that used to compute the difference between the logit of the real class and the logit of the output class. Since our task is a multi-class problem, multi-class cross entropy is generally used in such problem. However, a “label smoothing” technique was introduced for robust modeling [46], and it has been used in various state-of-the art-models ranging from image classification to language translation and speech recognition in an attempt to improve accuracy by using a weighted mixture of targets from the dataset with uniform distribution instead of the hard targets to compute cross entropy.

A. Dataset

We collected real ArSL images to be used in the planned experiments. The collection images were captured using diverse smartphones. More than twenty volunteers participated in the collection of this dataset. Each volunteer performed thirty gestures corresponding to ArSL alphabet. Each character from this alphabet is represented using ten instances. This yields a total of 5400 images. Fig. 3 shows sample of our dataset for the Arabic latter “Baa”. As one can see, those images exhibit different visual properties of their background. This choice is meant to make the ArSL recognition task even more challenging. In other words, this would prove the ability of the proposed model to capture the most relevant visual features in order to discriminate better between the different ArSL alphabet classes.

B. Training Framework

The network will be trained like any usual CNN, by feeding the training input and output are fed to the model. Then, the model will keep iterating the training process through forward propagation and backpropagation. However, using EfficientNet has some requirements to use the pre-trained layers’ weights and continue the training process on the added layers. Where the size of RGB-images of sign gestures will be reduced to 224 x 224 to enter the network in training phase. Then, the CNN-based feature extraction processes the captured imaged for feature maps detection. After that, these maps enter the Lightweight EfficientNet to produce potential hand gestures of the input. Finally, the scaled feature maps are conveyed to FC layer classifier to generate the ArSL alphabet prediction. Then the model, using the output, will backpropagate and adjust the weights.

Fig. 3. Sample Images of "Baa" Letter from our Collected ArSL Data.

IV. Result

In order to evaluate the proposed models, we conducted four experimental scenarios. In the first one, we investigated the effect of segmentation on the system performance. Specifically, we fed the designed model with original and segmented (just the rectangle of hand was used as input) images. In this scenario Adam algorithm [7] was used to optimize the loss function. In the second scenario, we analyzed the system behavior when a Cross Entropy function is used as loss function in addition to Cross Entropy Label Smoothing. Same as in the first scenario, Adam optimizer was used here too. The third scenario was dedicated to the application of different versions of EfficientNet Lite model to see which model yields the best performance. As previous, Adam optimizer was used. The fourth scenario was designed to analyze the system behavior when four optimizers were applied in addition to Adam. In this scenario, to speed up the experiences, the lightest model (EfficientNet Lite 0) was the base model and all images were segmented.

One should note that the dataset described in Section III. proposed method is split into three subsets: (i) A training set including 4320 (80%) of images, (ii) A test set that includes 540 (10%) of the total number of images, and (iii) A validation set that represents 540 (10%) of the total number of images. For deep learning, the best practice is to use train/test/validation split rather than cross validation. All experiments were conducted using Nvidia K-80 GPUs associated with 16 GBs of RAM. The EfficientNet Lite network as well as the training, validation and the test procedures were implemented using Pytorch library [47]. Moreover, five pre-trained models published in [48] were exploited in our experiments.

Table I summarizes the three first scenarios settings. Actually, there are 9 EfficientNet Lite versions; however, the pre-trained weights are available only for the first five models from Lite 0 to 4. We decided that reporting the results of the lighter and heavier model could be concise and more informative. Where Table II reports the performance measures obtained using these models and the test data outlined earlier, *Model 4* (the lightest version with segmented data and label smooth loss function) outperformed the other models, and achieved an accuracy of 94.30%.

A. Impact of Segmentation

The lightest models, Model 2 and Model 4, attained an accuracy of 94.30% and 84.54%, respectively. This shows the impact of segmentation of the model performance. This considerable difference in performance is expected. In fact, the segmentation eliminates the undesirable image parts and artifacts and allows the models focus and learn from the relevant content and discard the irrelevant regions. On the other hand, Fig. 4 shows the evolution of the accuracy during the training and validation phase. As one can see, the gap between the training and the validation curves is larger in the case of unsegmented data.

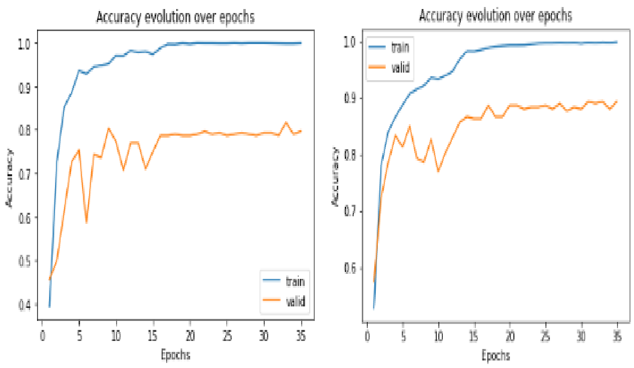
TABLE I. Model Settings.

Model name	w / Segm	w/o Segm	CrossEntropy	CrossEntropyLabel Smooth
EfficientNet_Lite 0 **Model 1**		✓	✓	
EfficientNet_Lite 0 **Model 2**		✓		✓
EfficientNet_Lite 0 **Model 3**	✓		✓	
EfficientNet_Lite 0 **Model 4**	✓			✓
EfficientNet_Lite 4 **Model 5**		✓	✓	
EfficientNet_Lite 4 **Model 6**		✓		✓
EfficientNet_Lite 4 **Model 7**	✓		✓	
EfficientNet_Lite 4 **Model 8**	✓			✓

TABLE II. Summary of Results Obtained by Model Evaluation on Testing Data

Model name	Accuracy (%)	F-measure (%)	GFlops	# params
EfficientNet_Lite 0 M1	79.73	80.21	0.4	3.41M
EfficientNet_Lite 0 M2	84.54	84.80	0.4	3.41M
EfficientNet_Lite 0 M3	87.83	87.53	0.4	3.41M
EfficientNet_Lite 0 M4	**94.30**	**94.24**	0.4	3.41M
EfficientNet_Lite 4 M5	88.60	88.83	2.64	11.76 M
EfficientNet_Lite 4 M6	87.26	87.36	2.64	11.76 M
EfficientNet_Lite 4 M7	91.95	91.96	2.64	11.76 M
EfficientNet_Lite 4 M8	91.28	91.18	2.64	11.76 M

params: Number of model parameters, *GFliops*: Floating operation per seconds.

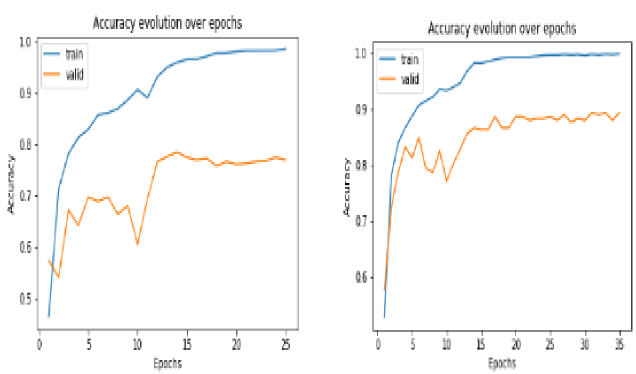


(a) (b)

Fig. 4. Accuracy over Epochs, (a) Model 2: Unsegmented Images were used. (b) Model 4: Segmented Images were used.

B. *Impact of Label Smooth on Loss Function*

The following results were obtained by evaluating the lightest pair of models, Model 3 and Model 4 in Table II. They illustrate the impact of the Label Smooth on the Cross Entropy loss function. An increase of 5% in accuracy can be noticed for the models using the Label Smooth. In fact, the Label smooth is typically used as a regularization technique that improves the model generalization. Particularly, Fig. 5 shows a larger gap between train and valid curves when no label smooth is used. This proves that the model over fits the training data before using this regularization technique.



(a) (b)

Fig. 5. Impact of Label Smooth Regularization: (a) Model M3: without Label Smooth3. (b) Model M4: with Label Smooth.

C. *Impact of Optimizers*

In these experiments, we tried to investigate five different optimizers known as Adam, RMSprop, AdamW, Adadelta and SGM optimizers. In order to figure out the best optimizer among them that had a good effect on the Model 4 to produce a good performance rate for recognition process of ArSL. Table III summarizes the results obtained from the fourth scenario when changing the optimizer in the training loop.

In fact, Model 4, the lightest version with optimal settings, is used for evaluation in this experiment. Namely, the optimizers used are Adam [7], RMSprop, AdamW, Adadelta and SGD algorithms. In RMSprop, stochastic gradient is applied with mini-batch and adaptive learning rates. Instead of accumulating prior gradient values, Adadelta adapts the learning rate using a moving window of gradient updates. Adam combines the good proprieties of RMSpro and Adadelta. It uses Momentum and adaptative learning rate, in other words, the learning rate is gradually adjusted over time. It remains the most prevalent optimizer used in deep learning. Similarly, AdamW optimizer is a variation of Adam optimizer in which the optimization is performed for the weight decay and learning rate separately. Under particular circumstances, it is assumed to have a faster convergence rate than Adam. SGD is an optimizer that updates the weights for each training sample over a limited-size subset of data. As expected, Adam outperforms all others.

TABLE III. MODEL PERFORMANCE ON TESTING DATA USING DIFFERENT OPTIMIZERS

Model 4		
Optimizer	**Accuracy**	**Loss**
Adam	**94.30%**	**0.054432**
RMSprop	93.29%	0.055978
AdamW	92.62%	0.055098
Adadelta	89.94%	0060984
SGD	86.93%	0.070220

D. *Impact of Data Augmentation*

We noticed in the experiment scenarios, the training and validation images are augmented using random horizontal flip transformation. Table IV shows the performance achieved with and without data augmentation. As it can be seen, the flip transformation increases the Accuracy and F-measure by more than 3%.

TABLE IV. IMPACT OF DATA AUGMENTATION

Settings: Model 4		
	Accuracy	**F-measure**
With Horizontal Flip transformation	**94.30%**	**94.24%**
Without Horizontal Flip transformation	91.16%	90.98%

E. *Result of EfficientNet_Lite 0 Model 4 on Arabic Sign Language*

In the following, we report the performance measures recorded for letters obtained by Model 4. Specifically, accuracy, the precision, recall and F-1 measure in Table V. Most letters are recognized correctly despite the similarities in some gestures like ("shin: ش " ,"sin: س ") , ("dhad: ض " ,"sad: ص "), ("thal: ذ " ,"dal: د "), and ("ray: ر " ,"zay: ز ") as depicted in Fig. 1 above. Further analysis showed that the 10 instances of each letter used in test are from the same signer. However, very few instances within one letter were not recognized. This is due to the intra-class variation or the inter-class similarities in gestures mentioned above. The least correctly classified letters are "Ra ر " (70%) and "Ain ع " (70%), where extra data could help to improve recognizing those two letters. On the other hand, there are 13 letters out of 30 with a 100% precision, where the rest of the alphabets did not have low precision except two letters with 70% as discussed before; applied different measures to assess Model 4 in order to provide an overview of its functionality and performance.

TABLE V. PERFORMANCE MEASURES OBTAINED USING MODEL 4

	Accuracy	Precision	Recall	F1
Model 4	94.30%	94.3%	94.46%	94.13%

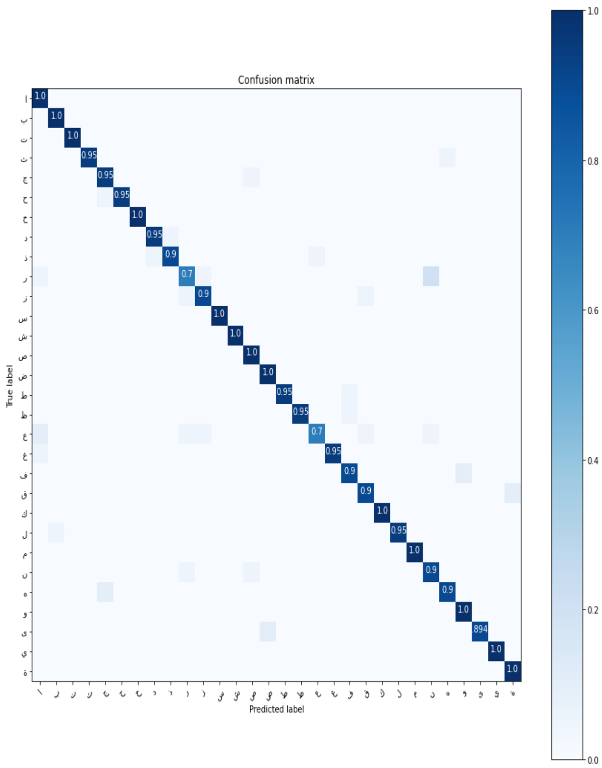


Fig. 6. Confusion Matrix obtained using Model 4.

Fig. 6 illustrates the confusion matrix obtained by the evaluation of the Model 4 by accuracy measure. Where it expresses a visual chart related to the predicted label of letters produced by Model 4 and the true label letters. As we can see here, two diagonal squares with a light color and an accuracy value of 0.7 corresponding to the letters (ra, 'ر') and (Ain 'ع'), respectively.

F. Comparison with other Existing Work

A comparison of the proposed recognition systems versus some of the most relevant work that cultivated the highest accuracy for ArSL recognition was conducted. We compared the results obtained by our best model, Model 4 (lightest version) to the work proposed in [48]. This existing work is a two Faster R-CNN technique based on VGG-16 and ResNET-18. As can be seen in Table VI, our model outperforms the existing works in term of accuracy by 1%. Furthermore, the size of our model is very small compared to the size of compared models.

TABLE VI. GLOBAL COMPARISON TABLE BETWEEN EXISTING WORK IN [48] AND THE PROPOSED METHOD

Model	**Accuracy**	**# of parameters**
Faster R-CNN(VGG-16)	93.2%	138M
Faster R-CNN(ResNET18)	93.4%	11M
EfficientNet_Lite 0 Model 4	94.30%	3.41M

V. CONCLUSION

Hearing-impaired people are suffering from the communication with others in an easy way since they have to learn the sign language, which considered as their formal language to interact with community. The overall performance of image-based solutions for this problem depends on the segmentation quality and the choice of the features that should encode the main visual properties of the sign language gesture. The existing solutions exhibit considerable rooms for improvement for the ArSL language recognition solutions which typically rely on heavyweight Convolutional Neural Network (CNN) models. The common hardware is insufficient for the existing models to deliver sufficient in real time in order to interpret the ArSL into Arabic spoken language. EfficientNet, introduced by Google, contains many CNNs layers that have high accuracy and also improve the efficiency of the models by reducing the number parameters and the computational cost. In this project, we proposed a system to recognize the Arabic sign language (ArSL) using a CNN based lightweight EfficientNet. Particularly, the proposed architecture contains two phases, feature extraction and classification in order to process the input image to produce the Arabic alphabet associated to that input. During this first phase of the project, we provided an overview on the background required for this research. Moreover, we conducted a literature review to survey existing relevant ArSL recognition system and approaches. We applied a real dataset to develop and deploy a system serving for the interpretation of letters of Arabic sign language. Then, standard performance measures adopted to assess the performance of the proposed system and compared its results with state-of-the-art approaches particularly VGG and ResNet.

As a future work, we propose to investigate Transformer for vision computer, a convolutional neural network (CNN) free deep learning architecture [49] based on self-attention mechanism. Where vision Transformer attains excellent results compared to state-of-the-art convolutional networks while requiring substantially fewer computational resources to train. Furthermore, we recommend extending this work to recognize the Arabic sign language words or common expressions.

ACKNOWLEDGMENT

This work was supported by the Research Center of the college of Computer and information Sciences at King Saud University, Riyadh, KSA. The authors are grateful for this support.

REFERENCES

[1] C. Martin, and N Chanda,"Mental Health Clinical Simulation: Therapeutic Communication" Int. J. of Clinical Simulation in Nursing, Vol.12, Issue 6, pages 209-214, Jun.2016.

[2] E. K. Elsayed, and D. R. Fathy, "Semantic Deep Learning to Translate Dynamic Sign Language", Int. J. of Intelligent Engineering and Systems, Jan. 2021.

[3] A. Horňakova, and A. Hudakova, "Effective communication with deaf patients", Int. J. of Original scientific article,VOL.4, NO.7, 2013.

[4] M. Mustafa, "A study on Arabic sign language recognition for differently abled using advanced machine learning classifiers", J. of Ambient Intelligence and Humanized Computing 12, 4101–4115, Mar. 2020.

[5] M. Mohandes, "Arabic sign language recognition", In International conference of imaging science, systems, and technology, Las Vegas, Nevada, USA, vol. 1, 2001.

[6] X. Chen, L. Zhang, T. Liu, and M. M. Kamruzzaman, "Research on deep learning in the field of mechanical equipment fault diagnosis image quality", Journal of Visual Communication and Image Representation, Vol. 62, pages 402–409, Jul. 2019.

[7] D. Melinte and L. Vladareanu, "Facial Expressions Recognition for Human–Robot Interaction Using Deep Convolutional Neural Networks with Rectified Adam Optimizer", MDPI Journal, Vol. 20, Issue. 13, Apr. 2020.

[8] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks", In International Conference on Machine Learning, pages 6105-6114, May.2019.

[9] A. Tharwat, T. Gaber, A. Hassanien, M. Shahin, M and B. Refaat, "Sift-based Arabic sign language recognition system", In Afro-european conference for industrial advancement, Vol.334, pages 359-370, 2015.

[10] M. Deriche, S. Aliyu, and M. Mohandes, "An intelligent arabic sign language recognition system using a pair of LMCs with GMM based classification" IEEE Sensors Journal, Vol.19, No.18, pages 8067-8078, Sep. 2019.

[11] R. Alzohairi, R. Alghonaim, W. Alshehri, S. Aloqeely, M. Alzaidan, and O. Bchir,"Image based Arabic sign language recognition system", International Journal of Advanced Computer Science and Applications (IJACSA), Vol.9, No.3, 2018.

[12] N. Tubaiz, T. Shanableh, and K. Assaleh, "Glove-based continuous Arabic sign language recognition in user-dependent mode", IEEE Transactions on Human-Machine Systems, Vol.45, No.4, pages 526-533, Aug. 2015.

[13] R. Naoum, H. Owaied, and S. Joudeh, "Development of a new arabic sign language recognition using k-nearest neighbor algorithm", Journal of Emerging Trends in Computing and Information Science, Vol.3, No.8, pages 1173-1178, Aug. 2012.

[14] E. Hemayed and A. Hassanien, "Edge-based recognizer for Arabic sign language alphabet (ArS2V-Arabic sign to voice)". International Computer Engineering Conference (ICENCO), pages121-127, Dec. 2010.

[15] A. El Alfi and S. Atawy, "Intelligent Arabic sign language to Arabic text translation for easy deaf communication" International Journal of Computer Applications, Vol.180, 2018.

[16] A. Youssif, A. Aboutabl, and H. Ali, "Arabic Sign Language (ArSL) recognition system using hmm", International Journal of Advanced Computer Science and Applications, Vol.2, No.11, Nov. 2011.

[17] J. Ravikiran , M. Kavi , M. Suhas, R. Dheeraj, S. Sudheender, V. Nitin and Pujari ,"Finger Detection for Sign Language Recognition", Proceedings of the International MultiConference of Engineers and Computer Scientists, Vol.I, Mar. 2009.

[18] T. Shanableh, K. Assaleh, and M. Al-Rousan, "Spatio-temporal feature-extraction techniques for isolated gesture recognition in Arabic sign language", IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), Vol.37, No.3, pages641-650, June 2007.

[19] A. Ahmed and S. Aly, "Appearance-based arabic sign language recognition using hidden markov models", International conference on engineering and technology (ICET), pages 1-6 , Apr. 2014.

[20] M. Sadek, M. Mikhael, and H. Mansour, "A new approach for designing a smart glove for Arabic Sign Language Recognition system based on the statistical analysis the Sign Language", In 34th National Radio Science Conference, pages 380-388, Mar. 2017.

[21] M. El-Gayyar, A. Ibrahim and M. Wahed, "Translation from Arabic speech to Arabic Sign Language based on cloud computing", Egyptian Informatics Journal, Vol.17, No.3, pages 295-303. Nov. 2016.

[22] N. Ibrahim, M. Selim, and H. Zayed, "An automatic arabic sign language recognition system (ArSLRS)", Journal of King Saud University-Computer and Information Sciences,Vol.30, Issue 4, pages 470-477, Oct. 2018.

[23] M. Kamruzzaman, "Arabic sign language recognition and generating Arabic speech using convolutional neural network", Wireless Communications and Mobile Computing, Vol.2020, Article ID 3685614, May. 2020.

[24] Y. Saleh and G. Issa, "Arabic Sign Language Recognition through Deep Neural Networks Fine-Tuning", International Journal of Online and Biomedical Engineering, Vol.16, No.5, pages 71-83, 2020.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.

[26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image Recognition", arXiv Organization, Cornell university, arXiv:1409.1556, Apr. 2015.

[27] S. Hayani, M. Benaddy, O. El Meslouhi, and M. Kardouchi, "Arab Sign language recognition with convolutional neural networks", International Conference of Computer Science and Renewable Energies (ICCSRE), pages1- 4, Jul. 2019.

[28] Y. LeCun, L. Bottoux, Y. Bengio and P. Haffner, "Gradient-Based Learning Applied to Document Recognition", In Proceedings of the IEEE, vol.86, no.11, pages 2278-2324, Nov.1998.

[29] G. Latif, N. Mohammad, R. AlKhalaf, R. AlKhalaf, J. Alghazo and M. Khan, "An Automatic Arabic Sign Language Recognition System based on Deep CNN: An Assistive System for the Deaf and Hard of Hearing", International Journal of Computing and Digital Systems, Vol.9, No.4, pages 715-724, Jul. 2020.

[30] H. Luqman and S. Mahmoud, "Transform-based Arabic sign language recognition", Procedia Computer Science, Vol.117, pages 2-9, 2017.

[31] M. Mohandes, S. Aliyu and M. Deriche, "Arabic sign language recognition using the leap motion controller", IEEE 23rd International Symposium on Industrial Electronics, pages 960-965, Jun. 2014.

[32] "Leap Motion Controller and Touchless Technology", Dartmouth Business Journal, [Online]. Available: http://dartmouthbusiness journal.com/2013/08/the-leap-motion-controller-and-touchless-technology. [Accessed: 15-Mar-2021].

[33] A. Elons, M. Ahmed, H. Shedid and M. Tolba, "Arabic sign language recognition using leap motion sensor" 9th International Conference on Computer Engineering & Systems (ICCES) , pages 368-373, Dec. 2014.

[34] N. El-Bendary, H. Zawbaa, M. Daoud, A. Hassanien and K. Nakamatsu, K. "Arslat: Arabic sign language alphabets translator", International conference on computer information systems and industrial management applications (CISIM), pages590-595, Oct. 2010.

[35] M. Packianather and P. Drake, "Comparison of Neural and Minimum Distance Classifiers in Wood Veneer Defect Identification", The Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, Sage Publications, Vol.219, No.11, pages 831-841, Nov. 2005.

[36] M. Almasre and H. Al-Nuaim, "A real-time letter recognition model for arabic sign language using kinect and leap motion controller v2", International Journal of Advanced Engineering, Management and Science, Vol.2, No.5, pages 239469, May. 2016.

[37] M. Dominio and P. Zanuttigh, "Hand gesture recognition with jointly calibrated Leap Motion and depth sensor", Multimedia Tools and Applications journal, Vol. 75, pages 14991–15015, Nov. 2016.

[38] J. Han, L. Shao, D. Xu and J. Shotton, "Enhanced computer vision with Microsoft Kinect Sensor: A review", IEEE Transactions on Cybernetics, Vol.43, No.5, pages 1318-1334, Oct.2013.

[39] M. Maraqa and R. Abu-Zaiter, "Recognition of Arabic Sign Language (ArSL) using recurrent neural networks", First International Conference on the Applications of Digital Information and Web Technologies (ICADIWT), Vol.4, No.1, pages478-481, Aug 2012.

[40] O. Al-Jarrah, and A. Halawani, "Recognition of gestures in Arabic sign language using neuro-fuzzy systems" Artificial Intelligence, Vol.133 No.1-2, pages 117-138, Dec. 2001.

[41] J. Jang, "ANFIS: Adaptive-Network-Based Fuzzy Inference System", IEEE Transactions on Systems Man Cybernetics, Vol.23, No.3, pages 665–685, Jun.1993.

[42] S. Aly and W. Aly, "DeepArSLR: A novel signer-independent deep learning framework for isolated arabic sign language gestures recognition", IEEE Access, Vol.8, pages 83199-83212, Apr. 2020.

[43] C.Wang, C. Chiu and J. Chang, "EfficientNet-eLite: Extremely Lightweight and Efficient CNN Models for Edge Devices by Network Candidate Search", arXiv Organization, Cornell University, Sep. 2020.

[44] T. Mingxing and V. Quoc,"EfficientNet: Rethinking model scaling for convolutional neural networks", In ICML, Vol.97, pages 6105–6114, 2019.

[45] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. “Rethinking the inception architecture for computer vision”. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2818–2826, 2016.

[46] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In: Advances in Neural Information Processing Systems 32 [Internet]. Curran Associates, Inc. pages 8024–35, 2019.

[47] RangiLyu, Pytorch implementation of Google's EfficientNet-lite. Provide imagenet pre-train models, source code [Source code]. https://github.com/RangiLyu/EfficientNet-Lite, 2020.

[48] R. Alawwad, O. Bachir and M. Ismail, Arabic Sign Language Recognition using Faster R-CNN, Proceeding of the International Journal of Advanced Computer Science and Application (IJACSA), Vol.12, No.3, 2021.

[49] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al. "An image is worth 16x16 words: Transformers for image recognition at scale", arXiv Organization, Cornell university, arXiv: arXiv:2010.11929v2, In ICLR, Jun.2021.

Review of Industry Workpiece Classification and Defect Detection using Deep Learning

Changxing Chen, Azween Abdullah, S.H. Kok, D.T.K. Tien
School of Computer Science and Engineering, Taylor's University, Subang Jaya, Malaysia

***Abstract*—Object detection and classification denotes one of the most extensively-utilized machine vision applications given the high requirements put forward for object classification and defect detection with the rise of object recognition scenes. Notwithstanding, conventional image recognition processing technology encounters specific drawbacks. Its benefits and limitations were duly compared upon selecting several typical conventional image recognition techniques. Resultantly, such recognition approaches required multiple manual participation elements and extensive manpower with restricted object identification. As a branch of machine learning, deep learning has attained more optimal results in the image recognition discipline. In the classification and defect detection of industrial workpieces, over 70 literature reviews of deep learning algorithms across multiple application scenarios for classical algorithm model and network structure assessment based on the deep learning theory. Relevant network model performance was compared and analyzed based on network intricacies parallel to natural image classification. Six research gaps were found based on the reviewed algorithm pros and cons. The corresponding six research proposal in workpiece image classification was highlighted with prospects on the workpiece image classification and defect detection direction development. It provides an empirical solution for the selection of workpiece classification and defect detection deep learning model in the future.**

Keywords—Convolutional neural network; image processing; image recognition; defect detection; deep learning

I. Introduction

In line with the proposal of artificial intelligence [1], optimal computational intelligence performance in mathematical theory and computing power enriches the artificial intelligence theoretical framework and catalyzes artificial intelligence development. Perceivably, artificial intelligence constitutes a part of computer science following the emergence of industry 4.0. Artificial intelligence could enhance the level of organizational astuteness with substantial implications across multiple sectors Deep learning and artificial intelligence machine learning depict a broad range of application disciplines, such as image recognition, network security, speech recognition, and natural language processing with significant breakthroughs. Various artificial intelligence recognition systems have been consistently developed with distinct functions and forms for economic and social advantages [2]. The fundamentals of artificial intelligence imply machine learning with algorithms. Multiple image recognition techniques require flexible adoption based on distinct application prerequisites to fulfill various image recognition task requirements in the practical application process. Under the computer vision category, image recognition, which simulates human vision using computers or image-based instruments, facilitates computers to comprehend the recognized entities with algorithms to substitute human eye functions. Conventional image recognition method development proved relatively slow pre-artificial intelligence development. Such recognition techniques were previously based on the object feature descriptor for image recognition and matching with limited discussions on conventional image recognition approaches [3]. The deep learning theory was derived from the conventional neural network under the deep neural network. This theory has eventually become the mainstream of image recognition methods with a distinct object recognition concept simultaneously, it is widely used in pattern recognition [4, 5], image recognition has made breakthroughs upon introducing deep learning into the image processing field while resolving multiple problems that could not be managed by conventional approaches.

Workpiece surface defect is one of the most important factors affecting the product quality of mechanical workpiece. The traditional manual visual inspection method is easy to be affected by manual experience and subjective factors, which lead to inaccurate test results and cannot meet the current inspection requirements and the on-line production requirements of automatic production line. Machine vision inspection has the advantages of high automation, high recognition rate and non-contact measurement. It has gradually become the mainstream method and development trend of surface defect detection. According to the current inspection and classification requirements of workpiece manufacturers, combined with the actual situation of the industrial site, starting from solving the actual impact, workpiece classification and defect recognition are carried out through machine vision. Machine vision classification and detection algorithm can solve the problems of many types and large quantities of workpieces.

A substantial number of factory workpieces (common components in industrial manufacturing) are extensively employed in industrial production. The prerequisites for workpiece recognition speed and accuracy continue rising as opposed to manual workpiece classification strategies with low efficiency and accuracy. In this vein, classification detection denotes high subjectivity. The recent emergence of artificial intelligence technology and computer vision and its application has been extensively employed in industrial sites to catalyze industrial parts classification development. This study summarized the common conventional image recognition approaches, presented specific common image recognition techniques under deep learning, and compared the method performance and effects on other applications.

II. Importance of Industry Workpiece Classification

Workpiece denotes a product manufacturing process component where the machining object in machining or generation (a single part or combination of specific ones) is assembled. The advent of artificial intelligence includes novel application prerequisites for the factory flow production mode given the perpetual improvement of labor cost on the industrial site and highly stringent product quality requirements. Automatic workpiece assembly is highly significant post-artificial intelligence development as the conventional artificial assembly line production mode failed to complement advanced industrial production. As such, industrial automation must be established to optimize manufacturing industry competitiveness. Accelerated automation transformation and the optimization of conventional sectors remain as one of the fundamental points to catalyze industrial development. The artificial intelligence-industrial site integration is inextricably linked for high productivity, novel changes, and industrial development possibilities. Industrial assembly is highly essential in the entire production process as each industrial site requires distinct parts. On another note, a sophisticated industrial production line encompasses workpiece classification and detection. Notably, the robot arm completes the assembly, sorting, and other relevant tasks involving various industrial parts post-classification and detection. Some industrial part sorting proves unsuitable for workers to sort and detect due to industrial site risks when the application scenarios are sufficiently enriched as follows: the monitoring state in the parts-sorting process, sorting process control, and workpiece classification emergency treatment in the industrial site. High requirements are reflected for industrial part detection accuracy and equipment process control stability.

The conventional manual workpiece sorting approach depends on manual operation for parts classification realization. This technique requires high worker's proficiency with substantial product quality implications. Specifically, the equipment of refined workpieces hampers production sorting. As product sorting and detection period in the entire production cycle proves time-consuming, production optimization depends on whether the automatic product line sorting process could be actualized. The means of scientifically controlling the industrial site workpiece classification denotes a complexity that must be regarded and resolved by relevant personnel in encouraging continuous industrial development through automation and intelligence. Machine vision systems, with image processing as one of the pertinent technologies, are increasingly implemented to resolve classification issues. It is deemed necessary to recognize workpieces for target workpiece classification realization. Image recognition is a processing application technology in deep learning and a fundamental task in computer vision. The diversified industrial parts demand substantially challenges the manufacturer's production and classification level given the adverse environment and intricate background within the industrial field. It is considered challenging to recognize the image and resolve the problem with conventional image feature selection based on interference factors: light and workpiece placement background. The industrial workpiece images to be classified are typically complex and ambiguous in practical production and application, thus rendering it intricate to structure an appropriate workpiece image classification approach. Observably, image classification denotes one of the difficult problems to be resolved in image classification and detection tasks.

III. General Open Source Workpiece Defect Datasets

Image datasets to highlight workpiece defects and classification remain lacking to date given the novelty of image defect detection studies. Current online public datasets generally constitute daily necessities, faces, and animals. Most recognition-oriented publications are performed on conventional image classification datasets [6]. Conventional image processing algorithms are typically incorporated into traditional surface defects and classification techniques. Artificial design features and classifiers are commonly implemented compared to the clear classification in computer vision. Specific datasets were adopted to complement the deep learning neural network training. The image formats typically encompass JPG and BMP and JPEG and RGB. Specific datasets in industrial disciplines, such as polished workpieces and customized CNC lathe workpieces do not possess public datasets. The lack of corresponding image training sets would inevitably restrict the promotion of deep learning applications in workpiece recognition. Data incongruence and dataset annotation need to be resolved despite the presence of constructed workpiece dataset pictures.

Current industrial dataset usage could accelerate the deep learning algorithm model development while public, reliable, and open-source industrial datasets could compare distinct deep learning detection algorithms. This section briefly elaborates extensively-utilized industrial datasets following the industrial workpiece classification strategy and the broad industrial dataset application for defect detection as shown in Table I.

TABLE I. General Open Source Workpiece Defect Dataset

Dataset	**Literature**	**Type Quantity**	**Image Size**	**Number of Images**
NEU surface defect database	[7]	6	200*200	1800
Severstal steel defect dataset	[8]	4	1600*256	12568
DAGM 2007 dataset	[9]	10	Grayscale 8-bit PNG format	11500
Kolektor surface defect dataset	[10]	50	width: 500 px height: from 1240 to 1270 px	399
Rail surface defect dataset	[11]	2	variable size	195

The dataset derived from the NEU surface defect database was gathered and generated by several Northeastern University teachers. Six surface defect types were demonstrated with each type entailing 300 picture samples with a total of 1800 gray-scale pictures provided through the bounding box. The picture size was 200*200 pixels. The datasets implied rolling scale

(RS), cracking (CR), pitting surface (PS), plaque (PA), inclusion (in), and scratch (SC), As shown in Fig. 1. Multiple defect types inevitably appeared on the metal workpiece surface during the production process [12]. The dataset was employed to train and classify the surface defect deep learning algorithm [13]. Notably, the artificial feature extraction approach could be integrated with the deep learning algorithm for optimal workpiece classification accuracy amidst insufficient samples.

The Severstal steel defect dataset provided by Severstal steel entails four strip steel surface defect types to locate and classify surface defects on steel plates. A total of 12568 training and 5506 test sets were identified with an image size of 1600*256 [14]. This dataset was analyzed with a per-pixel basis evaluation [15] with the potential as a high-quality defect detection baseline. As such, the deep learning model demonstrated better generalization and higher prediction accuracy in forecasting steel plate surface defects with the Severstal dataset [16]. The deep learning algorithm trained the data and optimized the defect detection learning model for gap identification compared to the other algorithms. In this vein, the deep learning algorithm model was structured while enhancing detection accuracy.

The DAGM 2007 dataset encompasses 10 defect image types with each containing 575 training images and 575 test images: training and test sub-datasets with the same size and a distinct number of label images, as shown in Fig. 2. Every picture encompassed the images saved in grayscale 8-bit PNG format for weak supervised industrial optical detection learning and training. The variance between dataset images proved to be minimal. The recognition algorithm model requirements are considerably high in accurately classifying the defects despite the presence of label files [9]. The accuracy and speed assessment of the fabric defect detection model algorithm using the dataset [17] catalyzed the elevation from the low-resolution feature map to the high-resolution fusion feature counterpart with iterations for optimal prediction outcomes.

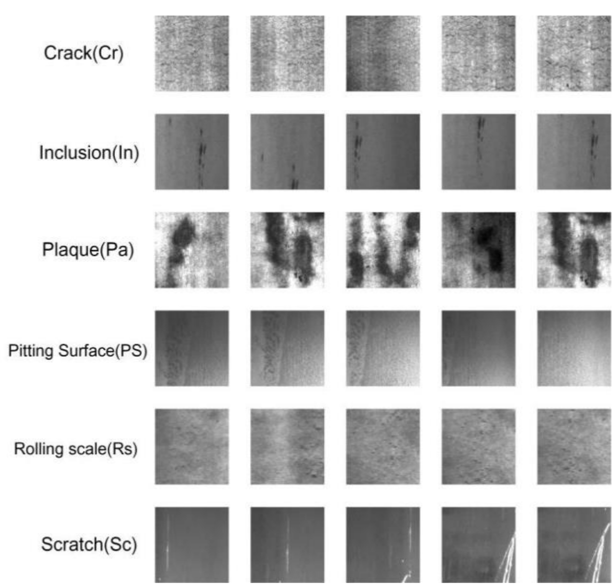


Fig. 1. The Six NEU Types of Surface Defect Images.

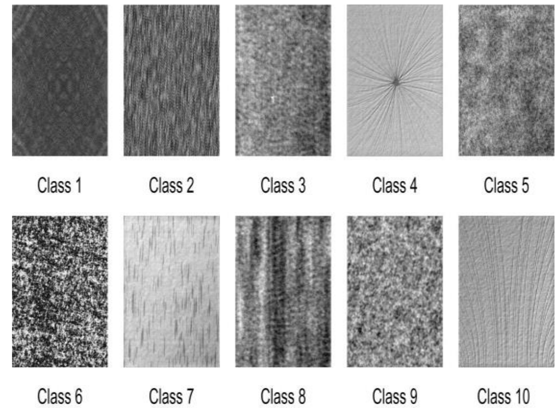


Fig. 2. DAGM Dataset:10 Types of Surface Defect Images.

The Kolektor surface defect dataset primarily gathered defect electronic commutator images, which were subsequently collected under uniform illumination. Labeled pictures were also provided, as shown in Fig. 3. The variance between the pictures proved to be minimal with only one defect picture in every defect type, such as a workpiece with small damages or cracks that are challenging to identify with human eyes. Only 52 defect images were visible in the entire dataset [10], thus significantly facilitating workpiece defect detection tasks.

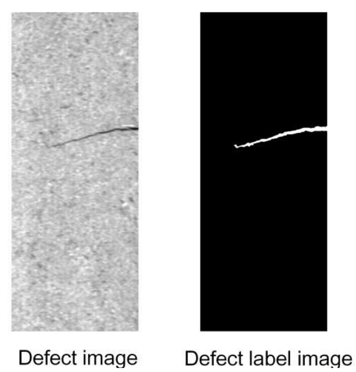


Fig. 3. Kolektor Defect Picture and Labeled Defect Picture.

The rail surface defect dataset implies the train track defects (train track surface crack image) marked by track surface inspection experts. The dataset encompassed 195 challenging images with every image entailing at least one defect and complex and noisy background [18]. Test and prediction datasets could be offered for the deep learning algorithm model to completely detect complex background information defects.

The detection and classification performance of the deep learning algorithm model correlates to image quality, which would then impact performance indicators involving model classification accuracy. Sample incongruence would also influence the classification outcomes with a substantial variance between training and test image quality [12]. Perceivably, the model prediction impact was associated with the model itself and the dataset [13]. More high-quality image data samples could facilitate optimal deep learning model algorithm development for high classification and detection performance.

IV. Traditional Application Methods of Image Recognition

Image recognition denotes computerized image processing, analysis, and understanding to determine multiple target and object types. This approach implies a practical deep learning algorithm application [19], such as online workpiece recognition for grinding burn and wheel wear following a self-clustering neural network [20].Feature matching was primarily utilized for workpiece object recognition in the preliminary stage. Computer key visual features [21] were employed for high detection rates. Meanwhile, Salve et al. recommended a means of shape measurement for object recognition [22]. Dalal et al. utilized the histogram of gradient (HOG) descriptor to compute each stage impact on its performance [23]. Effective gray and rotation invariant texture classification techniques were also incorporated in the early stage under the local binary mode [24]. Tuzel et al.'s research integrated region descriptors with target detection and texture classification [25]. As an astute classification system under machine vision, it essentially classifies the peeled open heart fruit core and shell [26].

K. Xia et al. structured a workpiece sorting system in line with a machine vision industrial robot to complete the sorting operation and fulfill subsequent requirements using image edge detection [27]. Some of the applications depicted in bottleneck identification, which varied from current intuitive approaches, structured a bottleneck identification model following the shortest completion delay time for the overflow load computation of every machine to fulfill each workpiece delivery and optimally determine bottlenecks [28]. Y. Guan et al. employed the affine scale invariant feature transformation (a-sift) technique to identify the rough matching feature points between the assessed and planned workpiece towards workpiece identification by making the identified affine change [29]. Hu's invariant moment was implemented to complement the extracted contour with the target counterpart within the template image for target workpiece identification [30].

Regarding workpiece detection and recognition, conventional approaches typically require manual feature selection and extraction to outline the features as vectors and utilize the similarity measurement function to match the (i) workpiece feature vectors to be identified and (ii) template workpiece [31]. The advent of image recognition remains stunted given its inapplicability in big-scale industries following the low efficiency of conventional sliding window approaches and feature robustness. Table II compares six typical methods and their subsequent categories.

Although conventional image recognition techniques primarily outline objects with artificially-designed features, it is deemed impossible to manually extract rich image feature information from objects with intricate feature designs, thus challenging the recognition problem. As such, a data-driven approach (convolutional neural learning network) proves necessary for image feature data comprehension and processing. The image classification approach with a convolutional neural network could derive the target feature value from the (i) image that is challenging in manual feature extraction or the (ii) image dataset encompassing significant noises compared to the traditional image recognition counterpart. For example, the workpiece dataset in the industrial field demonstrated good robustness to the training and recognition image upon deriving this eigenvalue with the convolution neural learning network. The extracted feature sequence was simultaneously conveyed to the deep neural learning network, which could further elicit the fuzzy features in the image convolution features and forecast the labeled workpiece image. Workpiece image recognition under the convolutional neural network could integrate the two-step workpiece detection and recognition into one, efficiently determine the novel workpiece information encompassed in the image, and save model space and computation with vital and practical significance for project implementation. Convolutional neural network, a deep neural network with a convolution structure, has recently been incorporated into multiple image recognition scenarios. This network inputs the original image into the network. Every network node conveys the image data post-data pre-processing and outputs the probability distribution on the category label with layer-by-layer weight iterative update and computation.

TABLE II. Traditional Methods of Typical Image Recognition

Methods	Reference	Recognition Object	Advantages	Disadvantages
SVM classifier	[32]	tomato	Detect with the least number of features	Extensive color analysis was required
Moment invariant method for edge detection	[33]	workpiece	Able to measure the roughness of a rotating workpiece	Specific measurement environments were required
Template matching method	[34]	workpiece	Simple, practical, and fast recognition speed	The recognition effect was deemed poor for the workpiece with ambiguous feature points
Feature extraction classification	[35]	workpiece	Effectively resolve the problems of unclear features and inaccurate recognition in a complex environment	The recognition robustness was deemed poor without fulfilling the universality of recognition
ROI extraction method	[36]	Air-bearing workpiece	A block matrix was developed using reference pixels and tolerances	Reference pixels were required
Method based on Zernike Moments	[37]	Complex workpiece	Insightful to detect ignored loading attachments and inaccurate assembly positions	The identified object needed to be located

V. IMAGE RECOGNITION METHOD BASED ON DEEP LEARNING

Deep learning constitutes a subclass of machine learning in traditional techniques. Object image feature extraction heavily relies on the manually-designed feature extractor, which requires expert designer knowledge to conduct intricate parameter adjustment process experiments in the model. Notably, the developed model could only determine objects in a particular environment with low generalization and robustness. The number of image feature parameters permitted in the feature extractor design is restricted following the developers' manual adjustment of model parameters. As a branch of artificial intelligence, deep learning neural network reflected higher adaptability with the advent of artificial intelligence as opposed to conventional machine vision techniques This network is deemed more extensive in the universality of article recognition given that the deep learning algorithm primarily entails data-driven image feature extraction for a deeper, more efficient, and accurate representation of the image dataset using the image learning of big samples compared to the conventional method. A series of image recognition techniques under deep learning could attain highly precise and optimal recognition to resolve multiple intricate image recognition scenarios. This section emphasizes four classical segmentation approaches based on deep learning: AlexNet [38-40], Yolo [41-43], VGg net [44-46], and ResNet [47-49].

A. AlexNet

Hinton's and Alex Krizhevsky's revolutionary AlexNet neural network (AlexNet) algorithm [38-40] championed the 2012 Imagenet competition. Specifically, Imagenet entails a large image recognition database encompassing marked pictures. AlexNet focuses on the full connection layer function with a total of eight layers: five convolutions and three full-connection. In a three-channel color map with 227 pixels in length and width (227*227*3), the image is incorporated into the first layer to be convoluted into 11*11*3. Every convolution kernel generates a novel pixel while all the convolution kernels subsequently slide through the 227*227*3 pixel picture with a stripe of four. Following the convolution output layer resolution computation, the convolution pixel layer data is duly computed with a convolution output of 55*55*96 in the first layer. The total convolution parameters of the first layer imply 35K as only the convolution kernel in the convolution layer entails neural network parameters post-calculation. Meanwhile, the second-layer characteristic map is transmitted to the third counterpart until the seventh-layer output data is fully connected with 1000 neurons from the eighth counterpart. The outcome was generated through softmax, which was utilized as a 1000 input image category for the classification score with the following attributes: (1) AlexNet algorithm converted the traditional neuron activation function $f(x)=(1+e^{-x})$ to $f(x)=max(0;x)$ with a rectified linear unit (relu) as an activation function that was extensively utilized in artificial neural network. The typical four-layer network with relus and tanh as an activation function attains the faster convergence speed effect involving relu in the CIFAR-10s experiment dataset compared to the conventional tanh activation function; (2) AlexNet utilized two techniques to resolve over-fitting issues: data enhancement and dropout. The original picture was cropped to be employed as network input in data enhancement while dropout was utilized to deter over-fitting and promote effective fusing. Regardless, the network model computing cost is exorbitant despite the feasible computation following the use of a graphics processor (GPU) in the training process.

B. YOLO

The You Only Look Once (Yolo) revolutionary neural network algorithm [41-43] resolves object detection (a regression problem) to avoid several reiterated prediction works and complete the input from the original image to the output of the image category following a separate end-to-end network. Yolo entails specific prerequisites to incorporate the image input size into the network, scale the image size to the specified size, classify the picture into S × S grid, and make predictions in every small grid. Based on Fig. 4, the category probability forecasted by each grid and the confidence predicted by each box were multiplied until the score correlated to every box and category. The non-maximum suppression approach was then utilized to derive the classification outcomes. Essentially, Yolo is deemed beneficial as it disregards the extraction process of region proposal and rapidly identifies objects with minimal background error detection rates and inaccurate background knowledge. This algorithm implies high generalization, which is unlikely to crash when incorporated into fields or unforeseen inputs. Notwithstanding, the error-prone S × S grid at the frame regression stage leads to inaccurate object positioning. A large missed detection rate is identified in the presence of multiple small targets in a network. The subsequent Yolo version continues to rectify such complexities. Specifically, Yolo V3 elevates detection performance, particularly in the multi-scale fusion approach, to resolve low detection performance post-defect optimization. On another note, a cross-layer connection is presented to optimize small target detection performance.

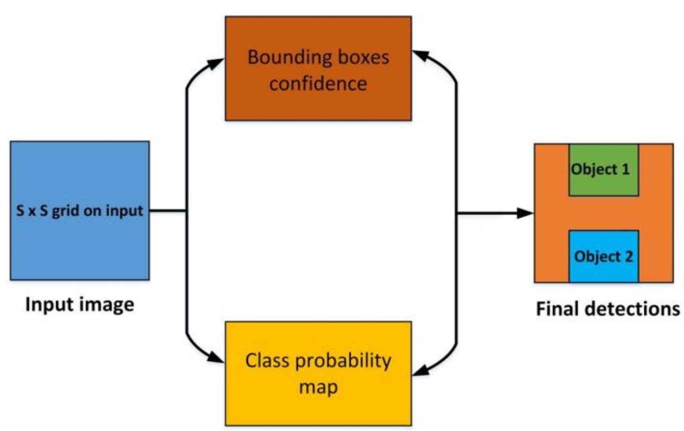


Fig. 4. Yolo Diagram Algorithm Process.

C. VGG Net

The VGG net [44-46] model is characterized by a substantial number of layers, including multiple network layers ranging in depth between 11 and 19. The deep learning-model performance correlation is examined to enhance the overall recognition performance by improving the network layer depth. This model aims to transform the convolution of the larger core layer into multiple smaller-layered convolution cores. Vggnet-16 and Vggnet-19 are extensively employed to render the entire network to be highly effective. The VGG also denotes a five-layer convolution and two full-connection layers for image feature extraction and one full connection layer for feature classification akin to the AlexNet framework. Fig. 5 illustrates the Vggnet-16 network structure diagram. The convolution layer kernel in the Vgg net structure is 3*3. Three groups of 3 *3 convolution layers connected with a 1*7*7 kernel were employed with the same effect. In terms of model benefits, the number of parameters is duly reduced. The original parameter (C*7*7) was transformed into 3*C*3*3 for a convolution layer with C kernels. Despite the presence of more parameters and deeper levels, VGG requires lesser iterations to initiate convergence given the depth and small filter size function as the post rule while the pre-initialization operation is performed on some layers. Such advantages could increase the non-linear correction layer, mitigate gradient disappearance and over-fitting issues, and optimize the model training speed. The network structure attributes simultaneously facilitated it to regulate the number of parameters while eliciting more image features to prevent over-computation and structure intricacy.

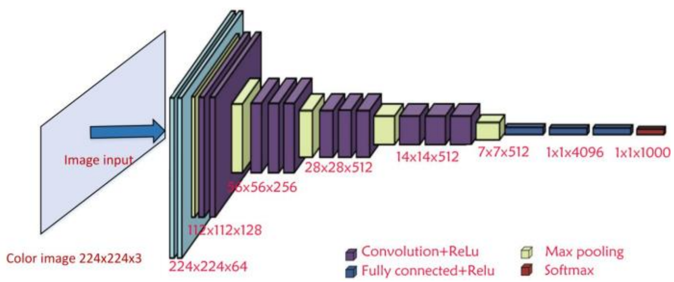


Fig. 5. Network Structure of VGG16 Adapted from [45].

D. *ResNet*

The proposed ResNet [47-49] algorithm model resolves the deepening of network layers despite the initially-enhanced accuracy. This precision would worsen if the number of network layers continues to increase. In other words, the network model could alleviate the degradation problem in network training. Regarding the conventional deep learning algorithm, the layers to which the gradient could not be conveyed are not trained when the number of network layer increases. Thus, the effect is not as robust as the shallow network with adequate layers as the error rate would rise with the increase in layers. The ResNet algorithm model puts forth the residual module in resolving the degradation problem. Based on the notion underpinning this method, the network at layer N is derived from the network at layer N-1 with conversion. It is connected to the upper-layer network for gradient propagation, which subsequently resolves the gradient disappearance caused by the neural network passing through depth. Thus, the residual structure is presented. Following Fig. 6, the output layer H (x) = F (x) is changed to H (x) = F(x) + x where the network loss function f (x, w) is extended to the multi-layer neural network using the back propagation gradient value formula. The front-layer network gradient becomes smaller with the increased number of layers 'n' in the neural network and the return of errors. Thus, the gradient would not disappear even with multiple network layers. On another note, ResNet could develop the residual module in the form of a small kernel with the other layers utilizing full convolution excluding the full connection layer for classification, which could substantially optimize the calculation speed. This residual block structure method is employed for reference as ResNet entails multiple algorithm types with a network depth of 50, 101, and 152. The ResNet model performance significantly varies with distinct sizes. Overall, this model needs to be structured based on the actual application context.

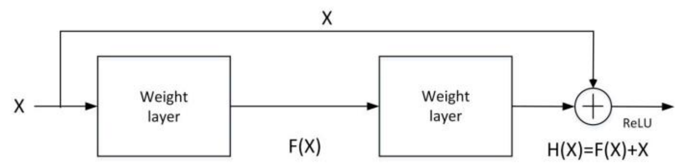


Fig. 6. Residual Block Structure.

VI. Performance Comparison

Based on recent research, image recognition technology could flexibly select multiple algorithms for recognition based on various application scenarios in actual recognition tasks regarding deep learning. Some might even need to integrate different recognition methods in obtaining the most optimal recognition accuracy. It is deemed pivotal to develop a set of robust and stable recognition algorithms with high market value and application possibilities following the intricate recognition environment, which is subject to more illumination or other interference. Fundamentally, relevant literature adopts four parameters to compute the following metrics (accuracy, precision, recall, and F1-score) as the evaluation indices of algorithm advantages and disadvantages. As the primary problem to be solved, accuracy could further optimize the empirical depth in the image recognition field with a highly positive effect on the efficient incorporation of multiple technologies and the development of relevant disciplines. This study primarily summarized the deep learning technology application to image recognition. Table III presents (i) the recent comparison of model precision and loss of value in deep learning algorithms, (ii) evaluations on the technological complexities encountered in the workpiece image classification application, and (iii) specific study gaps.

TABLE III. DEEP LEARNING ALGORITHM COMPARISONS

Literature	Model	Object	Accuracy	Loss rate
[50]	LeNet-5	Oil–water two-phase flow	88%	High (loss rate)
	VGG-16	Oil–water two-phase flow	98.3%	Median (loss rate)
	AlexNet	Oil–water two-phase flow	99.3%	Low (loss rate)
[41]	YOLO	Insulator	90%	3.8% (class fault)
	Fast R-CNN	Insulator	84%	4.1% (class fault)
[48]	ResNet-18	Apple leaf	99%	8.9%
	ResNet-34		97%	1.23%
	VGG		89%	2.43%
[51]	VGG-16	Chest X-ray images	97%	NA
	ResNet		86.7%	
	ResNet-50		98%	
	AlexNet		96.5%	
	ResNet-50+SVM		95.38%	
[40]	Alexnet	Power equipment	83.55%	NA
[52]	YOLO	Leucocyte	93.7%	NA
[53]	YOLO	Resistors and capacitors	92.57%	0.87% (error rate)
[47]	ResNet50	PatchCamelyon	98.8%	98.5% (AUC-ROC score)
	Vgg-16		95.9%	95.2% (AUC-ROC score)
	Vgg-19		97%	96.1% (AUC-ROC score)
[54]	ResNet-50	Bacterial image	99.9%	0.02% (loss rate)
	ResNet-34		99.3%	0.04% (loss rate)
[55]	AlexNet	Wireless capsule endoscopy	95%	NA
[56]	AlexNet	MNIST	98%	NA
[57]	GoogLeNet	Colour images	NA	90.31% (FMI)
	Alexnet		NA	87.69% (FMI)
	VGG16		NA	89.78% (FMI)
[58]	Deep Transfer Learning	Tongue images	95.92%	NA
[59]	LeNet	Microorganism image	98.66%	NA
[60]	transfer learning approach	Flower image	98.6%	98% (recall)
[61]	VGG16	Fundus images	91.3%	NA
[62]	RS-oriented error-tolerant deep learning	Noisy remote sensing image	95.9%	NA
[63]	deep learning and image recognition	Brake pad contour image	98%	NA
[64]	self-supervised super sample decomposition for transfer learning	COVID-19	99.8%	99.7% (sensitivity)
[65]	Meta-learning	Classical image	92.48%	NA

Notes: NA - Not Available

The accuracy rate in Table III denotes the percentage of the number of correctly predicted samples in the total number of samples or the proportion of precisely-predicted outcomes in the total number of samples. The formula of accuracy is shown in (1), essentially, TP implies a positive prediction with an actual prediction that is true. The TN denotes a negative and true forecast. The FP demonstrates a positive forecast with an actual prediction that is false. The FN denotes the forecast to be negative with an actual prediction that is false. The FP and FN reflect prediction error values, which means that the pre-determined target is not discovered. Several relevant detection indicators (accuracy) imply the proportion of accurate forecasting outcomes in all positive predictions. Recall rate denotes the proportion of accurately-predicted findings in all positive occurrences. In deep learning classification and detection tasks, the precision of appropriate techniques is typically assessed based on statistical findings. High TP and TN reflect high precision and optimal detection impacts of the deep learning model.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

Based on Table III, the deep learning algorithm model accuracy is fundamentally between 80% and 99%. The designed model evaluation accuracy is significantly enhanced compared to conventional techniques, thus implying the application effect of the deep learning model to be more ideal. Despite the diversification of model evaluation methods in theoretical and practical research, such studies remain considerably scattered given the primary utilization of the accuracy evaluation index for model assessment. Accuracy implies the accurately-forecasted image proportion in objective evaluation and classification. Although the high precision value in outcome detection or classification assumably reflects a high recall value, this evaluation index might prove contradictory in some cases. For example, an accurately-forecasted outcome in prediction would demonstrate a 100% precision value albeit with a significantly low recall value. Meanwhile, all the outcomes returned with a recall of 100% would denote substantially low precision. Hence, most studies do not employ such evaluation predictors for model assessment.

Despite there being no unified recognition technique for small target workpiece identification in workpiece recognition, relevant researchers have recommended multiple detection methods under deep learning to resolve the problems associated with small target detection. Notwithstanding, different study objects could ascertain whether precision or recall is high based on the required judgment following relevant research.

Detection approaches have undergone continuous optimization to manage multiple classification and detection problems involving deep learning image datasets and fulfill the application requirements in actual scenarios. The high-dimensional semantic features of image data could be elicited by convolution and non-linear layers, which is much better than traditional detection performance approaches. Image is more vulnerable to noise, thus significantly increasing workpiece classification and interpretation complexities in the industrial field. It is deemed pivotal to determine how to fully utilize the evaluation index information of in-depth learning in serving the workpiece identification and classification application requirements and alleviate data processing intricacies for optimal workpiece classification, detection, and interpretation within the industrial field.

VII. Research Gap

The research gaps are summarized in Table III.

1) The recognition algorithm still requires a specific optimization level, specifically through image feature extraction, to further enhance the image recognition rate following experimental data, computing equipment, and research time issues. Such image features could be conveniently extracted by the convolution filter following [50]. Empirically, the accuracy of the three aforementioned deep learning network algorithms proves relatively high. Simultaneously, a small number of target image misrecognition would not impact the traffic pattern recognition outcomes. Literature [61] could substantially optimize the recognition and classification rate with the image feature extraction approach.

2) The number of deep learning network parameters consumes much time, power, and hardware resources in the actual training process, which complicates the neural network application. The benefits of speed and accuracy in algorithm recognition are highlighted in [41]. Recognition accuracy could be further optimized given the increased number of trained images despite its time consumption and high cost.

3) In terms of computing power limitation, a higher image resolution could enhance the number of extracted image features through the algorithm given the low image resolution in the training dataset. Following [47], an optimized classical depth learning algorithm model could significantly enhance low image resolution performance. Test time augmentation (TTA) was adopted to improve the image while the derived image prediction accuracy data proved better than other depth learning algorithm models.

4) Concerning workpiece classification and recognition algorithms, the newly-proposed algorithm performance could be assessed and examined to accelerate the iterative update speed of the theoretical algorithm. The number and type of database samples failed to fulfill the prerequisites and match the workpiece defects generated by practical applications following the inadequate public datasets of some types based on the current workpiece defect database. Most of the algorithms could not be fairly compared given the absence of an agreed database standard [66]. Despite the presence of extensively utilized workpiece defect datasets, such as NEU [7], UCI [67], and Rail surface [68], the odious industrial setting poses substantial complexities to the workpiece defect image dataset of the actual industrial production line.

5) Deep learning training requires sufficient training datasets in the workpiece image dataset production. It is deemed necessary to ascertain the means of training a model that could precisely detect images through a restricted number of sample datasets. Data optimization served to enhance segmentation accuracy. The deep learning benefits minutely differed across multiple evaluation indicators with high precision. Fundamentally, the evaluation indices were not extremely low.

6) Several deep learning algorithm models might not apply to specific application scenarios given the emergence of more relevant counterparts. It is deemed feasible to develop a novel deep learning algorithm to enhance the impact of workpiece classification and recognition. For example, [69] integrated the computational advantages of a 2D FCN network and the ability to resolve 3D spatial consistency without influencing segmentation accuracy. Given the palpable limitations, the resolution of recognized pictures and speed calculation require improvement as network calculation precision serves to increase with the number of computations. This accuracy could be improved [40] through integrations with other methods albeit with low precision in some image recognition scenarios. Despite the improved effects, some room for improvement is still available. Regarding parameter optimization, [65] proposed that tests could be performed on non-trained tasks. Meanwhile, classification could be realized even with minimal training: a direction worthy of the effort.

VIII. Research Proposal

1) Workpiece image classification algorithms could strive to fuse image feature extraction into one step and enhance recognition accuracy. Regarding the recognition target limitation, the extracted image required a strong expression while the recognition and classification rate proved relatively low. This situation adversely influenced the recognition of intricate or unclear images given its complexity in feature image extraction. The algorithm model could be learned in depth from the noise dataset with the possible identification of low feature target images in line with [62], thus broadening the deep learning application range. Specifically, [63] employed the edge feature extraction approach to determine the internal and external features of the recognition target image contour. The image features were distinguished post-feature point extraction. The deep learning algorithm model was subsequently presented for feature training towards high recognition accuracy. Meanwhile, [56, 57] utilized the deep learning method to fuse image features and resolve the target image feature ambiguity or relatively indistinguishable recognition rates.

2) Given the presence of issues involving extensive training and prediction periods, future workpiece image classification algorithm studies could consider how to mitigate network model redundancy, optimize the number of network layers, and shorten the computation time while simultaneously ensuring recognition accuracy to some extent. Based on [48], shallow networks could also reflect optimal recognition accuracy and low error rates with even better recognition impacts than deep networks. Palpable target detection errors or reiterated identification in employing the same algorithm for target detection and limited training time resembles the drawbacks highlighted in conventional neural networks following [52, 53].

3) Deep learning algorithm could further enhance recognition accuracy with optimal hardware and image acquisition.AS affirmed by [59], low resolution and multiple datasets could be utilized for the deep learning algorithm to realize classification. The usage of more epochs improved training accuracy while surface accuracy could be optimized by the number of iterations. Regarding the disadvantage, the loss value would be too high in the iterative process. The same deep learning method demonstrated minimal variance in recognition accuracy under multiple levels, specifically in small target images, with improved computer algorithms [54].

4) Further enhance network generalization (particularly in restricted datasets), optimize small dataset detection, and integrate the conventional recognition. The means of developing an algorithm with strong applicability require further examination for thousands of object types. Regarding limited sample collection, a more widely-disseminated collection database was established in the public counterpart following [51] and a dataset in line with their actual application scene by gathering six chest X-ray image databases. The gathered database should be relevant to the research. Detection algorithm with convolution neural network to improve microfeature extraction and segmentation capacity and model accuracy in the follow-up. As deep learning technology was employed for dataset pre-processing albeit with relatively ideal experimental effects.

5) It is also rendered possible to optimize moving image recognition, develop a recognition model towards dataset expansion, and enhance the recognition model adaptability to the actual industrial setting in the process of moving image recognition. For example [55] could further determine the range of textures and features, such as color while [58] employed the deep transfer learning technique to assess tongue images and resolve the complexities in gathering adequate marker image samples. The recommended approach implied better classification accuracy. Essentially, [70] outlined the mapping relationship between image classification input feature vector and image category and structured a moving image recognition model.

6) The incorporation of deep learning algorithm benefits into specific fields or different but relevant disciplines or problems has garnered more attention towards completing or improving the learning effect of target fields or tasks. Ensemble learning is a promising and experimentally-proven technology. Based on [60], deep learning approaches significantly influence intricate tasks, such as image feature extraction, segmentation, and semantic classification. Meanwhile, [64] mitigated the network complexity by pruning the redundant depth algorithm model parameters to derive a small and efficient classification model, enhance the runtime reasoning speed of neural network, and elicit the favored classification effect with minimal calculation and computational workload in the future. The incorporation of meta-learning into deep learning denotes a viable method.

IX. Conclusion

This study reviewed the recent development of deep learning in image recognition, emphasized specific deep learning image method types, summarized and compared similar algorithm performance, examined the advantages and disadvantages of every deep learning algorithm based on different application scenarios, and flexibly selected deep learning classification methods to effectively improve the recognition effect. highlighting six research proposals for workpiece image recognition and defect detection. Empirical solutions are provided for the selection of future deep learning models for artifact classification and defect detection.

Many actual influencing factors and complex situations are present in workpiece recognition. In other words, it is challenging to apply the model derived by conventional methods to the actual circumstance. Research on the deep convolution neural network given its prevalence in computer vision tasks has made a significant breakthrough, thus proving the potential of deep learning in image classification. There are also limitations in the research of workpiece recognition, in terms of theoretical experiments on images, there is no research on image acquisition and recognition of workpieces in practical application scenarios, and it is necessary to build recognition systems to carry out research on practical workpiece recognition applications based on multiple factors such as light, angle and placement position.

In the future work, more methods will be reviewed to enhance the generalization ability of the model and improve the practical application ability of the model in the industrial field, which is also the improvement direction proposed by the future in-depth learning research work. At the same time, it can effectively and reducing the dataset is also one of the priorities of the future work.

References

[1] J. McCarthy, M. L. Minsky, N. Rochester, and C. E. Shannon, "A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955," AI magazine, vol. 27, no. 4, pp. 12-12, 2006.

[2] P. TianY, WangX, "Deep Learning Strong Partsfor Pedestrian Detection," presented at the IEEE International Confer-enceon Computer Vision, 2016.

[3] S. I. Krizhevsky A, Hinton G E, "ImageNet classification with deep convolutional neural networks," in Proceedings of the International Conference on Neural Information Processing Systems vol. 1097-1105, ed, 2012.

[4] R. Kaluri and C. Reddy, "An Enhanced Framework for Sign Gesture Recognition using Hidden Markov Model and Adaptive Histogram Technique," International Journal of Intelligent Engineering and Systems, vol. 10, no. 3, pp. 11-19, 2017.

[5] R. Kaluri and C. H. P. Reddy, "Optimized feature extraction for precise sign gesture recognition using Self-Improved Genetic Algorithm," 2018.

[6] S. S. Pandarge and V. R. Ratnaparkhe, "Image Processing based Linear Discriminant and Quadratic Discriminant Classifier for Feature Extraction Models," in 2019 International Conference on Intelligent Computing and Control Systems (ICCS), 15-17 May 2019 2019, pp. 1502-1505, doi: 10.1109/ICCS45141.2019.9065388.

[7] B. Zhao, M. Dai, P. Li, R. Xue, and X. Ma, "Defect Detection Method for Electric Multiple Units Key Components Based on Deep Learning," IEEE Access, vol. 8, pp. 136808-136818, 2020, doi: 10.1109/ACCESS.2020.3009654.

[8] Z. Xinzi, "BSU-Net: A Surface Defect Detection Method Based on Bilaterally Symmetric U-Shaped Network," in 2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE), 25-27 Dec. 2020 2020, pp. 1771-1775, doi: 10.1109/ICMCCE51767.2020.00388.

[9] H. Zhou, B. Jang, Y. Chen, and D. Troendle, "Exploring Faster RCNN for Fabric Defect Detection," in 2020 Third International Conference on Artificial Intelligence for Industries (AI4I), 21-23 Sept. 2020 2020, pp. 52-55, doi: 10.1109/AI4I49448.2020.00018.

[10] Y. Li and J. Li, "An End-to-End Defect Detection Method for Mobile Phone Light Guide Plate via Multitask Learning," IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-13, 2021, doi: 10.1109/TIM.2021.3053987.

[11] D. Zhang, K. Song, Q. Wang, Y. He, X. Wen, and Y. Yan, "Two Deep Learning Networks for Rail Surface Defect Inspection of Limited Samples With Line-Level Label," IEEE Transactions on Industrial Informatics, vol. 17, no. 10, pp. 6731-6741, 2021, doi: 10.1109/TII.2020.3045196.

[12] T. Benbarrad, L. Eloutouate, M. Arioua, F. Elouaai, and M. D. Laanaoui, "Impact of Image Compression on the Performance of Steel Surface Defect Classification with a CNN," Journal of Sensor & Actuator Networks, Article vol. 10, no. 4, pp. 73-73, 2021, doi: 10.3390/jsan10040073.

[13] X. Feng, X. Gao, L. Luo, C. Humbert, and R. B. Navarro, "X-SDD: A New Benchmark for Hot Rolled Steel Strip Surface Defects Detection," Symmetry (20738994), Article vol. 13, no. 4, p. 706, 2021, doi: 10.3390/sym13040706.

[14] J. Božič, D. Tabernik, and D. Skočaj, "End-to-end training of a two-stage neural network for defect detection," in 2020 25th International Conference on Pattern Recognition (ICPR), 10-15 Jan. 2021 2021, pp. 5619-5626, doi: 10.1109/ICPR48806.2021.9412092.

[15] F. Akhyar, C. Y. Lin, G. S. Kathiresan, B. Surianarayanan, and C. Y. Hsu, "Detectors++: The Robust Baseline for a Defect Detection System," in 2021 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), 15-17 Sept. 2021 2021, pp. 1-2, doi: 10.1109/ICCE-TW52618.2021.9603063.

[16] M. Abu, A. Amir, Y. H. Lean, N. A. H. Zahri, and S. A. Azemi, "The Performance Analysis of Transfer Learning for Steel Defect Detection by Using Deep Learning," Journal of Physics: Conference Series, vol. 1755, no. 1, p. 012041, 2021/02/01 2021, doi: 10.1088/1742-6596/1755/1/012041.

[17] H. Dong, K. Song, Y. He, J. Xu, Y. Yan, and Q. Meng, "PGA-Net: Pyramid Feature Fusion and Global Context Attention Network for Automated Surface Defect Detection," IEEE Transactions on Industrial Informatics, vol. 16, no. 12, pp. 7448-7458, 2020, doi: 10.1109/TII.2019.2958826.

[18] M. Niu, K. Song, L. Huang, Q. Wang, Y. Yan, and Q. Meng, "Unsupervised Saliency Detection of Rail Surface Defects Using Stereoscopic Images," IEEE Transactions on Industrial Informatics, vol. 17, no. 3, pp. 2271-2281, 2021, doi: 10.1109/TII.2020.3004397.

[19] X. Caiyun, "Overview of image recognition technology," (in China), Computer knowledge and technology, vol. (10), pp. 2446-2447, 2013,9.

[20] J. Shi, X. Zhang, B. Zhong, and H. Ren, "A study of on-line identification of grinding burn and wheel wear based on self-clustering neural network," in IEEE International Conference on Industrial Technology, 1996.

[21] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, 8-14 Dec. 2001 2001, vol. 1, pp. I-I, doi: 10.1109/CVPR.2001.990517.

[22] S. G. Salve and K. C. Jondhale, "Shape matching and object recognition using shape contexts," in 2010 3rd International Conference on Computer Science and Information Technology, 9-11 July 2010 2010, vol. 9, pp. 471-474, doi: 10.1109/ICCSIT.2010.5565098.

[23] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in IEEE Computer Society Conference on Computer Vision & Pattern Recognition, 2005.

[24] T. Ojala, M. Pietikinen, and T. Menp, "Gray Scale and Rotation Invariant Texture Classification with Local Binary Patterns," European Conference on Computer Vision, 2000.

[25] O. Tuzel, F. Porikli, and P. Meer, "Region Covariance: A Fast Descriptor for Detection and Classification," Springer, Berlin, Heidelberg, 2006.

[26] H. Nouri-Ahmadabadi, M. Omid, S. S. Mohtasebi, and M. S. Firouz, "Design,development and evaluation of an online grading system for peeled pistachios equipped with machine vision technology and support vector machine," Information Processing in Agriculture, vol. 4, no. 4, p. 9, 2017.

[27] X. Ke and Z. Weng, "Workpieces sorting system based on industrial robot of machine vision," in International Conference on Systems & Informatics, 2016.

[28] D. Wen, H. Li, and Z. Aixia, "A bottleneck resource identification method for completing the workpiece based on the shortest delay time," in 2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), 8-10 Aug. 2011 2011, pp. 9-13, doi: 10.1109/AIMSEC.2011.6010180.

[29] Y. Guan, J. Tian, and Y. Tang, "Workpiece Identification and Localization Based on Monocular Vision," in 2019 IEEE 9th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER), 2019.

[30] C. Huang, C. Dan, and X. Tang, "Implementation of Workpiece Recognition and Location Based on Opencv," in 2015 8th International Symposium on Computational Intelligence and Design (ISCID), 2016.

[31] Y. Wei, "Research on recognition and pose estimation methods for low texture workpiece," Changzhou University, [D] 2021.

[32] A. Sdk, B. Se, C. Sb, and B. Bk, "A microcontroller based machine vision approach for tomato grading and sorting using SVM classifier," Microprocessors and Microsystems, vol. 76, 2020.

[33] B. M. Kumar and M. M. Ratnam, "Machine vision method for non-contact measurement of surface roughness of a rotating workpiece," Sensor Review, vol. 35, no. 1, pp. 10-19, 2015.

[34] Z. Liu, P. Wan, L. Ling, L. Chen, and W. Zhou, "Recognition and Grabbing System forWorkpieces Exceeding the Visual Field Based on Machine Vision," Jiqiren/Robot, vol. 40, no. 3, pp. 294-300 and 308, 2018.

[35] X. X. L. Wang H, Zuo Y B, "HL-S workpiece recognition algorithm based on multi-feature fusion," Journal of Electronic Measurement and Instrument, vol. 33(12), pp. 94-99, 2019.

[36] P. Kunakornvong and P. Sooraksa, "Machine Vision for Defect Detection on the Air Bearing Surface," in 2016 International Symposium on Computer, Consumer and Control (IS3C), 4-6 July 2016 2016, pp. 37-40, doi: 10.1109/IS3C.2016.20.

[37] B. Guotang, C. Yuansong, Q. Renchao, and H. Wei, "An Defect Detection Method Based on Zernike Moments for Industrial CT Image," in 2007 8th International Conference on Electronic Measurement and Instruments, 16-18 Aug. 2007 2007, pp. 2-831-2-835, doi: 10.1109/ICEMI.2007.4350809.

[38] Q. Wang, J. Xu, T. Sun, Z. Lv, and G. Zong, "Research on Road Adhesion Condition Identification Based on an Improved ALexNet Model," Journal of Advanced Transportation, Article pp. 1-14, 2021, doi: 10.1155/2021/5531965.

[39] C. Wang, L. Yu, X. Zhu, J. Su, and F. Ma, "Extended ResNet and Label Feature Vector Based Chromosome Classification," IEEE Access, vol. 8, pp. 201098-201108, 2020, doi: 10.1109/ACCESS.2020.3034684.

[40] Y. Guo, Z. Pang, J. Du, F. Jiang, and Q. Hu, "An Improved AlexNet for Power Edge Transmission Line Anomaly Detection," IEEE Access, vol. 8, pp. 97830-97838, 2020, doi: 10.1109/ACCESS.2020.2995910.

[41] S. Wang, L. Niu, and N. Li, "Research on Image Recognition of Insulators Based on YOLO Algorithm," in 2018 International Conference on Power System Technology (POWERCON), 6-8 Nov. 2018 2018, pp. 3871-3874, doi: 10.1109/POWERCON.2018.8602149.

[42] G. Li, Z. Song, and Q. Fu, "A New Method of Image Detection for Small Datasets under the Framework of YOLO Network," in 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 12-14 Oct. 2018 2018, pp. 1031-1035, doi: 10.1109/IAEAC.2018.8577214.

[43] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 27-30 June 2016 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.

[44] M. Ye et al., "A Lightweight Model of VGG-16 for Remote Sensing Image Classification," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 6916-6922, 2021, doi: 10.1109/JSTARS.2021.3090085.

[45] H. Ge, Z. Zhu, R. Liu, and X. Wu, "Application of Multiscale Facial Feature Manifold Learning Based on VGG-16," Journal of Sensors, Article pp. 1-7, 2021, doi: 10.1155/2021/7129800.

[46] M. Rezaee, Y. Zhang, R. Mishra, F. Tong, and H. Tong, "Using a VGG-16 Network for Individual Tree Species Detection with an Object-Based Approach," in 2018 10th IAPR Workshop on Pattern Recognition in Remote Sensing (PRRS), 19-20 Aug. 2018 2018, pp. 1-7, doi: 10.1109/PRRS.2018.8486395.

[47] Z. Zheng, H. Zhang, X. Li, S. Liu, and Y. Teng, "ResNet-Based Model for Cancer Detection," in 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), 15-17 Jan. 2021 2021, pp. 325-328, doi: 10.1109/ICCECE51280.2021.9342346.

[48] X. Li and L. Rai, "Apple Leaf Disease Identification and Classification using ResNet Models," in 2020 IEEE 3rd International Conference on Electronic Information and Communication Technology (ICEICT), 13-15 Nov. 2020 2020, pp. 738-742, doi: 10.1109/ICEICT51264.2020.9334214.

[49] A. Mahajan and S. Chaudhary, "Categorical Image Classification Based On Representational Deep Network (RESNET)," in 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), 12-14 June 2019 2019, pp. 327-330, doi: 10.1109/ICECA.2019.8822133.

[50] D. Meng, Y. Hongyi, C. Xiaoyan, and W. Xinqiang, "Oil-in-Water Two-Phase Flow Pattern Identification From Experimental Snapshots Using Convolutional Neural Network," IEEE Access, article vol. 7, pp. 6219-6225, 01/01/ 2019, doi: 10.1109/ACCESS.2018.2888733.

[51] E. Cortés and S. Sánchez, "Deep Learning Transfer with AlexNet for chest X-ray COVID-19 recognition," IEEE Latin America Transactions, vol. 19, no. 6, pp. 944-951, 2021, doi: 10.1109/TLA.2021.9451239.

[52] Y. C. Yang and W. Chen, "An Improved YOLO Leucocyte Classification and Recognition Method," in 2021 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), 27-28 March 2021 2021, pp. 618-621, doi: 10.1109/ICITBS53129.2021.00157.

[53] J. Huang and Y. Lu, "A Method for Identifying and Classifying Resistors and Capacitors Based on YOLO Network," in 2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP), 19-21 July 2019 2019, pp. 1-5, doi: 10.1109/SIPROCESS.2019.8868519.

[54] T. Shaily and S. Kala, "Bacterial Image Classification Using Convolutional Neural Networks," in 2020 IEEE 17th India Council International Conference (INDICON), 10-13 Dec. 2020 2020, pp. 1-6, doi: 10.1109/INDICON49873.2020.9342356.

[55] S. Sunitha and S. S. Sujatha, "An Improved Bleeding Detection Method for Wireless Capsule Endoscopy (WCE) Images Based on AlexNet," in 2021 3rd International Conference on Signal Processing and Communication (ICPSC), 13-14 May 2021 2021, pp. 11-15, doi: 10.1109/ICSPC51351.2021.9451699.

[56] Y. Lu, "Image Classification Algorithm Based on Improved AlexNet in Cloud Computing Environment," in 2020 IEEE International Conference on Industrial Application of Artificial Intelligence (IAAI), 25-27 Dec. 2020 2020, pp. 250-253, doi: 10.1109/IAAI51705.2020.9332891.

[57] B. Lakshmipriya, N. Pavithra, and D. Saraswathi, "Optimized Convolutional Neural Network based Colour Image Fusion," in 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), 3-4 July 2020 2020, pp. 1-4, doi: 10.1109/ICSCAN49426.2020.9262439.

[58] C. Song, B. Wang, and J. Xu, "Classifying Tongue Images using Deep Transfer Learning," in 2020 5th International Conference on Computational Intelligence and Applications (ICCIA), 19-21 June 2020 2020, pp. 103-107, doi: 10.1109/ICCIA49625.2020.00027.

[59] T. Treebupachatsakul and S. Poomrittigul, "Microorganism Image Recognition based on Deep Learning Application," in 2020 International Conference on Electronics, Information, and Communication (ICEIC), 19-22 Jan. 2020 2020, pp. 1-5, doi: 10.1109/ICEIC49074.2020.9051009.

[60] N. Alipour, O. Tarkhaneh, M. Awrangjeb, and H. Tian, "Flower Image Classification Using Deep Convolutional Neural Network," in 2021 7th International Conference on Web Research (ICWR), 19-20 May 2021 2021, pp. 1-4, doi: 10.1109/ICWR51868.2021.9443129.

[61] X. Cheng, X. Feng, and W. Li, "Research on Feature Extraction Method of Fundus Image Based on Deep Learning," in 2020 IEEE 3rd International Conference on Automation, Electronics and Electrical Engineering (AUTEEE), 20-22 Nov. 2020 2020, pp. 443-447, doi: 10.1109/AUTEEE50969.2020.9315604.

[62] Y. Li, Y. Zhang, and Z. Zhu, "Learning Deep Networks under Noisy Labels for Remote Sensing Image Scene Classification," in IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, 28 July-2 Aug. 2019 2019, pp. 3025-3028, doi: 10.1109/IGARSS.2019.8900497.

[63] J. W. Chen, "Analysis of Subsampled Image Size for Detection and Identification of Brake Pad Contours by Using Deep Learning," in 2020 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE), 23-25 Oct. 2020 2020, pp. 174-176, doi: 10.1109/ECICE50847.2020.9301977.

[64] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "4S-DT: Self-Supervised Super Sample Decomposition for Transfer Learning With Application to COVID-19 Detection," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 7, pp. 2798-2808, 2021, doi: 10.1109/TNNLS.2021.3082015.

[65] C. So, "Exploring Meta Learning: Parameterizing the Learning-to-learn Process for Image Classification," in 2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 13-16 April 2021 2021, pp. 199-202, doi: 10.1109/ICAIIC51459.2021.9415205.

[66] X. Fang, Q. Luo, B. Zhou, C. Li, and L. Tian, "Research Progress of Automated Visual Surface Defect Detection for Industrial Metal Planar Materials," Sensors, vol. 20, no. 18, 2020.

[67] A. Abdullahi, N. A. Samsudin, M. R. Ibrahim, M. S. Aripin, and Z. A. Othman, "Towards IR4.0 implementation in e-manufacturing: Artificial intelligence application in steel plate fault detection," Indonesian Journal of Electrical Engineering and Computer Science, vol. 20, no. 1, p. 430, 2020.

[68] J. Gan, Q. Li, J. Wang, and H. Yu, "A Hierarchical Extractor-Based Visual Rail Surface Inspection System," IEEE Sensors Journal, vol. PP, no. 23, pp. 1-1, 2017.

[69] J. Duan et al., "Automatic 3D Bi-Ventricular Segmentation of Cardiac Images by a Shape-Refined Multi- Task Deep Learning Approach," IEEE Transactions on Medical Imaging, vol. 38, no. 9, pp. 2151-2164, 2019, doi: 10.1109/TMI.2019.2894322.

[70] Z. Wang, C. Tang, X. Sima, and L. Zhang, "Research on Application of Deep Learning Algorithm in Image Classification," in 2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), 14-16 April 2021 2021, pp. 1122-1125, doi: 10.1109/IPEC51340.2021.9421185.

Enterprise Architecture for Smart Enterprise System

A Quest for Chili Agrosystem

Meuthia Rachmaniah[1], Arif Imam Suroso[2]*, Muhamad Syukur[3], Irman Hermadi[4]
Computer Science Department, Bogor Agricultural University, Bogor, Indonesia[1, 4]
School of Business, Bogor Agricultural University, Bogor, Indonesia[2]
Agronomy and Horticulture Department, Bogor Agricultural University, Bogor, Indonesia[3]

***Abstract*—The chili agrosystem faces many challenges, and the enterprise architecture (EA) artifacts as the building block of a chili enterprise system (ES) are not exist. This research is a qualitative systematic literature review as part of developing intelligent ES to examine chili's production, consumption, and price. The first step toward ES development is recognizing worthy chili EA and EA frameworks that characterize existing chili market conditions. The study aims to answer three research questions (RQ) and uses a state-of-the-art approach, employing predetermined keywords, to six research databases and data gathered from the corresponding institutional agencies. The findings on RQ1 revealed eight dynamics chili main supply chain patterns and data segregation among institutional agencies. The RQ2 disclosed numerous studies on EA; however, none offered for the chili agrosystem. In addition, the RQ3 results are multiple and expose different EAF characteristics. Again, no study considers its applicability to the chili agrosystem. To conclude, the strength of enterprise architecture for the chili enterprise system is the resulting deliverables that fall into three categories. These are factors of the chili agrosystem, enterprise, and architecture factors. Of many available frameworks, the Zachman framework - The Ontology gives more offerings.**

Keywords—Chili agrosystem; enterprise architecture; intelligent enterprise; supply chain; Zachman framework

I. Introduction

Chili is an essential complementary ingredient for most Indonesian daily cuisines. Most chili farmers will plant chili when prices are high, resulting in an ample supply of chili and causing prices to fall. On the other hand, chili farmers do not grow chili when prices are low, resulting in a scarcity of chili, and eventually, prices will rise. Chili pests and diseases can also influence the supply of chili. Reference [1] analyzes the socio-economic and agro-ecological aspects of chili production. The results show that harvest losses due to pests and diseases are high. Also, the chili agribusiness risks for chili farmers are production, price, economic, and institutional risks [2]. These risks are interrelated. These risks that bring negative impacts are production risks due to plant diseases, markets due to price volatility, and monetary risks faced by chili farmers.

The chili supply chain has not been implemented efficiently and effectively in its management [3]. Most distribution chain actors face several barriers, such as losses, decreased products, and inefficiencies in the delivery period. The study of [4] stated that the chili supply chain is lengthy and unoptimized, including not entrusting the Village Owned Enterprises (called Bumdes) and the Indonesian Farmers Shop (called TTI). Reference [5] had identified 23 profiles of agro-industrial enterprises. The classification characteristics used are the location in the food chain and industry (field of activity), organizational forms and company law, company size, trade turnover, volume level, and technical and technological innovation base of the company (facilities and production technology used). Other characteristics are features of the company's structure (complexity, the presence of vertical and horizontal integration), the level and possible likelihood of production diversification, the company's price segment, the degree of product differentiation, the width and geography of market presence, the nature of influence in the market (the company's position in the market).

The importance of this research is longing for reconciling chili price volatilities due to supply and demand discrepancy. Here, establishing an intelligent enterprise system for chili agrosystem is an intended way of managing its supply to align with the market demand. Chili farmers and chili trade operators often faced obstacles in the distance, the high logistical costs to distribute seeds, fertilizer, pest medicines, and chili production from production centers to all areas in Indonesia. It creates a complex distribution of chili from producers to end consumers. Moreover, the condition of infrastructure is inadequate, especially in remote areas, and lack of technology to extend the shelf life of chilies to prevent them from rotting before reaching end consumers throughout Indonesia. The red chili supply chain requires a model compatible with various supply chain patterns occurring in Indonesia.

This study is part of the research on establishing intelligent enterprise systems to examine the production and consumption of chili. The first step towards its development is to recognize suitable chili enterprise architecture (EA) that characterizes existing conditions. The purpose of our study is incredibly inquisitive in:

- identifying the current problems in the existing chili agrosystem,
- investigating EA and its EA framework (EAF) suitable for the chili agrosystem, and
- determinating open challenges and areas for enhancement.

The study's direct audience for this paper is threefold. Firstly, we target researchers interested in a state-of-the-art overview of the area of the chili agrosystem. Secondly, we aim at researchers in the quest for the most helpful chili enterprise

*Corresponding Author.

architecture. Thirdly, we target chilies' farmers and or groups of farmers and its supply chain actors and stakeholders that would like to find out suitable enterprise architecture framework to improve the value of farmer exchange rates.

II. Proposed Method

This study used a state-of-the-art approach using previous latest ideas and methods from 2015-to 2020. To do this, we conducted a systematic literature review to lead our choice in determining the chili agrosystem enterprise architecture of available evidence and topic, including identifying shortcomings, inclinations, and voids in knowledge and indicating the direction it is beneficial to prompt further research. This section illustrates the foundation of this state-of-the-art by defining the state-of-the-art research questions and search keywords. We use a systematic literature review to answer the following state-of-the-art (STA) research questions.

A. State-of-the-Art Research Questions

EA and EAF are not new problems, and various methodologies, methods, and approaches offer to describe EA and EAF practices, including its implementation case studies. Our research aimed to ascertain the current issue of chili agrosystem and the quest for EA and EAF application for chili agrosystem, in particular, to answer the following research questions:

- RQ1 What is the Indonesia chili agrosystem outlook?
- RQ2 What is the state-of-art of EA?
- RQ3 What is the state-of-the-art of EAF applicable for the chili agrosystem?

In this study, the word ‘agrosystem outlook’ represents a series of activities and processes to characterize, develop, and maintain the delivery of fresh produce from farmers or grouped farmers to end consumers from the agri-business perspective. The words state-of-the-art of EA and EAF follow the description in Miriam Webster Dictionary. State-of-the-art determines the level of development (such as devices, procedures, processes, techniques, or science) that is achieved at a given time and can be used due to modern methods. Our study determines the state-of-the-art mostly in its implementation success factors.

B. Search Process

The search process is carried out by searching relevant articles using the list of keywords depicted in Fig. 1. The search applied to six research databases: IEEE Xplore, ACM Digital Library, Science Direct-Elsevier, Taylor and Francis, Springer Link, and Google Scholar, published between 2015 and 2020.

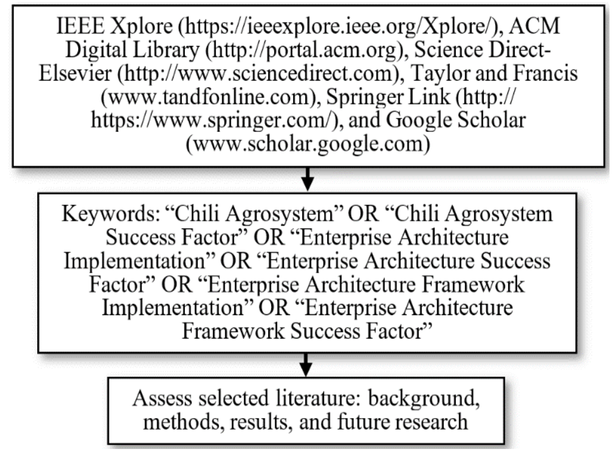


Fig. 1. Keywords Search flow Applied on Various Research Databases.

C. Scope of the State-of-the-Art Conduction

Our study defines the construct and guidelines for managing the STA as follow:

- Inclusion criteria: English peer-reviewed journal papers and conferences proceeding.
- Exclusion criteria: Book, Book Chapter, Indonesian peer-reviewed studies, government publication, studies irrelevant to the research questions, duplicate studies (by content and title), and short paper (e.g., poster).

Concerning the exclusion criteria, prospect papers that are not a specific approach for chili and practice are not within the scope of this research. This study intends to focus on practices and critical success factors of chili agrosystem and implementation of EA and EAF. Table I present the application of the query based on the keywords: “chili agrosystem,” “chili agrosystem success factor,” “enterprise architecture implementation,” “enterprise architecture success factor,” “enterprise architecture framework implementation,” or “enterprise architecture framework success factor.” Table II depicted the citations of the selected 32 papers obtained using Google Scholar.

TABLE I. Studies Retrieved through Various Search Engines Conducted in June 2020

Source	Paper Found	Candidate	Selected
IEEE Xplore	359	6	1
ACM Digital Library	7	6	3
Science Direct – Elsevier	9	6	4
Taylor and Francis	3	-	-
Springer Link	35	11	5
Google Scholar	333	51	19
Total	746	80	32

TABLE II. SELECTED PAPER RECENT CITATIONS

Reference	Cited	Reference	Cited	Reference	Cited
[1]	9	[16]	18	[27]	18
[2]	3	[17]	13	[28]	12
[3]	4	[18]	34	[29]	40
[4]	2	[19]	2	[30]	65
[5]	3	[20]	107	[31]	9
[10]	85	[21]	22	[32]	6
[11]	32	[22]	19	[33]	37
[12]	15	[23]	2	[34]	2
[13]	134	[24]	36	[35]	191
[14]	20	[25]	29	[36]	4
[15]	15	[26]	76		

III. RESULT AND DISCUSSION

A. RQ1 – Chili Agrosystem Outlook

The need for chili (red chili and cayenne pepper) for Indonesia's large cities with one million or more is around 800,000 tons/year or 66,000 tons/month. During the festive season or religious holiday, the need for chili usually increases by about 10-20% of regular requirements [6]. An enormous chili consumption per month indicates the need for national chili production and consumption system integrated with planting time management. The Ministry of Agriculture collects data and information on the production and area of chili commodity land, the Central Statistics Agency (BPS) handles export and import data, and the Ministry of Trade bears chili prices on the domestic and international markets.

Red chili land area from 2014-to 2018 experienced a decrease of -3.99%. On the contrary, the chili harvested area experienced an increase in growth of 22.50% (Fig. 2). Production of red chili and cayenne peppers grew by 0.04% for red chilies and 14.75% for cayenne peppers. The productivity of cayenne pepper experienced positive growth of 13.07%, while the productivity of red chili experienced a negative growth, which decreased by 0.13%. However, the productivity of red chili is better when compared to the productivity of cayenne pepper.

The consumption of red chili commodities is relatively high, especially in periods that coincide with religious holidays. According to [7] data, the highest red chili consumption per capita per month occurred in the province of West Sumatra (0.59 kg/month), followed by Bengkulu province (0.44 kg/month) and Banten province (0.42 kg/month) (Fig. 3.a). Whereas the highest level of consumption of red chili per ton per year in 2017 occurred in Banten province (93,234 tons/year), followed by West Java province (61,657 tons/year) and North Sumatra province (55,194 tons/year). Regardless of its consumption, in general, the national supply and demand projection showed a surplus (Fig. 3.b). Nevertheless, price fluctuation occurs every year.

Chili plantations spread in almost all provinces in Indonesia. According to [8], the three largest large chili production centers in 2018 are West Java (274,037 tons), Central Java (171,796 tons), and North Sumatra (155,835 tons) (Fig. 4). The three largest cayenne pepper production centers in 2018 are Banten (453,338 tons), West Kalimantan (20,530 tons), and DI Yogyakarta (141,771 tons).

Fig. 4 indicates that the province determines the priority of the type of chilies to be planted. For example, West Java chili production in 2017 was 134,910 tons, but in 2018 there was no data available for the production of cayenne pepper; on the contrary, in the year 2018, West Java's red chili production was the highest in Indonesia, which was 274,037 tons. In 2018 the production of red chili in Banten province was only 6,712 tons, but in the same year, Banten produced the highest production' of cayenne in Indonesia, which was 453,338 tons.

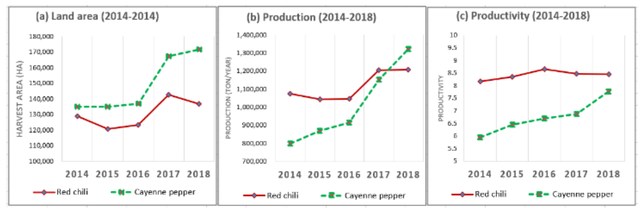


Fig. 2. (a) Land Area, (b) Production and (c) Productivity of Chili and Cayenne Pepper Year 2014-2018.

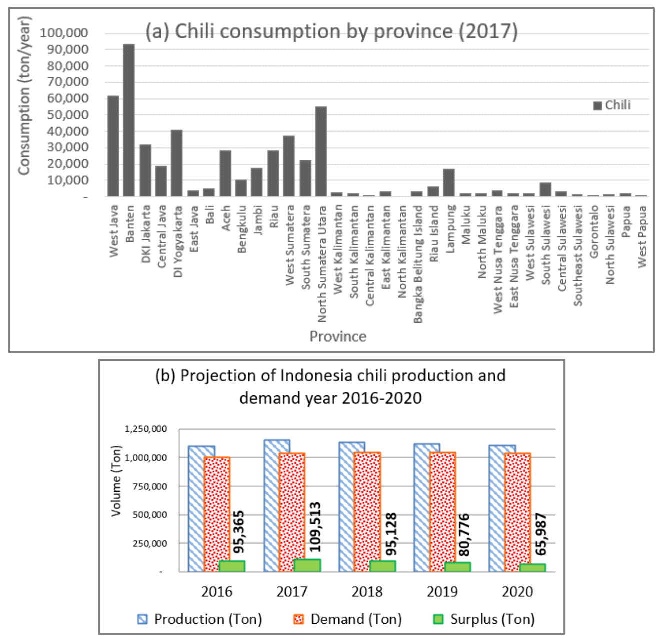


Fig. 3. (a) Year 2017 Consumption of Red Chili (Adapted from Agriculture Ministry, 2019) and (b) Year 2016-2020 Chili Projection of Supply and Demand (Adopted from [9]).

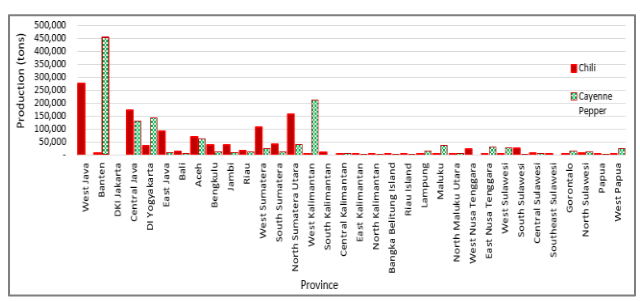


Fig. 4. Red Chili and Cayenne Pepper Production by Province Year 2018.

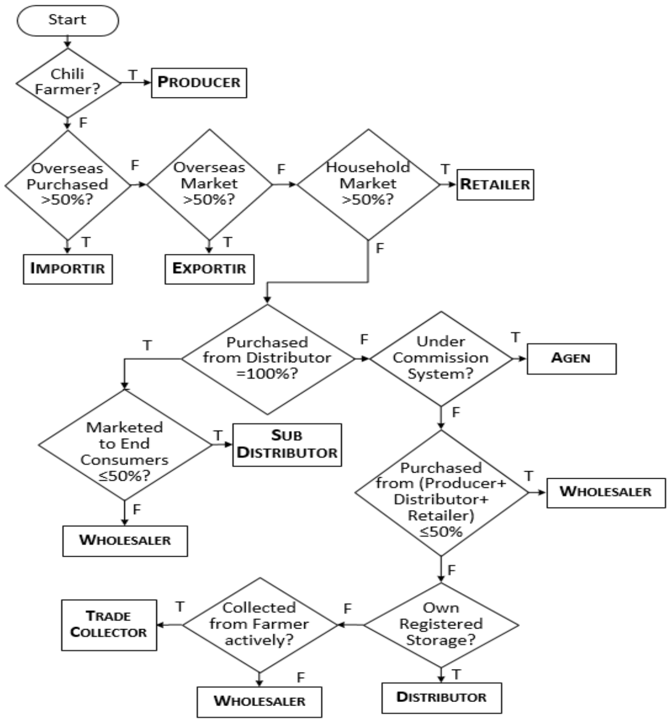


Fig. 5. Flowchart Determination of Chili Supply Chain Actors (T-True, F-False) (Adopted from [9]).

In 2018, the Indonesia Central Bureau of Statistics, known as Badan Pusat Statistik (BPS), surveyed the pattern of red chili distribution implemented by agrosystem red chili actors in 2009, 2015, 2017, and 2018 [9]. The actors of the red chili agrosystem consist of two groups, namely businesses and non-trade businesses. Trade businesses comprise medium, large and small companies that act as distributors, sub-distributors, agents, wholesalers, merchants, exporters, importers, or retailers (Fig. 5). Also, non-trading businesses/companies are red chili farmers as producers.

The BPS 2018 survey observed eight patterns of Indonesia's red chili trade system upstream to downstream. The most common red chili trade system pattern in 17 provinces in Indonesia is Pattern No. 1, which involves four actors in the red chili trade system (Table III). The most minor chili trade system operators are Pattern No. 6 and Pattern No. 8, which involve three red chili trade system actors. Almost all trading patterns involve retailers to end consumers. Meanwhile, DKI Jakarta has the most different supply chain patterns of the red chili trade system: red chili from outside the province rather than from farmers (producers) directly. DKI Jakarta obtains red chili from West Java, Central Java, and East Java provinces. In this survey, the definition of end customers is households, other business activities (restaurants, restaurants, catering businesses, hospitals, and hotels), processing industries, and government and non-profit institutions.

The BPS 2018 survey also analyzed the margins of trade and freight (MTF) merchant compensation as a supplier of goods. The MTF value is the difference between sales with the purchase value or the price difference from producers to final consumers. The MTF calculation considers the main trade patternsn*(see* Table III). Nationally, the MTF of red chili is 47.10% [9]. Regarding prices reaching the final consumer, the province with the lowest MTF is Riau Island province, with a total MTF of 15.25%, while the highest MPP occurs in South Kalimantan province, with a total MTF of 130.76% (Fig. 6). The MTF values greater than 100% also occur in Bengkulu and Maluku provinces. The high MTF value is mainly affected by transportation costs.

Data and information on the production and land area of strategic food commodities are managed separately by the Ministry of Agriculture, export and import data by the Ministry of Trade (Ministry of Trade), and the strategic food commodity price data by the Central Statistics Agency (knowns as BPS). The Ministry of Agriculture, BPS, and the Ministry of Trade has a separate work unit called the Data Center and Information System (Pusdatin). The Pusdatin manages the database of strategic food commodities following the domain of its authority. Also, to monitor price movements, in the year 2017, Bank Indonesia launched the National Strategic Food Price Information Center (PIHPS) on the hargapangan.id (Fig. 7). The PIHPS provides information on strategic food commodity prices daily; the enumeration is from Monday to Friday at 09:00-11:00 AM, reported to Bank Indonesia at 10:00-12:00 AM, and published at 01:00 PM West Indonesia Time. The number of samples is two retailers per traditional and modern market (primary market) per commodity in 82 districts/cities locations.

TABLE III. MAIN RED CHILI DISTRIBUTION PATTERN IN INDONESIA YEAR 2018 (ADOPTED FROM [9])

Pattern No.	Supply Chain Pattern	No. of Provinces
1.	Farmer → Trade Collector → Retailer → End Consumer	17
2.	Farmer → Trade Collector → Wholesaler → Retailer → End Consumer	4
3.	Farmer → Wholesaler → Retailer → End Consumer	5
4.	Farmer → Agent → Wholesaler → Retailer → End Consumer	2
5.	Other Provinces → Wholesaler → Retailer → End Consumer	1
6.	Farmer → Retailer → End Consumer	3
7.	Farmer → Distributor → Retailer → End Consumer	1
8.	Farmer → Trade Collector → End Consumer	1

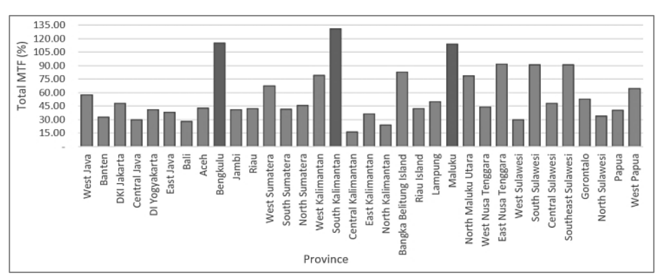


Fig. 6. Margin of Trade and Transportation Costs by Province (Adopted from [9]).

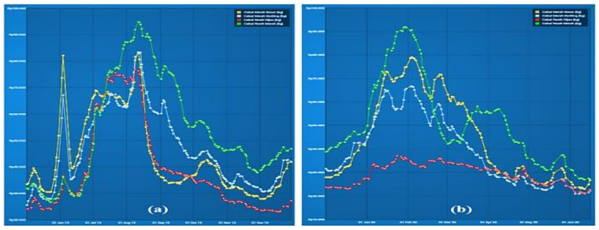


Fig. 7. The Menu Prices of Producers on Chili Commodities in West Java in the Graphic Report Format showed Fluctuating Prices: (a) Period May 2, 2019 until December 30, 2019 and (b) Period December 2, 2019, up to June 16, 2020.

B. *RQ2 – State-of-the-Art of Enterprise Architecture*

An enterprise architecture (EA) is a series of structured models that represent the building blocks of an enterprise system. The architecture framework simplifies processes and guides architects in all areas of architectural development, providing a set of conventions, principles, and practices. Reference [10] successfully identified 15 EA artifacts used by the 14 EA stakeholders interviewed. The artifacts are used to align with the goals and objectives of EA utilization in every part of the EA process and potential users that will use the produced artifacts. Further, [11] conducted an empirical analysis of 27 organizations and succeeded in identifying 24 EA artifacts that benefit organizations, explaining their practical use, and analyzing the empirical validity of the most popular EA conceptualization. Additionally, [12] reveal the post-implementation review apply to artifacts practices aiming to evaluate: the initiation of the implemented artifacts, the management and conducting EA artifact development, and the control of future change to the developed artifacts. The practices in the first category are business strategy, risk management, planning, and architectural method. The practices in the second category are governance, continuity, and stakeholder satisfaction. Lastly, the practices in the third category are alignment, architecture technique, management, and integration.

Reference [13] conducted a Monkey platform survey of 747 respondents (311 responded, 133 completed) and identified four success factors for EA management (EAM). These are the quality of EAM products, quality of EAM infrastructure, quality of EAM service and delivery, and EAM organization anchoring. Also, [13] suggested further research related to applying four EAM principles, which are determining EAM infrastructure, creating stakeholder awareness, providing high-quality EA products and services, and ensuring stakeholder commitment. Prior to that, [10] study review on over 100 special publications identifies six critical success factors (CSFs). Of the six CSFs, three CSFs successfully support the EA program implementation process. These are monitoring and compliance, commitment to using architecture, and consultation and communication.

The complexity of EA implementation and its bureaucratic business functions and complex IT structures become challenging problems for organizations. Reference [14] uses axiomatic design as a systematic approach for EA to analyze current enterprise capability and map the requirement of business, data, application, and technology layer of EA as the design domain aligned with the organization's strategic goals. Meanwhile, [15] measured implementation factors from the points of view of the experts and practitioners. The measurement comprises 27 factors that construct the 6-factor of internal process, 6-factor of learning and growth, 6-factor of authority support, 3-factor of cost, 3-factor of technology, and 3-factor of talent management. Analysis results have shown that there is no significant agreement between the experts and the practitioners except for rules and process of internal process category, the political influence of authority support category, and financial resources of cost category factors.

Further, [16] affirmed that EA program success derives primarily from how architecture is practiced rather than what is practiced. Meanwhile, [17] develop the CSF model of team capability, communication, top management commitment, technology and infrastructure, and governance. Reference [17] model indicated that governance gives the highest factor in successful EA implementation.

The study of [18] obtained 13 critical success factors (CSF) enterprise architecture implementation for the public sector (Fig. 8). In general, CSF that influences the successful implementation of EA in the public sector is technical development (68% articles) and frameworks and methodologies (50% articles). It is unavoidable that the people factor influences successful EA implementation. A study of [19] identified seven types of people factors. Five people factors have proven to have an association with successful EA implementation. These are skilled EA talent, centralized enterprise architect team, talent management plan, talent retention program, and EA learning culture. Surprisingly, the remaining two factors with a minor association with EA implementation are trained EA talent and certified EA talent.

To obtain benefits from EA, the findings of [20] highlight the importance of EA service capability and dynamic capability in creating benefits from EA. Reference [20] gives three recommendations. Firstly, how dynamic capabilities are activated through EA service capability and how projects and organizations benefit. Secondly, longitudinal studies are needed to fully understand how and why EA service capability develops over time, that is, reflecting the maturity of EA. Lastly, longitudinal studies will lead to a better understanding of the process to achieve organizational benefits. Also, [21] suggests three future works to carry out. Firstly, group the current EAIM problem into three main categories: modeling, development, and maintenance. Secondly, the identification of factors is evaluated from surveys of different EA project stakeholder groups (Fig. 9). Lastly, to carry out the effectiveness of EA implementation using factors and practices that affects the EA implementation.

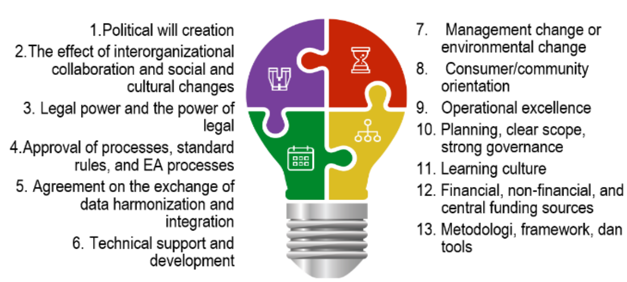


Fig. 8. CSF Implementation of EA in Public Sector (Adopted from [18]).

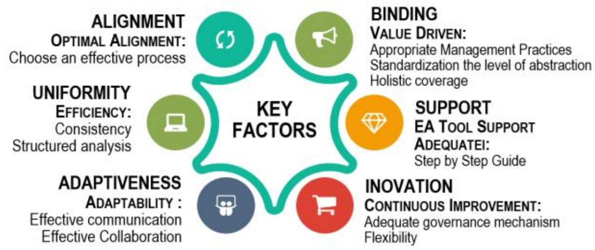


Fig. 9. Key Factors that Influence the Effectiveness of EAIM (Adopted from [21]).

The increasing flow of information and system integration in organizations and along the trade chain is one of the main challenges organizations need to address. Reference [22] proposed EA 4.0, which is an extended EA for the context of Industry 4.0. The EA 4.0 component consists of the EA model, data from enterprise information systems and IoT devices, and advanced analytics. His study reflects the Model4Insight platform that is still under development. For the record, EA 4.0 requires an EA model at a more detailed level, requires new skills, and works together with experts or data scientists.

Meanwhile, [23] uses a hybrid evaluation method for enterprise architecture that considers the organizational culture aspect encompassing the management sustainability variable consisting of governance, continuity, integration, and maintenance. Further, [24] uses EA to improve investment quality in information technology (IT). According to [25], top-quartile organizations use more EA artifacts to prepare IT investment decisions, specifically heat maps, policies, roadmaps, business capability models, and landscape diagrams.

One framework and approach that fits all does not apply in the case of EA because the different fields in which EA operates have unique/different field-specific requirements and specifications [25]. They suggest further research: (i) SOA requires further research on developing methods and models, and tools that directly measure service response time at higher accuracy; (ii) Issues and challenges to EA design; (iii) EA for healthcare needs to address aspects of organizational culture and professional culture; and (iv) Technology and methodology issues to guide the development of EA. Also, communication along the wide-distributed geography gives other challenges that might create a knowledge gap.

In [26] successfully identified current issues on cloud terminology, complexity theory, agile or adaptation, big data, things, entrepreneurship, intelligence, and sustainability. Based on their systematic literature review results, [26] recommended the need for further research on issues related to whether there was a mismatch between EA's academic efforts and EA in practice. Besides, trend analysis shows the number of publications in specific industries, such as health, manufacturing, and government issues. To fully understand the differences and similarities across the industry, further research has to be more detailed. Nevertheless, [27] developed Agile enterprise architecture (AEA) for geographically distributed Agile development (GDAD) environment. The study creates measurement model evaluation and 26 measures comprise of AEA (7-item), communication efficiency (5-item), communication effectiveness (4-item), on-time completion (2-item), on-budget completion (2-item), software functionality (3-item), and software quality (3-items).

C. RQ3 – State-of-the-Art of EA Framework

The architecture framework has several benefits [28]. The benefits are supporting stakeholders' decision-making about enterprise design and operations, improving users' trust that using reference architecture will be successfully applied to projects, and facilitating enterprise design communication. The architecture framework may be applied to various systems and enterprise scenarios. Another benefit of the architecture framework is building a general way to organize, interpret, and analyze architectural descriptions and identify architecture problems, generic stakeholders, viewpoints, and levels of abstraction. Besides, the architecture framework implements reuse and provide unified and unambiguous terminology definitions.

EA frameworks are differentiated according to their field of application [28]. Some EA frameworks frequently used are the Zachman Framework, the Open Group Architecture Enterprise (TOGAF) Architecture Development Method (ADM), Federal Enterprise Architecture Framework (FEAF), Department of Defense Architecture Framework (DoDAF). Ministry of Defense Architecture Framework (MODAF), Adaptive Enterprise Architecture Framework, and EA3 Cube Framework. In essence, the development of many EA Frameworks uses the Zachman Framework as its foundation (see Fig. 10 for EA framework evolution). Reference [29] mentioned that the existing EAF is too overwhelming for small-medium enterprises (SMEs). Hence, [29] developed a CHOOSE metamodel, an acronym for "maintain Control, employing a Holistic Overview, that is based on Objectives and kept Simple, of your Enterprise." Following evaluation and validation through five SMEs, CHOOSE metamodel includes only four essential concepts (goal, actor, operation, object), each applying to four primary EA focus (what, why, who, how). That supports [26] study that most EAF focus on EA implementation and EA artifacts development lacks implementation evaluation methods.

Reference [30] chose the Zachman framework in developing a system-of-system (SoS) architecture. DoDAF and MODAF are not suitable for SoS, while the TOGAF, FEAF, and Zachman frameworks are suitable. Based on the Zachman framework guidelines, the development of architectural SOS is best to use agent-based simulation integrated with SysML and UML. Meanwhile, [31] conducted an Agile modeling language study and provided a diagrammatic integration of machine-readable from several aspects of the Zachman framework. In [31] study, the 'What' aspect of dealing with the concept of Linked Enterprise Data environments, such as graph servers, graph databases, and RESTful HTTP requests with PHP-based programming languages with SPARQL queries for client requests. Agile modeling method engineering (AMME) creates and develops Agile modeling tools related to semantics, syntax, and functionality, namely environments such as ADOxx commonly used in prototyping. The primary key lies in the AMME, which provides tools to make prototyping in agile modeling [31] quickly.

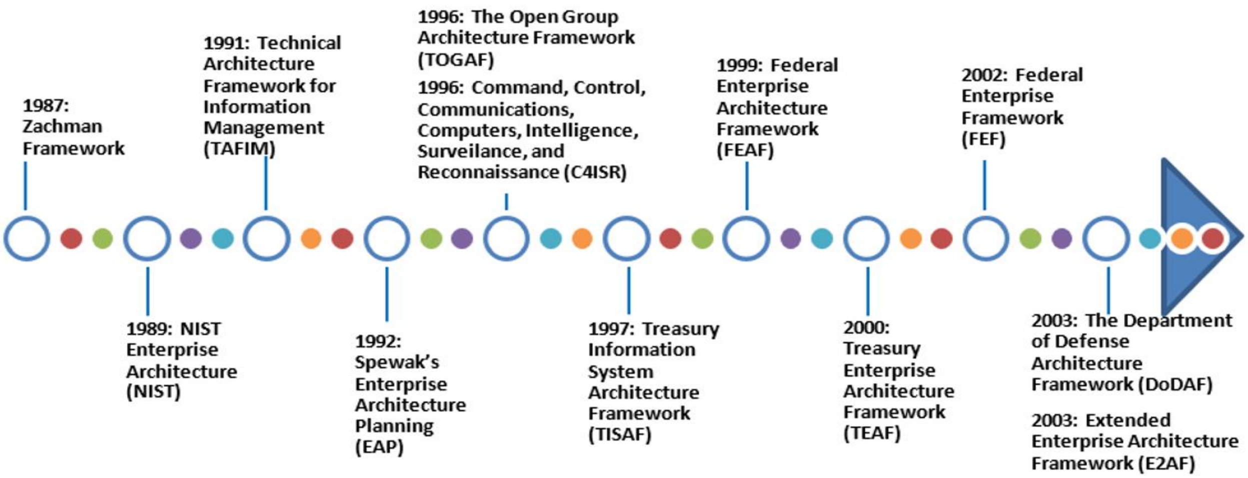
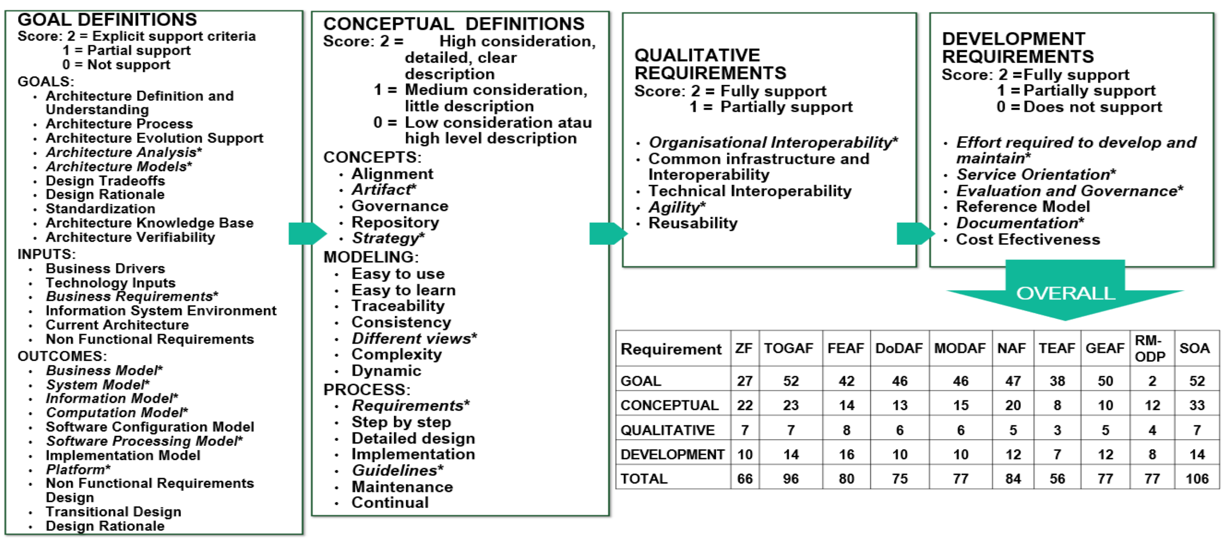


Fig. 10. The Evolution of EA Framework (Adopted from [28]).



Requirement	ZF	TOGAF	FEAF	DoDAF	MODAF	NAF	TEAF	GEAF	RM-ODP	SOA
GOAL	27	52	42	46	46	47	38	50	2	52
CONCEPTUAL	22	23	14	13	15	20	8	10	12	33
QUALITATIVE	7	7	8	6	6	5	3	5	4	7
DEVELOPMENT	10	14	16	10	10	12	7	12	8	14
TOTAL	66	96	80	75	77	84	56	77	77	106

Fig. 11. Stages of Comparative Analysis on ten EAFs and their Final Results (Adopted from [9]).

At present, there are more than 100 EAF available and used by four user groups: industry, government, open-source, and proprietary defense [32]. On critical IT infrastructure (CITI) design, [32] selected the EAF producing the ten most popular EAFs, resulting from usage percentage of the Zachman framework (25%), SOA (15%), TOGAF (11%), DoDAF (11%), FEAF (9%), British Ministry of Defense Architecture Framework (British MODAF) (2%), NATO Architecture Framework (NAF) (1%), TEAF (1%), Gartner EAF (GEAF) (3%), and ISO Open Distributed Processing-Reference Model (RM-ODP). Further, [32] scored each criterion on goals definitions, conceptual definitions, qualitative requirements, and development requirements (Fig. 11). Comparative analysis of [32] shows that the highest total overall score for CITI design is SOA and TOGAF. Almost all EAFs surveyed ignored or had less description of the rationale architecture design, even though needed in CITI design. Some EAFs include documentation of system boundaries and assumptions. Therefore [32] stated that the foundation of architecture must include criteria, benefits, and risks.

To propose an effective EA implementation methodology (EAIM) using the list of products in each EAIM phase (Fig. 12), [33] firstly identified EAIM criteria and compared its uses on four EAF (Table IV). Considering the complexities of EA implementation, [33] exploration found that no effective methodology existed. Nevertheless, the proposed EAIM aligns business and IT, integrates application and infrastructure, defines appropriate objectives and vision, governance plan, and step-by-step guidelines.

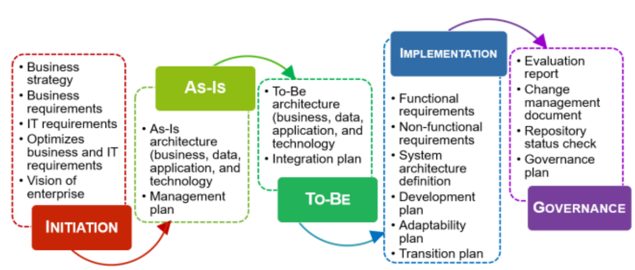


Fig. 12. List of Products by EAIM Phase to Measure Effective EAIM (Adopted from [33]).

TABLE IV. COMPARATION OF EAIM CRITERIA AMONG EAF

Identified EAIM Criteria	EAP	TOGAF	FEAF	DoDAF
Iterative	-	++	-	-
Management process	-	-	-	-
Maintenance process	-	+	-	-
Ability to work with other EAF	+	++	-	-
Requirement management process	-	++	-	-
Step-by-step guidelines	+	+	+	+
Easy to understand	+	-	+	+
Non-functional requirement	-	++	-	++
Complexity management	-	-	-	-
Supporting tool	+	+	-	+
Governance	+	++	+	-
Type-usage	++	++	-	-
Repository	++	++	++	++
Easy to implement	+	-	+	+

[a.] Note: ++ Fully consider/support/exist/all/easy;

[b.] + Partly consider/partly support/partial/somewhat/easy;

[c.] - Not consider/not support/not exist/particular/difficult

In [34] selected the EAF for e-Government by applying the Multiple Criteria Decision Making (MCDM) method with the Analytical Hierarchy Process (AHP) tool to select four EAFs, namely the Zachman framework, FEAF, TOGAF, and TEAF. The AHP results showed the highest ranking of EAF are Zachman framework (30.76%), FEAF (25.23%), TOGAF (24.13%), and TEAF (19.13%). The results of the matrix combination show the preference attributes for the Zachman Framework, which are the preferred framework attributes, knowledge base, and architecture evolution support.

In [35] explored EA in the future using the Zachman framework version 3.0 - The Enterprise Ontology perspective. The world's challenges are even more significant with a global market that affects social transformation and government instability. The results of the exploration of [35] are fourfold. First, various non-technical domains must contribute to the progress of EA. Secondly, insight into systems thinking and complexity science. Thirdly, there is no significant progress on EA related to new enterprise realities (e.g., virtual, boundaryless, heterogeneous culture, and retention knowledge). Lastly, many advances have been discussed but are still very early. Reference [36] identified 14 criteria to evaluate EAF artifacts. These criteria are applied for the Zachman Framework, TOGAF, FAF, EAP, The Enterprise Architecture, and DoDAF by three designated experts with overall perceived results 90.87% usable, 97.62% relevant, and 90.48% correct.

As the implications, further study using manageable chili enterprise systems requires bringing farmers closer to end consumers that enhance farmers' welfare [37]. Farmers might utilize a predictive analytics model to plan optimal chili production schedules and plan logistics and distribution chains to several regions to reduce the production cost and increase profit for the farmer [38] [39]. Chili agrosystems may be considered family companies and small-medium enterprises with a multidimensional concept with several variables, including actors, attributes (motivation and activities), and consequences or outcomes [40]. The smart enterprise system has to build knowledge foundations, bearing in mind the technology-assisted organization, leadership, and management models [41]. Therefore, enterprise architecture implementation and critical success factors and artifacts may address using various methods (Fig. 13).

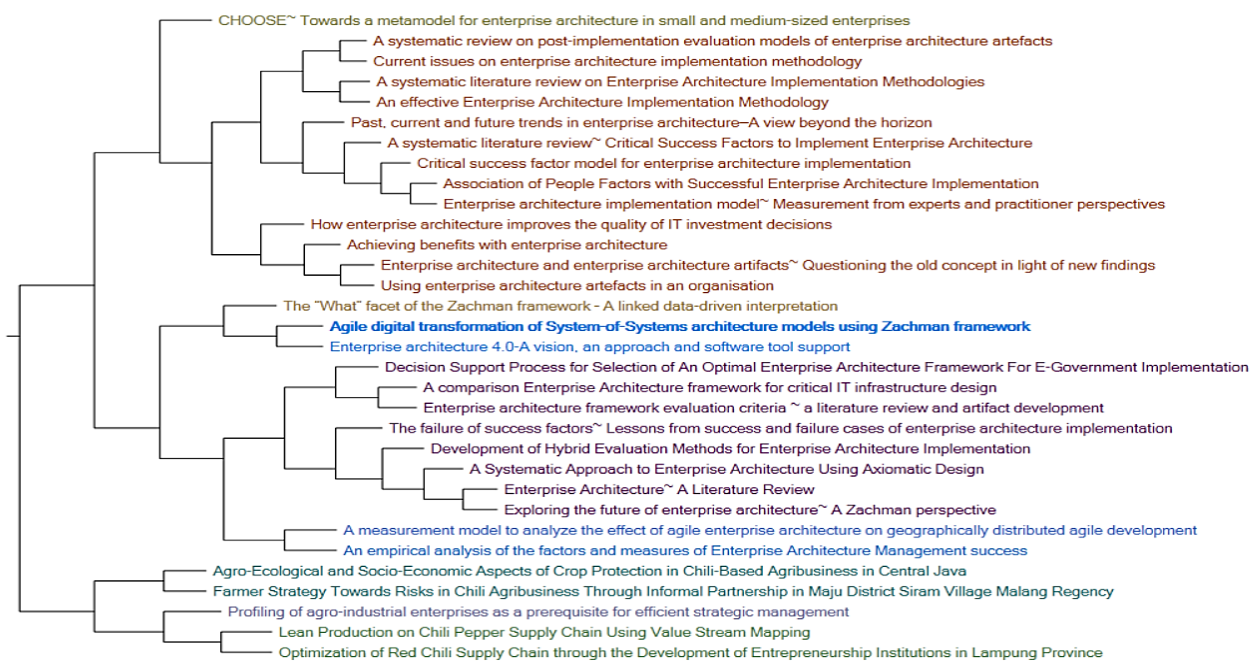


Fig. 13. Horizontal Dendogram of State-of-the-Art Results.

IV. Conclusion

This study's main findings are that using an enterprise architecture framework might provide better solutions to guide the supply chain management in governing the price volatility of the chili agrosystem. Even though the proposed methods do not represent the entire publication, they suffice the goals. Chili agrosystem implicates multiple mechanisms, such as the array of heterogeneous production centers and productivity, dynamic and variably supply chain patterns, unappealing trade and freight margin, and segregated chili datasets among associated institutions.

There are diverse open challenges and areas for further studies. Conforming to the results, we contribute three lines of work, particularly if we consider that some problems are to be solved. Firstly, enterprise architecture is unique depending on the nature of the enterprise: government institutions, non-governmental agencies, e-commerce, virtual organization, or any organizations. Hereupon, resolving the current problem can probably be separated into three primary classes containing modeling, development, and maintenance. Next, each enterprise architecture framework has its well-defined characteristics.

Consequently, enterprise architecture choices are not general but depend on the nature of the enterprise. Last, the chili agrosystem outlook demands the availability of an integrated creation of a chili dataset for chili farmers or groups to access. The existence of the chili dataset can better enhance the economic, social, institutional, and environmental appearances of the chili agrosystem.

In conclusion, the strength of enterprise architecture relies upon three factors of the resulting deliverables. These are enterprise, influencing, and architecture factors. Nevertheless, of many available frameworks, the Zachman framework gives more offerings for developing an intelligent enterprise system of fresh chili. We suggest a thorough study investigating these factors to achieve a suitable framework for the chili agrosystem. Further study is also required on various supply chain patterns using a Zachman enterprise architecture framework to improve the efficacy of the chili agrosystem.

Acknowledgment

This work was supported by the Computer Science Department, Faculty of Mathematics and Natural Science, IPB University.

References

[1] Mariyono J. Agro-Ecological and Socio-Economic Aspects of Crop Protection in Chili-Based Agribusiness in Central Java. Agriekonomika. 2017;6(2):120.

[2] Hidayat K. Farmer Strategy Towards Risks in Chili Agribusiness Through Informal Partnership in Maju District Siram Village Malang Regency. Agric Soc Econ J. 2017;17(1):6–15.

[3] Perdana T, Hermiatin FR, Pratiwi ASN, Ginanjar T. Lean Production on Chili Pepper Supply Chain Using Value Stream Mapping. Mimb J Sos dan Pembang. 2018;34(2):311–20.

[4] Barusman ARP, Soewito, Romli K. Optimization of Red Chili Supply Chain through the Development of Entrepreneurship Institutions in Lampung Province. Rev Integr Bus Econ Res. 2019;8(2):233–43.

[5] Lomachenko TI, Kokodey TA. Profiling of agro-industrial enterprises as a prerequisite for efficient strategic management. IOP Conf Ser Earth Environ Sci. 2020;421(2).

[6] Kementerian Perdagangan. Profil Komoditas Barang Kebutuhan Pokok dan Barang Penting - Komoditas Cabai. Jakarta: Kementerian Perdagangan; 2016.

[7] Badan Pusat Statistik. Distribusi Perdagangan Komoditas Cabai Merah Indonesia Tahun 2018. Badan Pusat Statistik. 2018. 1–97 p.

[8] Pusdatin - Pusat Data dan Sistem Informasi Pertanian. Outlook Cabai - Komoditas Pertanian Subsektor Hortikultura. Pus Data dan Sist Inf Pertan. 2020;1–51.

[9] Datamikro K, Statistik BP. Indonesia - Survei Sosial Ekonomi Nasional 2017 Maret (KOR). 2018.

[10] Niemi E, Pekkola S. Using enterprise architecture artefacts in an organisation. Enterp Inf Syst. 2017;11(3):313–38.

[11] Kotusev S. Enterprise architecture and enterprise architecture artifacts: Questioning the old concept in light of new findings. J Inf Technol. 2019;34(2):102–28.

[12] Nikpay F, Ahmad R, Rouhani BD, Shamshirband S. A systematic review on post-implementation evaluation models of enterprise architecture artefacts. Inf Syst Front [Internet]. 2016;1–20. Available from: http://dx.doi.org/10.1007/s10796-016-9716-0.

[13] Lange M, Mendling J, Recker J. An empirical analysis of the factors and measures of Enterprise Architecture Management success. Eur J Inf Syst. 2016;25(5):411–31.

[14] Behrouz F, Fathollah M. A Systematic Approach to Enterprise Architecture Using Axiomatic Design. In: Procedia CIRP. 2016.

[15] Bakar NAA, Harihodin S, Kama N. Enterprise architecture implementation model: Measurement from experts and practitioner perspectives. Colloq Inf Sci Technol Cist. 2016;0:1–6.

[16] Hope T, Chew E, Sharma R. The failure of success factors: Lessons from success and failure cases of enterprise architecture implementation. SIGMIS-CPR 2017 - Proc 2017 ACM SIGMIS Conf Comput People Res. 2017;21–7.

[17] Rouhani BD, Ahmad binti R, Nikpay F, Mohammaddoust R. Critical success factor model for enterprise architecture implementation. Malaysian J Comput Sci. 2019;32(2):2019.

[18] Ansyori R, Qodarsih N, Soewito B. A systematic literature review: Critical Success Factors to Implement Enterprise Architecture. Procedia Comput Sci. 2018;135:43–51.

[19] Bakar NAA, Hussien SS. Association of People Factors with Successful Enterprise Architecture Implementation. Int J Eng Technol. 2018;7(4.31):52–7.

[20] Shanks G, Gloet M, Asadi Someh I, Frampton K, Tamm T. Achieving benefits with enterprise architecture. J Strateg Inf Syst. 2018;27(2):139–56.

[21] Nikpay F, Ahmad binti R, Rouhani BD. Current issues on enterprise architecture implementation methodology. Internattional J Soc Educ Econ Manag Eng. 2015;9(1):112–5.

[22] Aldea A, Iacob ME, Wombacher A, Hiralal M, Franck T. Enterprise architecture 4.0-A vision, an approach and software tool support. Proc - 2018 IEEE 22nd Int Enterp Distrib Object Comput Conf EDOC 2018. 2018;1–10.

[23] Noviansyah B, Arman AA. Development of Hybrid Evaluation Methods for Enterprise Architecture Implementation. Proceeding - 2018 Int Conf ICT Smart Soc Innov Towar Smart Soc Soc 50, ICISS 2018. 2018;1–4.

[24] Van den Berg M, Slot R, van Steenbergen M, Faasse P, van Vliet H. How enterprise architecture improves the quality of IT investment decisions. J Syst Softw. 2019;152:134–50.

[25] Gorkhali A, Xu L Da. Enterprise Architecture: A Literature Review. J Ind Integr Manag. 2017;02(02):1750009.

[26] Gampfer F, Jürgens A, Müller M, Buchkremer R. Past, current and future trends in enterprise architecture—A view beyond the horizon. Vol. 100, Computers in Industry. 2018.

[27] Alzoubi YI, Gill AQ, Moulton B. A measurement model to analyze the effect of agile enterprise architecture on geographically distributed agile development. J Softw Eng Res Dev. 2018;6(1).

[28] Kale V. Enterprise Process Management Systems: Engineering Process-Centric Enterprise Systems using BPMN 2.0. CRC Press Taylor and Francis Group. 2019.

[29] Bernaert M, Poels G, Snoeck M, De Backer M. CHOOSE: Towards a metamodel for enterprise architecture in small and medium-sized enterprises. Inf Syst Front. 2016;18(4):781–818.

[30] Bondar S, Hsu JC, Pfouga A, Stjepandić J. Agile digital transformation of System-of-Systems architecture models using Zachman framework. J Ind Inf Integr [Internet]. 2017;7:33–43. Available from: http://dx.doi.org/10.1016/j.jii.2017.03.001.

[31] Harkai A, Cinpoeru M, Buchmann RA. The "What" facet of the Zachman framework – A linked data-driven interpretation. Lect Notes Bus Inf Process. 2018;316:197–208.

[32] Dorohyl Y, Tsurkan V, Telenyk S, Doroha-Ivaniuk O. A comparison Enterprise Architecture framework for critical IT infrastructure design. Inf Tchnology Secur. 2017;5(2(9)):90–118.

[33] Nikpay F, Ahmad RB, Rouhani BD, Mahrin MN, Shamshirband S. An effective Enterprise Architecture Implementation Methodology. Inf Syst E-bus Manag. 2017;15(4):927–62.

[34] Mokone CB, Eyitayo OT, Masizana A. Decision Support Process for Selection of An Optimal Enterprise Architecture Framework For E-Government Implementation. J e-Government Stud Best Pract. 2019;2019:1–14.

[35] Lapalme J, Gerber A, Van Der Merwe A, Zachman J, Vries M De, Hinkelmann K. Exploring the future of enterprise architecture: A Zachman perspective. Comput Ind. 2016;79.

[36] Hadaya P, Leshob A, Matyas-balassy PMI. Enterprise architecture framework evaluation criteria : a literature review and artifact development. Serv Oriented Comput Appl. 2020;1–20.

[37] Rachmaniah M, Suroso AI, Syukur M, Hermadi I. Strategic Food Risks – Chili'S Agrosystem Perspective. J Manaj dan Agribisnis. 2021;18(1):19–31.

[38] Siregar JJ, Suroso AI. Big Data Analytics Based Model for Red Chili Agriculture in Indonesia. In: Lecture Notes on Data Engineering and Communications Technologies. Springer Science and Business Media Deutschland GmbH; 2021. p. 554–64.

[39] Suroso AI, Syukur M, Hermadi I, Rachmaniah M. Sistem Enterprise Komoditi Pangan Strategis. Bogor Indonesia: PT Penerbit IPB Press; 2022. 218 p.

[40] Wahyudi I, Suroso AI, Arifin B, Syarief R, Rusli MS. Multidimensional aspect of corporate entrepreneurship in family business and SMES: A systematic literature review. Economies. 2021;9(4):1–17.

[41] Maulana MM, Suroso AI. Designing a simple house of knowledge management. Digital Transformation Management. 2022. 107–126 p.

Virtual Reality Application for Pain Management: User Requirements

Ioan Alexandru Bratosin[1], Ionel Bujorel Pavaloiu[2], Nicolae Goga[3], Andreea Iuliana Luca[4]
Department of Engineering in Foreign Languages, University Politehnica of Bucharest, Bucharest, Romania[1, 2, 3]
Faculty of Psychology and Education Science, University of Bucharest, Bucharest, Romania[4]

***Abstract*—The usage of fully immersive Virtual Reality applications for pain relief is still in an exploratory stage. In consequence there is a need to understand the user perspective and wishes regarding this kind of product. To address this issue, this paper presents quantitative research in order to establish the functional and non-functional requirements of our application. Voluntary response sampling was used for the research (N = 55). The inquiry form contained questions regarding serious game content, eagerness for testing, performance, resource consumption optimization, portability, data security, accessibility. The questionnaire was shared via Google Forms. The answers were collected and interpreted. The study revealed that a significant part of the participants was willing to test the application and that they would use an immersive Virtual Reality application during a normal treatment session if the opportunity is available. As functional requirements, the following were considered important: the presence of animals in game, a bright environment and nature-based background sounds. The following non-functional requirements were considered important: game optimization, portability, data security, accessibility, graphics quality and a short learning curve.**

***Keywords*—Virtual Reality; user requirements; serious games; therapy**

I. Introduction

A subject of major significance for researchers is the usage of software application like serious games in fields like psychology [1,2], economics [3], medicine [4,5], automotive industry [6], etc. One of the ideas for serious games is to examine research hypotheses [7] in a safe and controlled environment. Another purpose is to facilitate the instruction of personnel with the objective of significantly decreasing the material and human resource costs [7].

In order to provide the best user experience during different test phases of the application or during personnel training, there is a prerequisite to ensure that the functional and non-functional requirements are properly defined for the type of application being developed. A user questionnaire can be used to examine the suitability of the requirements [8, 9], depending on the type of target population, while the method of sampling can range from convenience or voluntary sampling to random simple, stratified, or clustered sampling [10].

Virtual reality (VR) pain therapy is a new domain [11], where the applications developed are not created based on a predefined template or established requirements. The applications used in VR therapy (VRT) for pain relief range from non-interactive VR applications [12], interactive VR application not meant for pain relief [13] to custom made applications [11]. All of them had one factor in common, they provided a certain efficiency in pain relief and all users would recommend the therapy. Another thing to notice is the variety in the applications used, they do not have a standard and they are not based on predefined requirements for virtual reality therapy [12, 13].

The main benefit of this research and the approach presented is the definition of a functional and non-functional requirement template on which VR software applications for pain relief therapy can be developed. This would ease the creation of VR applications meant for pain relief as well as improving the effects of VR on pain treatment as the applications created would be developed around the user's needs.

This paper is divided in two main sections after presenting Section 1, Introduction. The second section presents a literature review containing the state of the art in the use of user questionnaires for obtaining relevant data. This section is divided in 2 sub-sections. The first sub-section deepens the subject towards a more specific topic; functional and non-functional requirements (FRs and NFRs) for software application and the second sub-section that describes the use of serious games, VR applications in medicine and the need for establishing a suitable protocol for defining the FRs and NFRs. The third section contains the evaluations done on the user questionnaire in order to verify its validity and relevance towards the presented research subject. The section starts with a description of the user's opinion on multiple functional and non-functional requirements and continues with an evaluation of their relevance by taking into consideration statistical parameters of the responses.

II. Literature Review

Assessing the usability of an application is a very important part in the development process [14] and it can be assessed through various methods, such as quantitative measures. Although questionnaires are among the best-known measures of user experience (UX), due to the high heterogeneity of the instruments [15], choosing the best tool to assess UX became a difficult task. Each of these formats has advantages and disadvantages which will be discussed further. In this regard, an extensive analysis has been made [15], indicating that the majority of usability questionnaires contain short items measured on a Likert scale. The System Usability Scale (SUS) is a short – 10 items – questionnaire that can be used to assess the general impact of an application. This questionnaire is highly employed due to the short completion time and flexible structure and it can be used in many practical settings, such as medical field [16], including the adaptation on applications for pain [17].

The System Usability Scale was created to assess the usability of a product or service and has been widely used in numerous areas such as VR rehabilitation and health services [18], VR learning [19] and VR training, and also for VR locomotion techniques [20] and it is one of the most widely used standardized questionnaires for the assessment of perceived usability [21]. A plethora of studies proved the reliability of SUS scale [22]. The SUS has many advantages, one of the most important being its flexibility [23] that allows it to assess a wide range of technologies.

Usability tests are conducted with either samples of participants who meet specific health criteria [18] or the general population [23], and most of the time is based on convenience sampling [24] a type of non-probability sampling used in research [25]. Apart from the sampling method, another issue to be taken into consideration is the sample size, which can be determined through statistical techniques by estimating the variance of the dependent measure(s) – for task-level measurements – or through more complex formulas and techniques in the case of problem-discovery usability measurement [26].

Overall, the heterogeneous state of art for UX questionnaires raises several issues to be taken into consideration when analysing perceived usability. However, as discussed above, questionnaires are one of the most widely used tools in assessing UX due to their efficiency and practicality. In addition to those, in the paper [27] is indicated that the usability research may benefit from VR technology for user research and human-product interaction.

A. Functional and Non-functional Requirements for Medical Applications

The process of developing software applications needs to fulfil several requirements, which are usually divided into functional requirements (FRs) and non-functional requirements (NFRs). FRs correspond to the capabilities of the system, whereas NFRs (e.g., being user-friendly, capability, performance, stability etc.) describe the overall proprieties that a system [28] must have and they may not be directly related to specific system components [29]. Due to their nature, NFRs – also called "system features" [30] - are more abstract requirements, therefore more difficult to define and quantify.

A recent analysis of requirements, made by Stamm et al. [31] for a VR intervention on a group of geriatric patients with chronic back pain in order to determine their importance, analysed the requirements for the overall system - software, hardware and gamification or game integration. Among the most important requirements, the participants listed: an individual briefing for the system, presenting instructions in a detailed manner and that the length of the exercise to be a maximum of 30 minutes (i.e., in accordance with the safe and healthy warning of the oculus advising 10-15 minutes break every 30 minutes). Also, the "user-friendly handling of the system" was important to ensure a safety regulation. Another important factor, highlighted by the authors, is the age of the participants, as elder participants have special requirements [32] and would need additional training in using the equipment.

B. Virtual Reality in Medicine

Virtual Reality is a domain that is continuously expanding its utility in the medical domain [33] due to its capabilities of simulating environments that closely resemble reality. The rendering of almost photorealistic scenes and the use of controllers with haptic feedback allows virtual reality surgical simulators to increase the skills of surgeons without endangering the patient. Medical applications can range from dentistry [34], intravenous-insertion, chest-tube insertion, central venous placement catheter simulators to therapeutic treatment of physical affection [35] or mental health disorders [36] like phobias.

VR and augmented reality (AR) are studied for their capabilities of increasing surgical accuracy, decreasing the overall length of surgery and improving surgery techniques [37] where the author shows the use of the enhanced view provided through the digital generated images. One of the research reviews was focused on pedicle screw placement where it was found that AR increased both accuracy and efficiency with thoracic pedicle screws. It was used as well in training up to 51 residents in this type of surgery and it was demonstrated that they had benefited from the AR technology. A clinical trial that used 20 patients for spinal fixations displayed a high level of 94.1% accuracy for the thoracic pedicle screw operation with the help of AR [38]. For the training mechanism in VR for spine surgery, in paper [39] it is revealed that the group which used the "ImmersiveTouch" VR simulator had better performance levels in all statistics like trajectory, depth of screw error and breach, compared with the control group that used the usual training methods. The success was determined by the 3D anatomical precision and representation of the model. The overall use of VR technology was reviewed [34], regarding the simulation of different medical procedures (i.e., laparoscopic [40] with LapSim, Lap Mentor, orthopaedics with TraumaVision, Procedicus KSA VR or other surgeries with Visible Ear Simulator). The HipNav VR simulator was developed for orthopaedics, and it contains a kinematic model of the hip joint and tools for specific procedures [41]. The main use of surgical simulators is to provide advanced training, automatic scoring and analysis of the intervention with objective metrics, so the VR surgical simulators improve overall performance and lead to less injury to patients during the real operation [42].

Another application of VR technology is for pain management, due to the lack of negative effects after the treatment unlike opioids. To validate the effects of VR on acute or chronic pain, a significant number of randomized trials are required. A randomized trial was performed on 120 subjects, half VR, the other control to test the efficacy of VR [43]. The patients were hospitalized with an average pain score greater than 3 on a scale from 1 to 10. They have used an Oculus Gear set with a set library of VR application. The outcome was patient-reported pain using a numeric scale and then compared pre and post intervention and after 48 and 72 hours. The result obtained was a positive one, as there was an observed difference in favour of VR. The result was more noticeable on patient with pain score greater than 7. The major areas where VR interventions are used include burn and wound care, intravenous insertions, dental procedures and surgery, where the first target is anxiety reduction [43]. Most of the results support the use of Virtual Reality for pain management in paediatric populations. But as the study points out, there is a need for the standardization in this domain to facilitate the creation, and research of VR software and hardware for pain management purposes. The paper [44] shows a review of multiple studies that have conducted clinical studies to verify the efficiency of VR with encouraging results. It proposes as well the combination of VR with Music Treatments (MT) to further reduce pain through mechanism that implicates distractions, mood regulation, and engagement. Further research is required to explore the combination of the two treatment types and assess their effectiveness.

A major issue with the previously described applications in the review studies [11] and [33] is the definition of their requirements. The main component that is stressed and showed importance in reducing pain in VR therapy is the immersive quality of the application, ease of use and environment. These qualities can be vastly improved by applying a combination of standardized and customized user forms to obtain the expectations of the users before the actual implementation of the application. This would lead to an improved design method for VR software application as well as the improvement of the effect of VR on pain therapy.

The main objective of the current research is to establish a template on which the initial stage of development for VRT applications can be eased and improved. This would lead to an improvement in the quality of VRT software and to its effects on therapy.

In conclusion, Virtual Reality is a new technology that requires continuous research through randomized clinical trials, creation of viable functional and non-functional requirements and a robust and standardized framework to analyse its effects.

III. Game4Pain Requirements

Following the previous discussion, in this paper the focus is on establishing the proper functional and non-functional requirements for the current VR application that is being research and developed by us. The end purpose of this application is to be applied as a pain relief alternative or as a secondary option to opioids in order to decrease the consumption of pain ameliorating drugs.

The study used a sample population of 55, from which 32.7% (18) male and 67.3% (37) female. The majority were in the age group of 18-24 years old - 83.6% (46), with a few in the age group of 25-34 - 12.7% (12), while the 35-44 and 45-60 groups of ages contained each 1.8%, one person.

The questionnaire provided to the volunteers included a set of 3 questions regarding their experience with VR applications, if any and their willingness to test a VR application for pain relief if they had the occasion (Table I). The first question asked was if the participants have received any VR treatment before.

The following answers were recorded: 36 of them did not receive any, 5 received full immersive VR, 6 Mobile VR, 7 persons Games and only one person TV based treatment.

The majority of participants expressed their enthusiasm to try the VR application for pain relief if they had the occasion, 34 people gave a positive answer, 18 were undecided and only 3 persons would not like to try.

The second part of the questionnaire in Table II consisted in establishing the functional and non-functional requirements, for this part there were 13 questions with answers on a scale from 1 to 5. In this part of the form, the questions were about the importance of various elements in the game e.g., background music, sunlight level, complexity level, the importance of wildlife and NFRs e.g., data security, accessibility, graphics quality, learning curve.

A. Functional and Non-functional Inquiry Form Analysis

From the analysis of the form questions mean and standard deviation results from Table II, the results show that the majority of the respondents are enthusiastic regarding the test of a fully immersive VR application given the mean score of 4.27 out of 5 with a relatively low standard deviation of 0.95. For the next item there is a similar high interest in having animals in the game with a mean score of 4.25. Regarding the fishing system the average is almost in the middle with 3.33 but as the standard deviation reach 1.44 it signifies that the opinions are likely opposed. A majority would like to have a fishing system implemented as a functional requirement while others are not interested.

Another interesting finding is that the subjects would prefer a luminous setting that simulates sunlight as it seen in question 5 and 6 where they affirm that the presence of sunlight affects their mood in a positive way.

TABLE I. User Attitude Towards VR for Pain Relief Therapy

Question	**Answer variant**	**Percentage of N=55**
Have you ever received a Virtual Reality treatment? If so, what type?	None	65.5%
	Full Immersive VR (e.g. Oculus Headset VR)	9.1%
	Mobile VR	10.9%
	Games	12.7%
	TV	1.8%
Would you prefer a full immersive VR experience over a normal TV/Game intervention during treatment?	Yes	61.8%
	No	5.5%
	Maybe	32.7%
Would you like to try a Virtual Reality game for the amelioration of pain, alongside usual treatment?	Yes	69.1%
	No	9.1%
	Maybe	21.8%

TABLE II. FR AND NFR FORM RESPONSE ANALYSIS

Question	Mean/ Standard deviation
1.How Interested would you be in testing a fully immersive VR application during treatment? (1 = Not at all, 5 = Very much)	4.27/0.95
2.Are you interested in having animals in the game? (1 = Not at all, 5 = Very much)	4.25/1.004
3.Are you interested in having the option to fish in the game? (1 = Not at all, 5 = Very much)	3.33/1.44
4.Would you prefer music or nature sounds to run in the background? (1 = Music; 5 = Nature)	3.65/1.40
5.Do you think that the level of sunlight in a game affects your mood in a positive way? (1 = Not at all, 5 = Very much)	3.96/1.05
6.Do you think that the level of sunlight in real life affects your mood in a positive way? (1 = Not at all, 5 = Very much)	4.22/1.08
7.How important you consider optimizing resource consumption. (How well the game runs on your device). (1 = Not important at all, 5 = Very important)	4.38/0.87
8.How important you consider the game having a short learning curve. (to be easy to learn) (1 = Easy, 5 = Hard)	3.33/1.14
9.How Important is Portability (ability to run on different operation system devices) (1 = Not important at all, 5 = Very important)	3.89/0.99
10.How Important is Data Security. (1 = Not important at all, 5 = Very important)	4.67/0.58
11.How Important is Accessibility (1 = Not important at all, 5 = Very important)	4.60/0.60
12.How Important is the Graphics quality (1 = Not important at all, 5 = Very important)	4.40/0.87
13.How would you consider the application having an in-depth Game Mechanics (the number of actions you can do in a game and the complexity) (1 = Simple Game, 5 = Complex Game)	3.98/0.83

As non-functional requirements regarding the application difficulty in question 8, the opinions are pointing towards an average learning curve, but with a high game complexity in question 13. There is a need to take in consideration the age of the group as its mostly young adults, and they would lean towards a challenging game, but due to the nature of the application as it is meant to be enjoyable to the target audience of persons that suffer from acute or chronic pain, therefore a slightly simpler application with a short learning curve would be more appropriate. In regard to NFRs for accessibility, graphics quality, data security, portability and resource consumption, all the results with high means of 4.40 and deviations under 1 indicate that they are essential in developing the application.

B. *Scale Assessment*

To validate the measurements from (Table III) a Principal Component Analysis (PCA) with varimax rotation was used [45]. By Kaiser's criterion, 5 principal components were extracted with Eigen values over the cutoff of 1. The final components chosen based on having at least three variables with a minimal loading of 0.3 [46] where Component 1 (I1, I5, I6, I7) and Component 2 (I9, I10, I11, I12), resulting in a total variance of 37.84%.

TABLE III. FR AND NFR FORM QUESTIONNAIRE

Item Number	Question
I1	How Interested would you be in testing a fully immersive VR application during treatment?
I2	Are you interested in having animals in the game?
I3	Are you interested in having the option to fish in the game ?
I4	Would you prefer music or nature sounds to run in the background?
I5	Do you think that the level of sunlight in a game affects your mood in a positive way?
I6	Do you think that the level of sunlight in real life affects your mood in a positive way?
I7	How important you consider optimizing resource consumption.(How well the game runs on your device).
I8	How important you consider the game having a short learning curve.(to be easy to learn)
I9	How Important is Portability(ability to run on different operation system devices)
I10	How Important is Data Security.
I11	How Important is Accessibility
I12	How Important is the Graphics quality
I13	How would you consider the application having an in depth Game Mechanics(the game and the complexity)

Next, to identify the latent factors measured, an exploratory factor analysis (EFA) was applied by using IBM SPSS v.25 software [47], with a principal axis factor extraction method and varimax rotation. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy had an acceptable value of (KMO = 0.57) using the minimum acceptable value of 0.50 [48], and Bartlett's Test of Sphericity was significant ($p < .001$).

By Kaiser's criterion [48], there were 5 factors with Eigen values above the cutoff of 1; however the procedure could not be completed due to a Haywood case [49]. As advised in IBM SPSS statistics a downside for this criterion may be the overestimation of the number of factors to retain. Most of the present communalities are lower than 0.70, therefore, Kaiser's rule might be inappropriate for this type of data. By Cattell's criterion [50], the scree plot and the previous principal component analysis suggest 2 extraction factors. Therefore, the analysis was executed with 2 fixed extraction factors. According to [51], only the items with a interpreted factor loading with an absolute value more than 0.3 were taken in consideration. Based on this cutoff, three items were not included in the analysis due to having too much uniqueness (I2, I3, I8), and the final structure included Factor 1 (I1, I4, I5, I6, I7) and Factor 2 (I9, I10, I11, I12, I13). The EFA resulted in a 10-item scale, explaining 34.53% of the total variance (Table IV).

Furthermore, the reliability analysis indicated an acceptable internal consistency for Factor 1 ($\alpha = 0.78$) and a modest one for Factor 2 ($\alpha = 0.65$). Although, a sample size of at least 50 is adequate for behavioral sciences, the internal consistency coefficient proved to be sensitive to both sample size and number of items, especially if the Eigen value for the first factor is lower than 6 [51]. Moreover, [52] showed that for sample sizes of 50, the loadings should be around 0.80 for a stable structure. In this case, a higher sample size may be advisable in the future.

TABLE IV. ITEM ANALYSIS AND FACTOR LOADINGS

Rotated Factor Matrix		
	Factor	
	1	2
I5	0.90	-
I6	0.78	-
I7	0.70	-
I1	0.56	0.34
I4	0.40	-
I11	-	0.78
I12	-	0.57
I9	-	0.55
I10	-	0.51
I13	-	0.31
Variance registered after rotation in %	34.53%	

Overall, the findings indicate that an adequate VRT application is required to have a simplified user interface to allow a fast comprehension of the system by the user with a robust framework that provides data security, accessibility and portability. A second important criterion is to have a simple objective, well described with simple game mechanics to avoid confusion. A third trait is immersion characterized by a high graphic quality, as immersion is the main goal of any VRT application. The analysis in this paper provides insight for how future work in this domain should be approached as well as what criterions VRT applications should follow.

IV. CONCLUSION

This article explored the establishment of functional and non-functional requirements for a Virtual Reality application that can be used for pain relief treatments based on the answers obtained from the user questionnaire using voluntary random sampling. The following high priority FRs and NFRs have been observed: immersive environment, presence of an illumination system that simulates daytime lighting, natural background sounds, interactive animal type objects, engaging mechanics, resource optimization, secure data storage, accessibility and improved graphics quality.

For future work, the goal is to implement the learned findings and requirements in a VR application meant for pain therapy. This application will be tested in a real environment where its effects can be observed. The experiment will be followed by a second questionnaire for user experience using the voluntary random sampling method, in order to verify the initial findings concerning the FRs and NFRs as well as its effect on pain relief.

REFERENCES

[1] I. C. Stănică, M. I. Dascalu, A. Moldoveanu, C. N. Bodea, S. Hostiuc, "A survey of virtual reality applications as psychoterapeutic tools to treat phobias", In Proceedings of ELSE 2016 The 12th International Scientific Conference eLearning and Software for Education, 2016, pp. 392-399.

[2] R. C. Popa, D. Gavajiuc, D. Grigore, N. Goga, A. Zaharia, "*A benchmark based on the automatic generation of ontology between the psychological and theological domains*", SSERS International Conference on Emerging Market Trends in Economics, International Relations, Business Management & Social Science Research, 2019, Reference ID: BCAT-09-P28.

[3] A. Bucciol, F. Landini, M. Piovesan, "Unethical behavior in the field: Demographic characteristics and beliefs of the cheater", Journal of Economic Behavior & Organization, 2013, Vol.93, pp. 248–257.

[4] I. A. Bratosin, I. B. Pavaloiu, A. Vasilateanu, N. Goga, G. Dragoi, D. Gavajuc, "Virtual reality therapy for pain", eLearning & Software for Education, 2020, Vol. 3, pp. 178-185.

[5] C.V. Marian, "*Artificial Intelligence Expert System Based on Continuous Glucose Monitoring (CGM) Data for Auto-Adaptive Adjustment Therapy Protocol – How to Make Sensors and Patients to Think Forward and Work Together?*", 2021 International Conference on e-Health and Bioengineering (EHB), 2021, pp.1-4.

[6] I. B. Pavaloiu, M. Y. Mzoughi, A. Vasilateanu, N. Goga, "Learning to drive cars in large cities", The 14th eLearning and Software for Education Conference – eLSE, 2018, Bucharest, Romania, pp 313-319.

[7] K., Mikhail, A. Obukhov, D. Dedov, "Formalization of the burning process of virtual reality objects in adaptive training complexes", Journal of Imaging, 2021, Vol. 7, no. 5: 86. https://doi.org/10.3390/jimaging7050086.

[8] M. Maguire, N. Bevan, "User requirements analysis a review of supporting methods", IFIP World Computer Congress, 2002, pp. 133-148.

[9] L. Lehtola, M. Kauppinen, "Suitability of requirements prioritization methods for market-driven software product development", Software Process: Improvement and Practice, 2006, pp. 7-19, https://doi.org/10.1002/spip.249.

[10] H. Taherdoost. "Sampling methods in research methodology; How to choose a sampling technique for research", International Journal of Academic Research in Management, 2016, Vol. 5, pp. 18-27.

[11] Z. Trost, C. France, M. Anam, C. Shum, "Virtual reality approaches to pain: toward a state of the science", PAIN, 2021, Vol.162, Issue 2 pp. 325-331 doi: 10.1097/j.pain.0000000000002060.

[12] C. G. Ford, E. M. Manegold, C. L. Randall, A. M. Aballay, C. L. Duncan, Assessing the feasibility of implementing low-cost virtual reality therapy during routine burn care, Burns, June 2018, Volume 44, Issue 4, pp 886-895, https://doi.org/10.1016/j.burns.2017.11.020.

[13] S. Benham, M. Kang, N. Grampurohit, Immersive Virtual Reality for the Management of Pain in CommunityDwelling Older Adults, OTJR: Occupation, Participation and Health, December 2018, Vol 39, Issue 2, pp. 90-96, https://doi.org/10.1177/1539449218817291.

[14] R. C. Popa, A. Vasilateanu, A. Doncescu, M. Goga, "*A semantic engine for organizational documents: A quantitative research for requirements definition*", 10th annual International Conference of Education, Research and Innovation, 2017, pp. 3968-3975.

[15] J. Brooke, "SUS: A "Quick and Dirty" Usability Scale", In Usability Evaluation, In Industry. CRC Press, 1996.

[16] M. Sathiyanarayanan, S. Rajan, "MYO armband for physiotherapy healthcare: A case study using gesture recognition application", 8th International Conference on Communication Systems and Networks (COMSNETS), 2016, pp. 1–6, doi.org/10.1109/COMSNETS.2016.7439933.

[17] F. Spyridonis, J. Hansen, T. M. Grønli, G. Ghinea, "PainDroid: An android-based virtual reality application for pain assessment", Multimedia Tools and Applications, 2014, Vol. 72(1), pp. 191–206, doi.org/10.1007/s11042-013-1358-3.

[18] R. Lloréns, E. Noé, C. Colomer, M. Alcañiz, "Effectiveness, usability, and cost-benefit of a virtual reality-based telerehabilitation program for balance recovery after stroke: A randomized controlled trial", Archives of Physical Medicine and Rehabilitation, 2015, Vol. 96(3), pp.418-425.e2. https://doi.org/10.1016/j.apmr.2014.10.019.

[19] A. Ewais, O. D. Troyer, "A usability and acceptance evaluation of the use of augmented reality for learning atoms and molecules reaction by primary school female students in palestine", Journal of Educational Computing Research, 2019, Vol. 57(7), pp. 1643–1670. https://doi.org/10.1177/0735633119855609.

[20] C. Boletsis, "A user experience questionnaire for VR locomotion: Formulation and preliminary evaluation", International Conference on Augmented Reality, Virtual Reality and Computer Graphics, 2020, pp. 157–167. https://doi.org/10.1007/978-3-030-58465-8_11.

[21] B. Klug, "An overview of the system usability scale in library website and system usability testing", Weave: Journal of Library User Experience, 2017, Vol. 1(6). https://doi.org/10.3998/weave.12535642.0001.602.

[22] A. Bangor, P. T. Kortum, J. T. Miller, "An empirical evaluation of the system usability scale", International Journal of Human-Computer Interaction, 2008, Vol. 24: 6, pp. 574-594.

[23] P. Lavrakas, "Encyclopedia of survey research methods", Sage Publications, Inc., 2008, https://doi.org/10.4135/9781412963947.

[24] H. Taherdoost, "Sampling methods in research methodology; How to choose a sampling technique for research", Social Science Research Network. 2016, https://doi.org/10.2139/ssrn.3205035.

[25] J. R. Lewis, "The system usability scale: Past, present, and future", International Journal of Human–Computer Interaction, 2018, Vol. 34(7), pp. 577–590, https://doi.org/10.1080/10447318.2018.1455307.

[26] J. Lewis, "Psychometric evaluation of the PSSUQ using data from five years of usability studies", Int. J. Hum. Comput. Interaction, 2002, Vol. 14, pp. 463–488. https://doi.org/10.1080/10447318.2002.9669130.

[27] F. Rebelo, P. Noriega, E. Duarte, M. Soares, "Using virtual reality to assess user experience", Human Factors, 2012, Vol. 54(6), pp. 964–982. https://doi.org/10.1177/0018720812465006.

[28] J. Burge, D. Brown, "NFRs: Fact or fiction?", 2002.

[29] F. Manola, "Providing systemic properties (Ilities) and quality of service in component-based systems", Institute for Defense Analyses, 1999, URL: http://www.objs.com/aits/9901-iquos.html.

[30] F. Khan, S. R. Jan, M. Tahir, S. Khan, F. Ullah, "Survey: Dealing non-functional requirements at architecture level", VFAST Transactions on Software Engineering, 2016, Vol. 4(1), pp. 27–33, https://doi.org/10.21015/vtse.v9i2.410.

[31] O. Stamm, R. Dahms, U. Müller-Werdan, "Virtual reality in pain therapy: A requirements analysis for older adults with chronic back pain", Journal of NeuroEngineering and Rehabilitation, 2020, Vol. 17, Artiicle no. 129, https://doi.org/10.1186/s12984-020-00753-8.

[32] A. Holzinger, G. Searle, T. Kleinberger, A. Seffah, H. Javahery, "Investigating usability metrics for the design and development of applications for the elderly", In Computers Helping People with Special Needs Springer, 2008, pp. 98–105. Springer. https://doi.org/10.1007/978-3-540-70540-6_13.

[33] J. Mazurek, P. Kiper, B. Cieślik, S. Rutkowski, K. Mehlich, A. Turolla, J. Joanna Szczepańska-Gieracha, "Virtual reality in medicine: a brief overview and future research directions", Human Movement, 2019, Vol. 20(3), pp.16-22. doi:10.5114/hm.2019.83529.

[34] T. Joda, G.O. Gallucci, D. Wismeijer, N.U. Zitzmann, "Augmented and virtual reality in dental medicine: A systematic review", Computers in Biology and Medicine, 2019, Vol. 108, pp. 93-100. https://doi.org/10.1016/j.compbiomed.2019.03.012.

[35] M. Costa, L. P. Vieira, E. O. Barbosa, L. Mendes Oliveira, P. Maillot, C. A. Ottero Vaghetti, M. Giovani Carta, S. Machado, V. Gatica-Rojas, R. S. Monteiro-Junior, "Virtual reality-based exercise with exergames as medicine in different contexts: A short review", Clinical practice and epidemiology in mental health: CP & EMH, 2019, Vol. 15, pp. 15–20. https://doi.org/10.2174/1745017901915010015.

[36] G. Riva, B. K. Wiederhold, F. Mantovani, "Neuroscience of virtual reality: From virtual exposure to embodied medicine", Cyberpshychology, Behaviour, and Social Networking, 2019, Vol. 22, pp. 82-96, DOI: 10.1089/cyber.2017.29099.gr.

[37] J. S. Yoo, D. S. Patel, N. M. Hrynewycz, T. S. Brundage, K. Singh, "The utility of virtual reality and augmented reality in spine surgery", Ann Transl Med., 2019; Volume 7 Suppl 5: S171, doi:10.21037/atm.2019.06.38.

[38] A. Elmi-Terander, R. Nachabe, H. Skulason, "Feasibility and accuracy of thoracolumbar minimally invasive pedicle screw placement with augmented reality navigation technology", Spine (Phila Pa 1976), 2018, Vol. 43, pp. 1018-1023. 10.1097/BRS.0000000000002502.

[39] J. Gasco, A. Patel, J. Ortega-Barnett, "Virtual reality spine surgery simulation: An empirical study of its usefulness", Neurol Res, 2014, Vol. 36, pp. 968-973. 10.1179/1743132814Y.0000000388.

[40] G. Ahlberg, L. Enochsson, A. G. Gallagher, L. Hedman, C. Hogman, D. A. McClusky, S. Ramel, C. D. Smith, D. Arvidsson, "Iconography: proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies", Am J Surg. 2007, Vol. 193, pp.797–804.

[41] A. M. Digioia, D. A. Simon, B. Jarama, M. Blackwel, E. Morgan, R. V. O'Toole, B. Colgan, E. Kischel, "HipNav: preoperative planning and intra-operative navigational guidance for acetabular implant placement in total hip replacement surgery", Proc Computer Assisted Orthopaedic Surgery Symp. 1995.

[42] N. E. Seymour, A. G. Gallagher, S. A. Roman, M. K. O'Brien, V. K. Bansal, D. K. Andersen, R. M. Satava, "Virtual reality training improves operating room performance: results of a randomized, doubleblinded study", Ann Surg., 2002, Vol. 236, pp. 458–463.

[43] B. Spiegel, G. Fuller, M. Lopez, T. Dupuy, B. Noah, A. Howard, M. Albert, V. Tashjian, R. Lam, J. Ahn, F. Dailey, B. T. Rosen, M. Vrahas, M. Little, J. Garlich, E. Dzubur, W. IsHak, I. Danovitch, "Virtual reality for management of pain in hospitalized patients: A randomized comparative effectiveness trial", PLoS One, 2019 Aug 14; Vol. 14(8):e0219115. doi: 10.1371/journal.pone.0219115. PMID: 31412029; PMCID: PMC6693733.

[44] E. Honzel, S. Murthi, B. Brawn-Cinani, G. Colloca, C. Kier, A. Varshney, L. Colloca, "Virtual reality, music, and pain: developing the premise for an interdisciplinary approach to pain management", Pain, 2019, Vol. 160(9), pp. 1909–1919. https://doi.org/10.1097/j.pain.0000000000001539.

[45] D. L. Streiner, "Figuring Out Factors: The Use and Misuse of Factor Analysis." The Canadian Journal of Psychiatry, 1994, Vol. 39(3), pp. 135–140. doi:10.1177/070674379403900303.

[46] H. F. Kaiser, & Rice, J. "Little Jiffy, Mark Iv." Educational and Psychological Measurement, 1974, Vol. 34(1), pp. 111–117. doi:10.1177/001316447403400115.

[47] A. Field, (2013). Discovering statistics using IBM SPSS statistics (4th ed.). SAGE Publications.

[48] H. F. Kaiser, "An index of factorial simplicity", Psychometrika, 1974, Vol. 39, pp. 31–36.

[49] H. B. Heywood, "On finite sequences of real numbers", Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 1931, Vol. 134(824), pp. 486–501

[50] R. B. Cattell, "The Scree test for the number of factors", Multivariate Behavioral Research, 1966, Vol. 1(2), pp. 245–276. doi:10.1207/s15327906mbr0102_10.

[51] H. Yurdugül, "Minimum Sample Size for Cronbach's Coefficient Alpha: A Monte-Carlo Study", Hacettepe University Journal of Education, 2008, Vol. 35, pp. 397-405.

[52] E. Guadagnoli, W. F. Velicer, "Relation of sample size to the stability of component patterns.", Psychological Bulletin, 1988, Vol. 103(2), pp. 265–275. https://doi.org/10.1037/0033-2909.103.2.265.

Integration of Convolutional Neural Networks and Recurrent Neural Networks for Foliar Disease Classification in Apple Trees

Disha Garg, Mansaf Alam
Computer Science Department
Jamia Millia Islamia
New Delhi, India

***Abstract*—Automated methods intended for image classification have become increasingly popular in recent years, with applications in the agriculture field including weed identification, fruit classification, and disease detection in plants and trees. In image classification, convolutional neural networks (CNN) have already shown exceptional results but the problem with these models is that these models cannot extract some relevant image features of the input image. On the other hand, the recurrent neural network (RNN) can fully exploit the relationship among image features. In this paper, the performance of combined CNN and RNN models is evaluated by extracting relevant image features on images of diseased apple leaves. This article suggested a combination of pre-trained CNN network and LSTM, a particular type of RNN. With the use of transfer learning, the deep features were extracted from several fully connected layers of pre-trained deep models i.e. Xception, VGG16, and InceptionV3. The extracted deep features from the CNN layer and RNN layer were concatenated and fed into the fully connected layer to allow the proposed model to be more focused on finding relevant information in the input data. Finally, the class labels of apple foliar disease images are determined by the integrated model for apple foliar disease classification, experimental findings demonstrate that the proposed approach outperforms individual pre-trained models.**

Keywords—Artificial intelligence; deep learning; transfer learning; convolutional neural network; recurrent neural network; LSTM

I. Introduction

Agriculture began in neolithic era and has continued to this day. However, the situation of agriculture is deteriorating day by day as a result of harsh variations in climatic conditions and human malpractice. Diseases in trees, animals, and grains are dangerous and can kill agriculture if not detected early. Various plant diseases can spoil the surrounded plants and trees. But due to a lack of knowledge and expertise, it is not possible to diagnose such diseases. As a result, experts are doing everything possible to assist farmers in overcoming these diseases.

Apple (Maluspumila) is commercially the most important fruit and is fourth among the most widely produced fruits in the world [1]. The growth of apple trees is affected by the assault of numerous foliar diseases such as scab, black rot, and rust [2]. Because of resemblance in appearance in the early stages of tree growth, correctly classifying and identifying all forms of foliar diseases in apple trees is a difficult challenge for farmers. Early disease identification is important for effective and timely disease control. Incorrect and/or delayed diagnosis may result in disease transmission because small insects can damage the whole tree and rapidly it can become a larger and more costly issue.

The manual classification of these diseases can be unnecessarily cumbersome, time-consuming, and costly. Speeding up this method has many advantages, including lower costs, reduced effort. Various plant disease classification systems are designed to support non-expert and non-botanist users in automatically distinguishing plant diseases. Because of this, cost and time can be reduced with accurate identification. Machine learning (ML) [3][4] and deep learning (DL) [5-11] have received a lot of interest in recent years for automatic disease and pest detection in trees and plants. Usama Mokhtar et al. [3] implemented SVM-based disease detection of tomato leaves. They used the dataset of 800 healthy and unhealthy leaves for disease classification. Authors performed feature extraction using Grey-Level Co-occurrence matrix (GLCM and achieved a classification accuracy of 99.83%. Shima Ramesh et al. [4] implemented plant disease detection, and they achieved the highest classification accuracy (70.14%) using the random forest technique among the various machine learning algorithms. These machine learning techniques use manual feature extraction, but deep learning algorithms solve this problem effectively.

For automatically detecting crop diseases, convolutional neural networks (CNNs) have become common. Moreover, numerous studies on the identification of apple diseases also have been published in the literature using pre-trained CNN models since 2017. Table I summarizes the contents of these articles.

The problem with these models is that these models cannot extract some relevant image features of the input image. In this paper, the performance of combined CNN and RNN models is evaluated by extracting relevant image features. For this, transfer learning was used in which the pre-trained weight parameters on the ImageNet dataset are transferred to CNN and combined with the extracted features of RNN in the model.

TABLE I. STUDIES TO CLASSIFY APPLE LEAVE DISEASE USING PRE-TRAINED CNN MODEL

Reference	Task	Dataset	Method	Accuracy	Pros and Cons
(Peng Jiang et al. 2019) [1]	Real time apple leaf diseases classification	26,377 images of apples diseased leaves	VGG-INCEP	97.14 %	Detects disease spots of various sizes on the same leaf
(Yong Zhong, Ming Zhao 2020) [12]	Identify apple leaf diseases	2462 images of 6 apple leaf diseases	DenseNet-121	93.71 %	Cannot overcome the impact of data imbalance
(Xiaofei Chao et al. 2020) [13]	Early diagnosis and accurate identification of apple tree leaf diseases	2970 images of diseased and healthy leaves	Xception and DenseNet (XDNet)	98.82 %	The diversity of dataset is not enough
(GuanWang et al. 2017) [15]	Image-Based Plant Disease Severity Estimation	50,000 images of healthy and diseased crops-38 class labels	VGG16 trained with transfer learning	90.4%	Lacking number of parameters to compare the results
(Bin Liu et al. 2017) [16]	Identify apple leaf diseases	13,689 images of diseased apple leaves	AlexNet, image generation technique	97.62 %	Image generation technique can enhance robustness of CNN model

Contributions of the CNN-RNN combined network are as follows:

- To classify leaf diseases, Pretrained CNN models and RNN were used.
- The performance of combined network was evaluated.
- The suggested integrated model performs better than pre-trained networks, according to the findings of the experiments.

The remaining part of the article is laid out as follows: Section 2 presents a review of current scholarly publications relevant to this article. Section 3 contains the methodology, dataset collection and description, data preprocessing, an integrated network, and its layers. In Section 4, the training and fine-tuning of an integrated model is described. Section 5 contains the experimental setup as well as the model classification findings. In Section 6, the proposed model is compared with existing techniques. Section 7 concludes the paper.

A. *Motivation*

When the current work is examined on image classification, it was observed that researchers have paid a lot of attention to combining multiple models. Deep models with multi-modality perform better in terms of classification accuracy, precision, recall, and f1 scores. JananiVenugopalan et al. used a combination of models to diagnose Alzheimer's disease at an early stage[17].Md. Zabirul Islam et al. [18] used X-ray images to combine CNN and LSTM for the identification of a novel coronavirus (COVID-19).Aydin Kaya et al. [19] used the combination of RNN and CNN models for plant classification. It is also found that deep feature concatenation is extremely efficient. This sparked the interest in combining RNN and CNN, as well as the notion of concatenating features from pre-trained CNN and RNN models.

II. RELATED WORK

Many pieces of research on the identification of plant diseases have been carried out so far. Genetic algorithms [20] [21], artificial neural networks [22], Bayes classifiers [23], and fuzzy logic [24] have been used in previous studies to identify and classify plant leaf diseases having a higher level of accuracy.

DL has been extensively researched in recent years for disease and pest identification in plants [9] [10] [25] [26]. Previous research has also shown that CNN is ideally suited for high-accuracy detection and diagnosis of plant diseases automatically. Some of these studies used CNN-based models that had been pre-trained, while others created their models based on CNN or pre-trained models.

Alvaro Fuentes et al. (2017) [5] proposed a faster region-based CNN for tomato diseases and pests recognition in real-time. Authors utilized a huge data-set for testing and training, comprising of 5000 images to develop these systems. These systems are capable of detecting nine distinct diseases and pests. The authors used a faster RCNN, but this model has a poor identification rate and a lot of pattern variability in certain disease groups.

Belal A. M. Ashqar and Samy S. Abu Naser (2018) [6] used deep learning approach to develop tomato leaves disease detection system. They used 9000 images of diseased and healthy tomato leaves for training of deep CNN to classify 5 diseases. They achieved 99.84 % accuracy. To perform detection and diagnosis of diseases in plants, Konstantinos P. Ferentinos (2018) [7] developed a convolutional neural network model using diseased and healthy 87,848 images of plant leaves. They trained various models. Among all CNN models, VGG achieved an accuracy of 99.53% with the best performance. The authors used the VGG model on diseased images and had a decent success rate, but this model is far from being a generic method.

Aravind Krishnaswamy Rangarajan et al. (2018) [8] have performed the classification of disease in tomatos by using pre-trained DL and its algorithm. Authors have used AlexNet and VGG16 pre-trained architectures on PlantVillage consisting of 13262 images and achieved 97.29% classification accuracy for VGG16 net and 97.49% for Alex-Net. However, other transfer learning models need to be evaluated for this dataset.

Additionally, Geetha ramani G. and Arun Pandian J (2019) [9] proposed a model to identify diseases in plants using Deep

CNN. For training, they used a dataset that included 39 different plant leaf classes as well as background images. Authors used the following 6 types of data augmentation methods: principal component analysis (PCA), gamma correction, flipping, noise injection, color augmentation, scaling, and rotation. They achieved a classification accuracy of 96.46% for this proposed model. Despite their high success rate, authors were unable to completely exploit important image features using deep CNN.

Moreover, Edna Chebet Too et al. (2018) [10] performed the comparison of InceptionV4, VGG16, ResNet with 101, 152, 50 layers, and Dense-Nets with layer number as 121. Among all structural designs, Dense-Nets achieved the testing accuracy of 99.75 percentages. With an increasing number of epochs, DenseNets has the potential to increase accuracy without overfitting. However, computational time can be improved.

TABLE II. PREVIOUS STUDIES BASED ON DEEP LEARNING FOR THE OTHER PLANT SPECIES

Author	Plant species	Model used	Best Results	No. of images
Mehmet MetinOzguven, Kemal Adem [27]	Sugar beet	Faster R-CNN	92.89%	155
Aravind-Krishna swamy Rangarajan, Raja Purushothaman, Aniirudh Ramesh [8]	Tomato	AlexNet VGG16 net	97.49% 97.29%	13,262
Geetharamani G., ArunPandian J [9]	Different plant disease	Deep CNN	96.46%	61,486
Karthik R et. al. [28]	Tomato	Attention embedded residual CNN	98%	95999
Sladojevic et al. [20]	Plant leaves	CNN based model	96.3%	2 589
Honghua Jiang et.al [29]	weed and crop recognition	GCN-ResNet-101	97.8%	6000
Gensheng et al. [26]	tea leaf	VGG16	90%	4 980
Ashqar et al. [6]	Tomato	Deep CNN	99.84%	9000
Jihen Amara et al. [11]	Banana	LeNet	98.61%	3700
Srdjan Sladojevic et al. [30]	Different plant disease	deep CNN	96.3%	30880
Sharada P. Mohanty et al. [31]	Different crops	GoogLeNet	99.35%	54306

One of the most significant gaps of the current studies in the field of plant disease detection is a significant decrease in classification performance [12, 15, 26] of the models on real images collected in fields compared to images taken in a controlled environment so there should be availability of large public datasets [14] [27]. Table II summarizes the findings of related investigations.

III. METHODOLOGY

The overall method for detecting foliar diseases in an apple tree dataset is represented in Fig. 1 in several phases. The apple foliar disease dataset was utilized as the training data for the CNN-RNN network. The training accuracy, validation accuracy, training loss, and validation loss were calculated at each epoch. Confusion matrix, accuracy, precision, recall, and f1-score were used to evaluate the performance of the proposed system. It is also observed the minimum training and minimum validation loss and corresponding accuracies.

A. Dataset

High-resolution and real-life symptom images of different apple foliar diseases have been manually captured with angles, illumination, noise, and surfaces. This dataset for the Plant-Pathology Challenge is made available and open to the group through "https". It's an integral component of FGVC (Fine-grained visual categorization)which is a workshop at CVPR(Computer -Vision and Pattern- Recognition) 2020.This dataset has four types of classes (a) Apple scab, (b) Black rot (c) Cedar apple rust (d) healthy. Sample images of each class are given in Fig. 2. Proposed model, however, is not limited to this dataset and can be used on a variety of plant disease datasets. Table III shows the count of total images in the dataset, which has previously been separated into the train set and test set.

B. Preprocessing

Dataset is enhanced to increase the model accuracy and reduce over-fitting. Apple foliar disease dataset contains RGB images of arbitrary sizes. All of the images were scaled to a resolution of 299 by 299 pixels and to be compatible with the initial values of the network, all pixel values were divided by 255.After that, sample-by-sample normalization was carried out. The efficiency of end-to-end training can be greatly improved by normalization. Finally, the training images are subjected to a variety of random augmentations, such as random rotation, shearing, zooming, and flipping.

C. Development of the Integrated Model

A proposed integrated CNN-RNN model is based on pre-trained CNN. The models mainly contain different layers of pre-trained CNN models, one RNN layer/LSTM layer and one softmax layer. The images are re-sized to fit the input layers of the pre-trained models such as Xception, VGG16, and InceptionV3. After this, deep features were extracted from these pre-trained models and then different shape of the output is found for different pre-trained models. For inceptionv3 the output shape is found as (None, 8, 8, 2048). For VGG16, the output shape is received as (None, 9, 9, 512) and for the Xception model (None, 10, 10, 2048) is received as an output shape. The input is provided to the LSTM layer using the reshape technique. After analyzing the time characteristics, a softmax layer/fully connected layer predicted class labels of the input images that were categorized into four categories: scab, black rot, cedar apple rust, healthy.

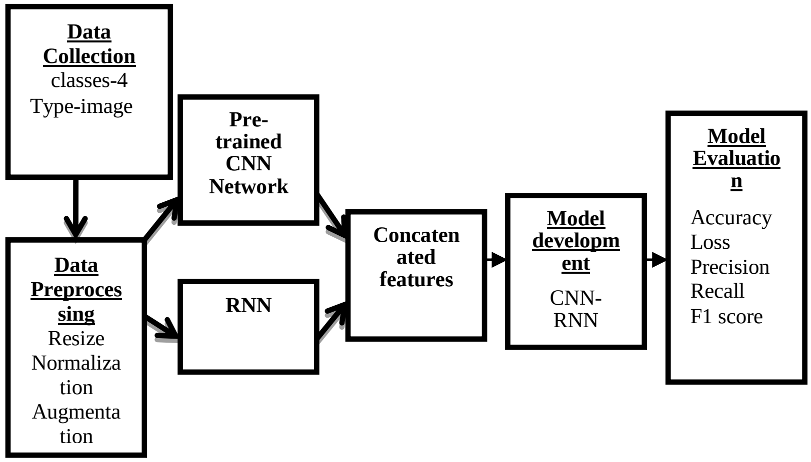


Fig. 1. Proposed System.

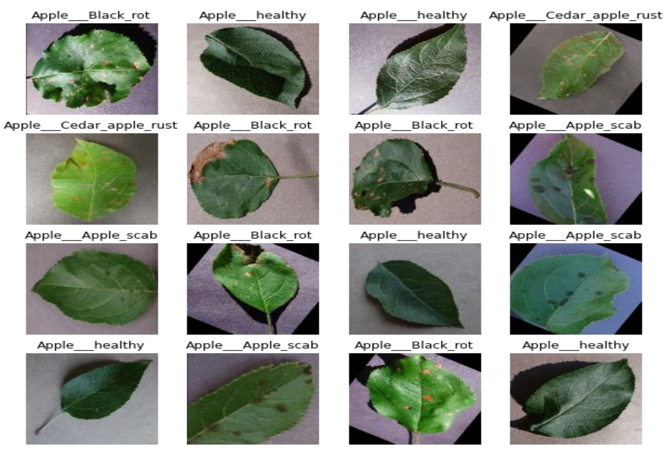


Fig. 2. The Apple Foliar Dataset Sample for Four Classes.

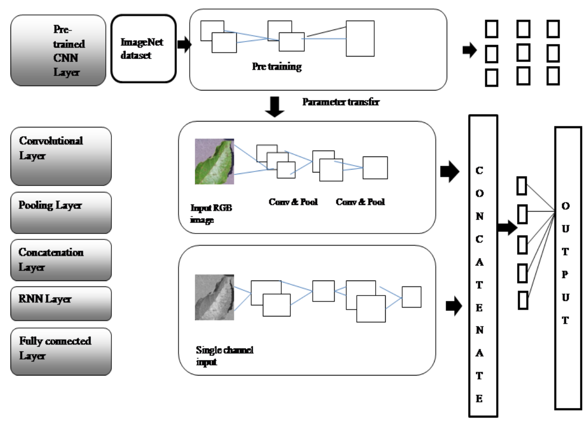


Fig. 3. Integrated Model.

TABLE III. IMAGE COUNT IN THE DATASET

Dataset	Apple scab	Black rot	Cedar apple rust	Healthy	Total
Training	4032	3974	3520	4016	15542
Testing	1008	994	880	1004	3886
Total	5040	4968	4400	5020	19428

RNNs can handle non-sequential data as well as sequential data, according to several pieces of research [32]. Improved RNN models, such as LSTMs allow for long sequence training, conquering issues such as vanishing gradients. Vanishing gradients is a problem where a deep neural network is not able to transmit valuable gradient information from the output to the input end of the model. RNN methods for sequentially processing the data of variable-length and fixed sizes have been presented in a few recent papers [33]. LSTM may be used to collect the images having discriminating regions for fine-grained classification [34]. LSTM (Long Short-Term Memory networks) has a "processor" that determines whether or not the information is useful. The Integrated model is illustrated in Fig. 3 with four layers: Pre-trained CNN layer, RNN layer, concatenation layer, softmax, or fully connected layer.

All layers of an integrated model are briefly described as follows:

1) Pre-trained CNN layer: For initializing weights of CNN model, pre-trained weight parameters are used. ImageNet is used for pre-training and it is a common dataset to develop different architectures. It is broad enough (1.2 million images) to construct a generic model. Transfer learning indicates that these pre-trained networks can generalize to images that are not part of the ImageNet dataset. By fine-tuning the model, changes can be made to it that was not previously possible.

2) Convolution layer: Convolution operations with feature maps are performed with this layer using convolution windows of various sizes. The effect of applying a convolution over an image is a feature map. The row index and column index of resulted matrix is denoted by u and v respectively. The number of weight parameters varies as the window size changes. Convolution and pooling are two types of computations that can be done in the early layers of CNN. The convolution layer has the following operation as in (1):

$$S(u, v) = (M*F)(u, v) = \sum\sum M(u+m, v+n)\, F(m, n) \quad (1)$$

In (1), M indicate the input matrix. F represents the m×n size 2D filter of and S denotes the 2D feature map output. M*F denotes the convolutional layer operation. Rectified linear unit (ReLU) is the activation function which is used after the convolution operation to improve nonlinearity in feature maps. ReLU calculates activation by keeping the threshold at zero. It may be mathematically represented as in (2):

$$f(y) = \max(0, y) \tag{2}$$

3) Pooling layer: To minimize the number of parameters, this layer down samples a given input representation. By using max-pooling which is the most frequent approach, the size of resulted image can be reduced by retaining image information. As a convolution layer, the pooling layer also uses filters of different sizes.

4) RNN/LSTM Layer: An input layer, a hidden layer, and an output layer are present in RNN and CNN. The most significant aspect of RNN is the interaction of these hidden layers. The input and hidden layer nodes are linked. Output layer takes the outcome of the hidden layer. Hidden layer receives information from the output layer node. There exist various adjacent hidden layer nodes.

In this article, LSTM is used which is a form of RNN. LSTM is made up of memory blocks and these blocks are recurrently connected blocks. Each has 3 multiplicative gates (input, output, forget gate) and one or two memory cells that are recurrently connected. The input gate monitors how much data it should read, forget gate monitor whether it should forget the value of the current cell, and the output gate monitor whether it should output the new cell value.

5) Concatenation layer: The concatenation layer concatenates features derived from the CNN and RNN. Feature concatenation is a powerful technique for combining many features to improve classification accuracy.

6) Softmax layer or fully connected layer: The features of RNN and CNN are concatenated and forwarded to the softmax layer, which generates the input image class labels.

D. Long Short-term Memory (LSTM)

A cell with three gates (input, output, and forget) makes up an LSTM unit. LSTM can choose which information is lost and which is remembered. An LSTM layer is made up of memory blocks, which are recurrently connected blocks. It can be assumed that these blocks are a differentiable variant of a computer's memory chips. Every block has one or more than one recurrently linked memory cells, as well as 3 multiplicative units. These units continuously provide analogs of reading, Reset and Write operations for the cells. The following are the gate equations:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{3}$$

$$f_t = \sigma\left(W_f[h_{t-1}, x_t] + b_f\right) \tag{4}$$

$$O_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{5}$$

$$gt = tanh(W_i[h_{t-1}, x_t] + b_i) \tag{6}$$

i_t,f_t, and o_t represent input gate, forget gate, and output gate respectively. σ denotes the sigmoid function. The weight for the respective gate(x) neurons is represented by W_x. The output of the previous LSTM block (at timestamp t-1) is h_{t-1}. x_t is the input at the current timestamp. bx is the bias for the corresponding gate(x). In (3), h_{t-1} and x_t are passed through the sigmoid layer and used to decide that which part of the `information is needs to be added. (6) Obtains new information when h_{t-1} and x_t are passed through the tanh layer in (8):

$$c_t = i_t g_t + f_t c_{t-1} \tag{7}$$

$$h_t = o_t tanh(c_t) \tag{8}$$

The new and previous cell states are denoted by ct and c_{t-1}, respectively.g_t is the current information. g_t and C_{t-1} are combined in (7). The final output O_t is calculated in (5) and it gets multiplied with c_t which transfers new information through the tanh layer.

IV. CNN-RNN Network Training and Fine-tuning

To classify apple foliar diseases, the combination of CNN models and RNN was used. There were a training phase and a testing phase in the proposed method illustrated. In the training process, firstly on the ImageNet dataset, CNN model was pre-trained and the parameters of this pre-trained network were used for initializing the new CNN model, this is called fine-tuning. When the entire layers of CNN have been frozen, the RNN model was trained. After the completion of training, layers of the CNN model were unfrozen and the whole CNN-RNN model was trained.

Test images that were pre-processed were fed into the CNN-RNN model. By using the softmax layer, the classification results were obtained. In this experiment, the proposed model is built on two layers: Conv and pool. These are the first few layers of pre-trained models, already trained on ImageNet dataset. In this proposed model, the above two layers were moved to the same place to allow the transmission of features. The CNN transfer learning model uses RGB input, whereas RNN uses a single-channel input. The aforementioned layers were trained on jointly on the apple foliar disease dataset. The training process was terminated after predetermined epochs.

V. Experimental Results

The proposed work is implemented in two parts:

Part 1: Implementing transfer learning using CNN

Part 2: Implementing RNN

RNN uses randomly initialized parameters while CNN uses pre-trained weight parameters. During the training phase, the gradient of the cross-entropy loss function is used to iteratively update these weights. Firstly, all layers of CNN were frozen then RNN and the last classification layer of CNN were trained. Then training samples are calculated with the help of an RMSProp optimizer. Then we defrosted all layers of CNN and the whole CNN-RNN model was trained then Adam optimizer was used to calculate the samples of training. For the whole network, 0.0001 was used as a learning rate.

Since the problem is multi-class, the categorical cross-entropy/softmax loss function is used. Keras and TensorFlow2.5.0with python 3 were used on an Intel(R) Core(TM) i5-2.9 GHz CPU for implementing models. Furthermore, we were carried out the experiments on an NVIDIA Tesla T4 GPU with 2.32 GB of RAM.

A. Evaluation Indices

The following indices were used to compare the performance of the individual and integrated systems: True Positive (TP): The class of interest was correctly categorized. True Negative (TN): Classified correctly as not the class of interest. False Positive (FP): The class of interest was incorrectly categorized. False Negative (FN): Not the class of interest was incorrectly categorized. Accuracy is represented in (9), Precision is represented in (10), Recall is represented in (11) and F1 score is represented in (12). Mathematical expression of all these metrics is represented below:

$$\text{Accuracy} = \frac{(TP+TN)}{(TN+FP+TP+FN)} \tag{9}$$

$$\text{Precison} = \frac{(TP)}{(TP+FP)} \tag{10}$$

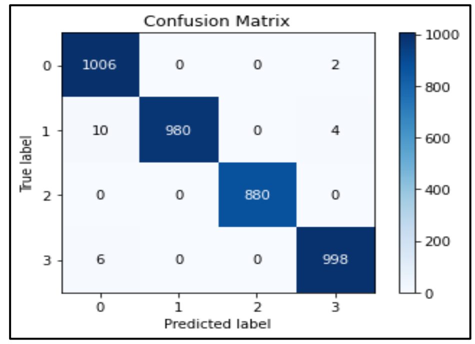
$$\text{Recall} = \frac{TP}{(TP+FN)} \tag{11}$$

$$\text{F1 score } \frac{2*TP}{(2*TP+FP+FN)} \tag{12}$$

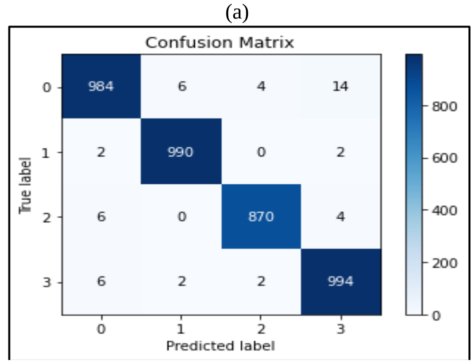
B. Result Analysis and Discussion

Fig. 4 illustrates the confusion matrix of the pre-trained CNN model and the proposed CNN-RNN models for apple foliar disease classification. Fig. 4(a) represents Xception-LSTM, Fig. 4(b) represents Xception, Fig. 4(c) represents VGG16-LSTM, Fig. 4(d) represents VGG16, Fig. 4(e) represents InceptionV3-LSTM and Fig. 4(f) represents Inception V3. In the confusion matrix, apple_scab is represented as class 0, black_ rot is represented as class 1, cedar_ apple_ rust is represented as class 2, and healthy leaves are represented as class 3. The Xception model misclassified 44 of 3886 test images, whereas the proposed Xception-LSTM model misclassified 22 of them. Among the 3886 test images, 266 were misclassified by the VGG16 model, whereas the proposed VGG16-LSTM model misclassified 45 of them. Similarly, 60 were misclassified by the InceptionV3 model, whereas the proposed InceptionV3-LSTM model misclassified a total of 14 images. The proposed integrated CNN-RNN networks outperformed the individual pre-trained CNN networks in terms of TP and TN values, as well as reduced FN and FP values. Table IV shows the observed results of pre-trained CNN models and CNN-RNN models in terms of minimum training loss with corresponding training accuracy, and minimum validation loss with corresponding validation accuracy and Xception-LSTM accuracy and loss are depicted in Fig. 5(a) and 5(b).

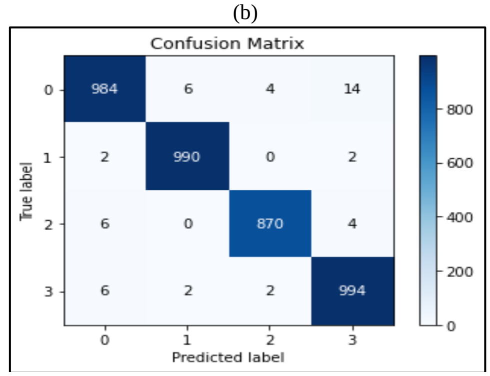
Fig. 6(a) and 6(b) depict accuracy and Xception loss, respectively. VGG16-LSTM accuracy and loss are depicted in Fig. 7(a) and 7(b). VGG16 accuracy and loss are depicted in Fig. 8(a) and 8(b). InceptionV3-LSTM accuracy and loss are depicted in Fig. 9(a) and 9(b). InceptionV3 accuracy and loss are depicted in Fig. 10(a) and 10(b). Performance evaluation for pre-trained CNN models and CNN-RNN models may be seen in the loss and accuracy figures.



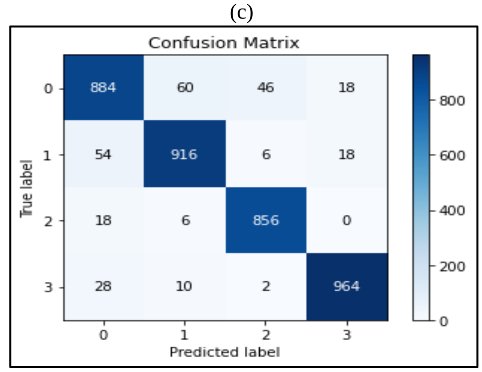
(a)



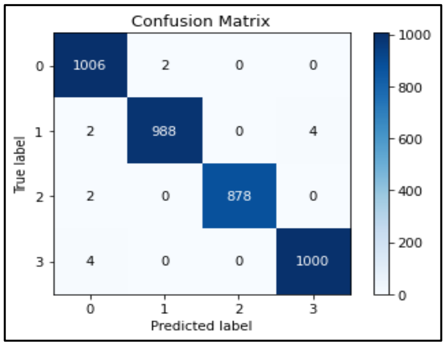
(b)



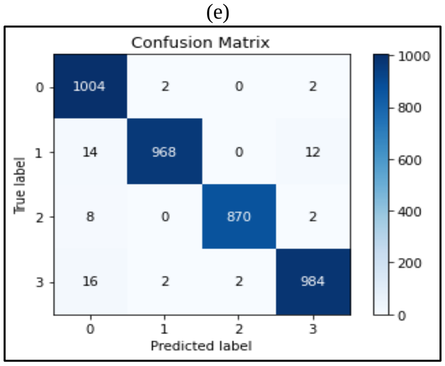
(c)



(d)



(e)



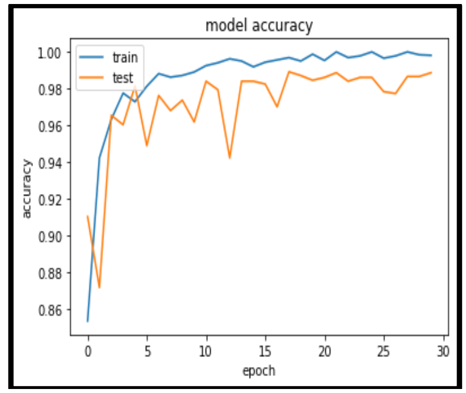
(f)

Fig. 4. (a) Xception-LSTM (b) Xception (c) VGG16-LSTM (d) VGG16 (e) InceptionV3-LSTM (f) InceptionV3.

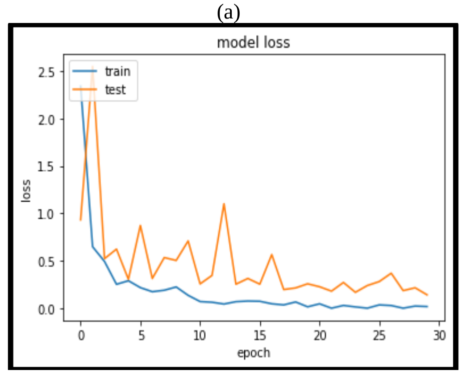
TABLE IV. COMPARISON OF INTEGRATED MODELS AND INDIVIDUAL MODELS IN TERMS OF MINIMUM LOSS AND CORRESPONDING TRAINING AND VALIDATION ACCURACY

Model	Training		Validation	
	Loss	**Accuracy**	**Loss**	**Accuracy**
Xception	0.0144	99.87	0.1408	98.86
Xception-LSTM	0.001	99.94	0.0111	99.69
VGG16	0.0345	98.87	0.206	93.08
VGG16-LSTM	0.0028	99.94	0.0468	99.07
InceptionV3	0.0464	99.59	0.3042	98.24
InceptionV3-LSTM	0.0011	99.94	0.0027	99.9

Among the total number of 30 epochs, Minimum training loss and corresponding training accuracy of the Xception were found at .0144 and 99.87% respectively at epoch 20 whereas .0010 and 99.94 for Xception-LSTM model at epoch 10. The minimum validation loss and corresponding validation accuracy of Xception model were found .01408 and 98.86% respectively at epoch 30 whereas .0111 and 99.69 for the Xception-LSTM model at epoch 13. Similarly, the minimum training loss and corresponding training accuracy of the VGG16 model were found at .0345 and 99.87% respectively at epoch 27 whereas .0028 and 99.94 for the VGG16-LSTM model at epoch 22. The minimum validation loss and corresponding validation accuracy of the VGG16 model were found at .206 and 93.08% respectively at epoch 25 whereas .0468 and 99.07 for the VGG16-LSTM model at epoch 29. Minimum training loss and corresponding training accuracy of the InceptionV3 model were found .0464 and 99.59% respectively at epoch 19 whereas .0011 and 99.94 for InceptionV3-LSTM model at epoch 12. The minimum validation loss and corresponding validation accuracy of the InceptionV3 were found .3402 and 98.24% respectively at epoch 10 whereas .0027 and 99.9 for InceptionV3-LSTM model at epoch 10.

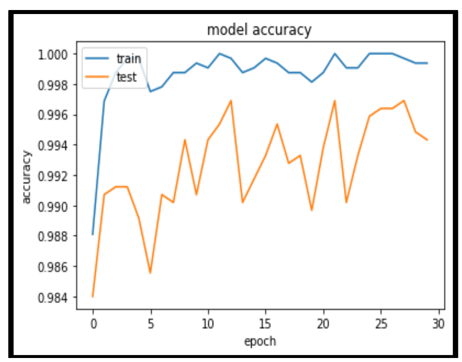


(a)

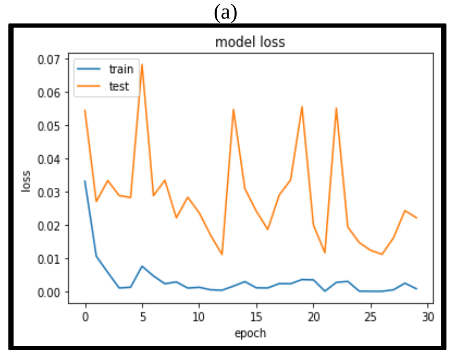


(b)

Fig. 5. (a) Accuracy (Xception-LSTM) (b) Loss (Xception-LSTM).

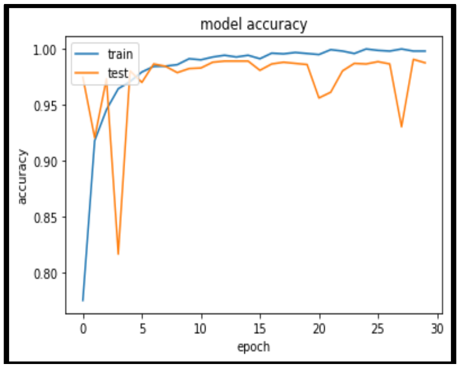


(a)

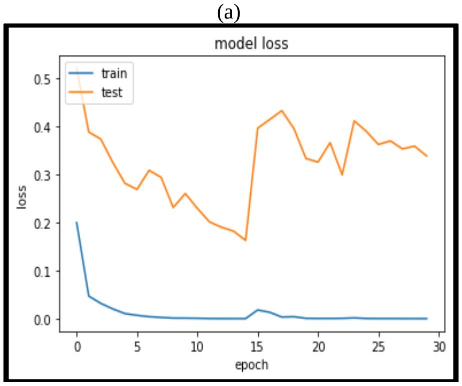


(b)

Fig. 6. (a) Accuracy (Xception) (b) Loss (Xception).



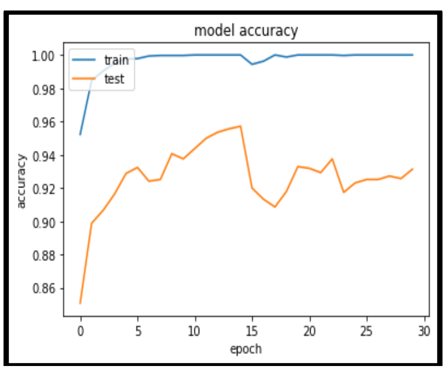
(a)



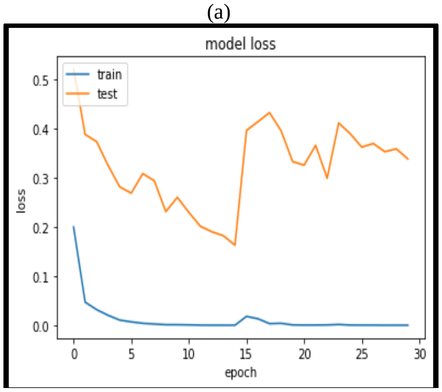
(b)

Fig. 7. (a) Accuracy (VGG16-LSTM), (b) Loss (VGG16-LSTM).

Results presented in Table IV shows that combined models have given the highest accuracy when the loss was minimum as compared to individual model. It can be seen in both the cases, training results as well as validation results.

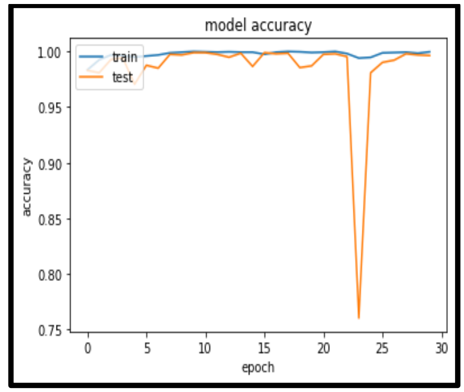


(a)

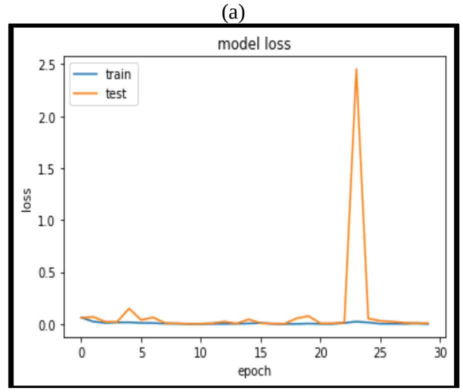


(b)

Fig. 8. (a) Accuracy (VGG16) (b) Loss (VGG).

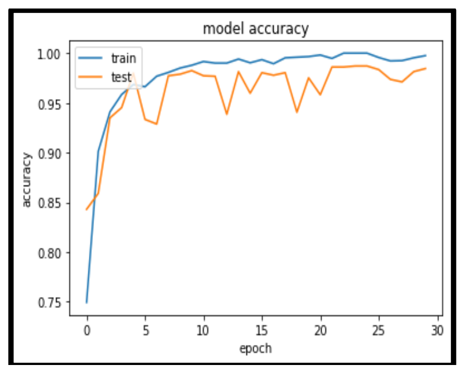


(a)

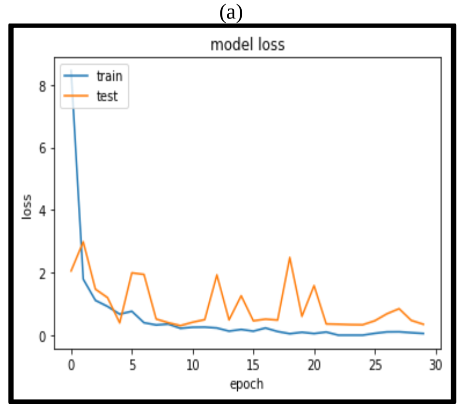


(b)

Fig. 9. (a) Accuracy (InceptionV3-LSTM) (b) Loss (InceptionV3-LSTM).

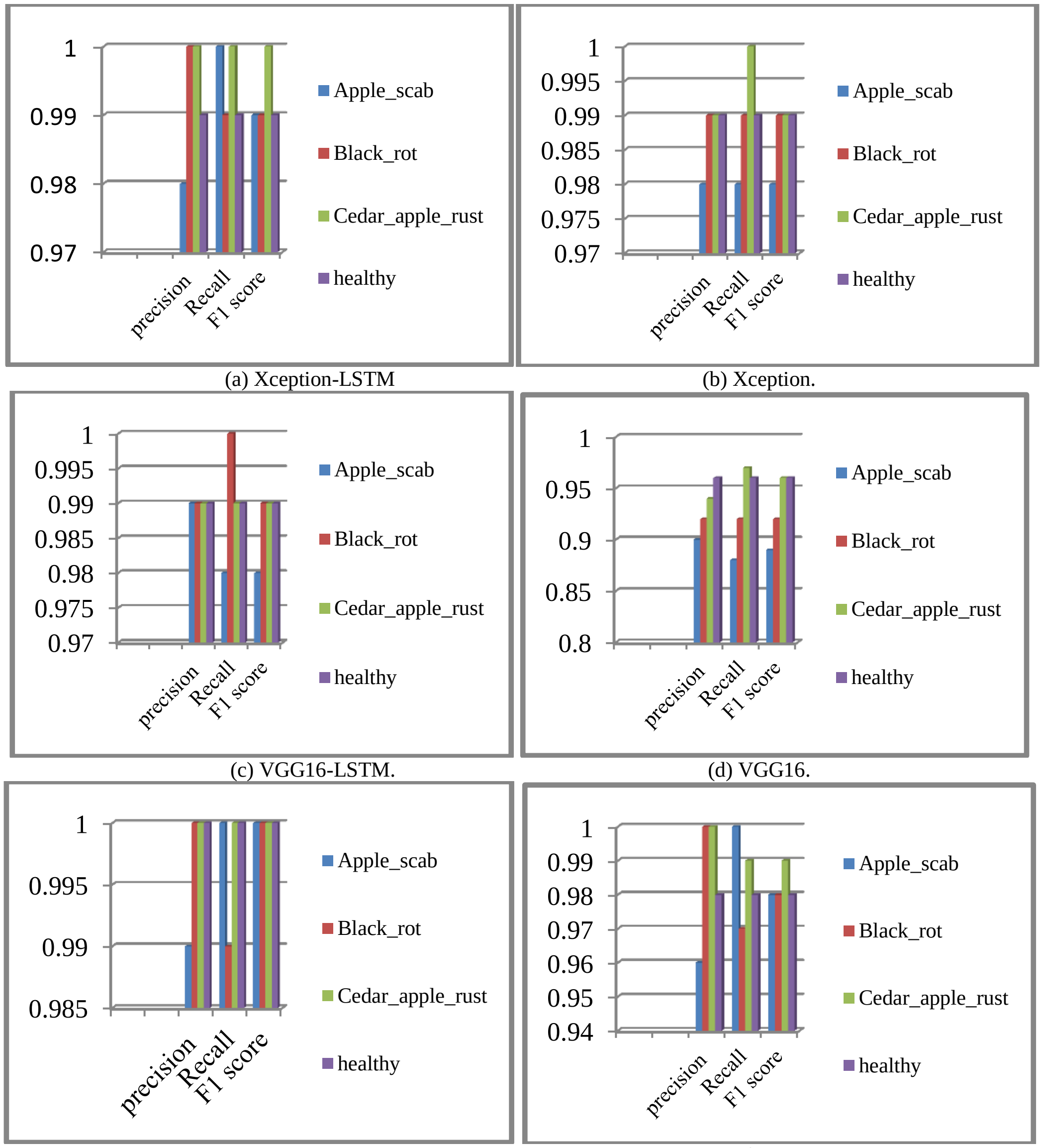


(a)



(b)

Fig. 10. (a) Accuracy (InceptionV3) (b) Loss (InceptionV3).



(a) Xception-LSTM (b) Xception.

(c) VGG16-LSTM. (d) VGG16.

(e) InceptionV3-LSTM. (f) InceptionV3.

Fig. 11. Precision, Recall, and f1 Score of CNN-LSTM Models and pre-trained CNN Model.

Fig. 11 depicts precision, recall, and f1 score findings. As per the results, integrated models achieved good score in terms of precision, recall, and F1 score as compared to individual Pretrained CNN models. Xception scored 98.75% precision, 99 % recall, and 98.75% f1 score for all four classes, whereas Xception-LSTM achieved 99.25 % precision, 99.5 % recall, and 99.25% f1 score for all four classes. Similarly, VGG16 model achieved 93% precision, 93.25% recall and 93.25% recall and VGG16-LSTM model achieved 98.75% precision, 99% recall and 98.76% f1 score. InceptionV3 achieved 98.5% precision, 98.5% recall, 98.25% f1 score and InceptionV3-LSTM achieved 99.75% precision, 99.75% recall, 99.9% f1 score for all the classes.

VI. Comparison with Existing Methods

Many deep CNN architectures have recently been suggested for the classification of apple diseases. Earlier, Densenet-121, XDNet (Xception and DenseNet), VGG-INCEP, VGG16 model, AlexNet were used. For comparisons, pre-trained models were randomly selected such as Xception, VGG16, and Inceptionv3 for apple diseases detection. As seen in Table V, the suggested approach outperformed all other pre-trained models. Among three experiments, InceptionV3-LSTM produces the highest accuracy score as 99.8%.The benefits of the suggested integrated model are as follows:

- Pre-trained CNN models were used for reducing training complexity that saved time to train a model.
- The deep features in images can be extracted and classified automatically using proposed system.
- Based on the apple foliar disease dataset, proposed approach performed the best in comparison to current approaches in terms of classification.

TABLE V. A COMPARISON OF THE PROPOSED SYSTEM TO EXISTING METHODS FOR CLASSIFYING APPLE LEAF DISEASES

Author	Deep Learning model	Overall Accuracy
(Yong Zhong, Ming Zhao 2020) [12]	Densenet-121	93.71%
(Xiaofei Chao et al. 2020) [13]	XDNet (Xception and DenseNet)	98.82%
(Peng Jiang et al. 2019) [1]	VGG-INCEP	97.14%
(Marko Arsenovic et al. 2019) [14]	Pretrained CNN models	93.67%
(Saraansh Baranwal et al. 2019) [2]	GoogLeNet	98.42%
(GuanWang et al. 2017) [15]	VGG16model	90.4%
(Bin Liu et al. 2017) [16]	Deep CNN based on AlexNet	97.62%
Proposed work	**Xception-LSTM**	**99.5%**
	VGG16-LSTM	**99%**
	InceptionV3-LSTM	**99.8%**

The main purpose of this research is to achieve good results in detecting foliar diseases in apple trees. The experimental results revealed that the proposed CNN-RNN network outperforms the individual CNN models.

VII. CONCLUSION

To create combined neural network architecture for apple foliar disease image classification, the features of CNN and RNN models are merged in this work. When comparing separate models to the integrated models, the assessment findings reveal that the integrated models generated better outcomes. InceptionV3-LSTM received the maximum accuracy score of 99.8%Second highest accuracy score of 99.5% was found for Xception-LSTM and the lowest accuracy score was found for VGG16-LSTM as 99%.Despite this performance, these integrated models may have some improvement also. Training on 30 epochs was performed. We can increase the number of epochs in the future. Another improvement can be investigating the success of these models for other agricultural applications, such as weed identification and plant classification, other plant disease detection. Other deep neural network models will be utilised in future studies to detect apple leaf illnesses in real time, such as Faster RCNN (Regions with Convolutional Neural Network), YOLO (You Only Look Once), and SSD (Single Shot MultiBox Detector). Furthermore, more forms of apple leaf illnesses and thousands of high-quality natural images of apple leaf diseases must be collected in order to identify more diseases in a timely and efficient manner.

REFERENCES

[1] Jiang, P., Chen, Y., Liu, B., He, D., & Liang, C. (2019). Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks. IEEE Access, 7, 59069-59080.

[2] Baranwal, S., Khandelwal, S., &Arora, A. (2019, February). Deep learning convolutional neural network for apple leaves disease detection. In Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM), Amity University Rajasthan, Jaipur-India

[3] Mokhtar, U., El Bendary, N., Hassenian, A. E., Emary, E., Mahmoud, M. A., Hefny, H., &Tolba, M. F. (2015). SVM-based detection of tomato leaves diseases. In Intelligent Systems' 2014 (pp. 641-652). Springer, Cham.

[4] Ramesh, S., Hebbar, R., Niveditha, M., Pooja, R., Shashank, N., &Vinod, P. V. (2018, April). Plant disease detection using machine learning. In 2018 International conference on design innovations for 3Cs compute communication control (ICDI3C) (pp. 41-45). IEEE.

[5] A. Fuentes, S. Yoon, S.C. Kim, Park D.S., A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition, Sensors 17 (2022) (2017).

[6] B.A.M. Ashqar, S.S. Abu-Naser, Image-based tomato leaves disease detection using deep learning, Int. J. Acad. Eng. Res. 2 (12) (2018).

[7] K.P. Ferentinos, Deep learning models for plant disease detection and diagnosis, Comput. Electron. Agric. 145 (2018) (2018) 311–318.

[8] A.K. Rangarajan, R. Purushothaman, A. Ramesh, Tomato crop disease classification using pre-trained deep learning algorithm, ProcediaComput.Sci. 133 (2018) (2018) 1040–1047.

[9] G. Geetharamani, P.J. Arun, Identification of plant leaf diseases using a nine-layer deep convolutional neural network, Comput. Electr. Eng. 76 (2019) (2019) 323–338.

[10] E.C. Too, L. Yujian, S. Njuki, L. Yingchun, A comparative study of fine-tuning deep learning models for plant disease identification, Comput. Electron. Agric. 161 (2019) (2019) 272–279.

[11] J. Amara, B. Bouaziz, A. Algergawy, A deep learning-based approach for banana leaf diseases classification, BTW Workshop, Stuttgart, 2017, pp.79–88.

[12] Zhong, Y., & Zhao, M. (2020). Research on deep learning in apple leaf disease recognition. Computers and Electronics in Agriculture, 168, 105146.

[13] Chao, X., Sun, G., Zhao, H., Li, M., & He, D. (2020). Identification of Apple Tree Leaf Diseases Based on Deep Learning Models. Symmetry, 12(7), 1065.

[14] Arsenovic, M., Karanovic, M., Sladojevic, S., Anderla, A., &Stefanovic, D. (2019). Solving current limitations of deep learning based approaches for plant disease detection. Symmetry, 11(7), 939.

[15] Wang, G., Sun, Y., & Wang, J. (2017). Automatic image-based plant disease severity estimation using deep learning. Computational intelligence and neuroscience, 2017.

[16] Liu, B., Zhang, Y., He, D., & Li, Y. (2018). Identification of apple leaf diseases based on deep convolutional neural networks. Symmetry, 10(1), 11.

[17] Venugopalan, J., Tong, L., Hassanzadeh, H. R., & Wang, M. D. (2021). Multimodal deep learning models for early detection of Alzheimer's disease stage. Scientific reports, 11(1), 1-13.

[18] Islam, M. Z., Islam, M. M., &Asraf, A. (2020). A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images. Informatics in medicine unlocked, 20, 100412.

[19] Kaya, A., Keceli, A. S., Catal, C., Yalic, H. Y., Temucin, H., &Tekinerdogan, B. (2019). Analysis of transfer learning for deep neural network based plant classification models. Computers and electronics in agriculture, 158, 20-29.

[20] Arya, M. S., Anjali, K., &Unni, D. (2018, January). Detection of unhealthy plant leaves using image processing and genetic algorithm with Arduino. In 2018 International Conference on Power, Signals, Control and Computation (EPSCICON) (pp. 1-5). IEEE.

[21] Singh, V., & Misra, A. K. (2015, March). Detection of unhealthy region of plant leaves using image processing and genetic algorithm. In 2015 International Conference on Advances in Computer Engineering and Applications (pp. 1028-1032). IEEE.

[22] Awate, A., Deshmankar, D., Amrutkar, G., Bagul, U., &Sonavane, S. (2015, October). Fruit disease detection using color, texture analysis and ANN. In 2015 International Conference on Green Computing and Internet of Things (ICGCIoT) (pp. 970-975). IEEE.

[23] Kusumo, B. S., Heryana, A., Mahendra, O., &Pardede, H. F. (2018, November). Machine learning-based for automatic detection of corn-plant diseases using image processing. In 2018 International Conference on Computer, Control, Informatics and its Applications (IC3INA) (pp. 93-97). IEEE.

[24] Cintra, M. E., Meira, C. A., Monard, M. C., Camargo, H. A., & Rodrigues, L. H. (2011, November). The use of fuzzy decision trees for coffee rust warning in Brazilian crops. In 2011 11th International Conference on Intelligent Systems Design and Applications (pp. 1347-1352). IEEE.

[25] H. Gensheng, W. Haoyu, Z. Yan, W. Mingzhu, A low shot learning method for tea leaf's disease identification, Comput. Electron. Agric. 163 (2019) (2019) 104852.

[26] Hu, G., Wu, H., Zhang, Y., & Wan, M. (2019). A low shot learning method for tea leaf's disease identification. Computers and Electronics in Agriculture, 163, 104852.

[27] Mehmet MetinOzguven a, Kemal Adem, Automatic detection and classification of leaf spot disease in sugar beet using deep learning algorithm.

[28] Karthik, R., Hariharan, M., Anand, S., Mathikshara, P., Johnson, A., &Menaka, R. (2020). Attention embedded residual CNN for disease detection in tomato leaves. Applied Soft Computing, 86, 105933.

[29] Jiang, H., Zhang, C., Qiao, Y., Zhang, Z., Zhang, W., & Song, C. (2020). CNN feature based graph convolutional network for weed and crop recognition in smart farming. Computers and Electronics in Agriculture, 174, 105450.

[30] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, D. Stefanovic, Deep neural networks based recognition of plant diseases by leaf image classification, in: Computational Intelligence and Neuroscience, Vol. 2016, Hindawi Publishing Corporation, 2016, p. 11

[31] S.P. Mohanty, D.P. Hughes, M. Salathé, Using deep learning for image-based plant disease detection, Front. Plant Sci. 7 (2016).

[32] Karpathy, A. (2015). The unreasonable effectiveness of recurrent neural networks.[online] karpathy. github. io. Online verfügbarunter: http://karpathy. github. Io/2015/05/21/rnn-effectiveness/[Zugriff 13 Aug. 2018]. TH Köln Gustav-Heinemann-Ufer, 54, 50968.

[33] Lee, S. H., Goëau, H., Bonnet, P., &Joly, A. (2020). Attention-based recurrent neural network for plant disease classification. Frontiers in Plant Science, 11.

[34] Zhao, B., Wu, X., Feng, J., Peng, Q., & Yan, S. (2017). Diversified visual attention networks for fine-grained object classification. IEEE Transactions on Multimedia, 19(6), 1245-1256.

Modern City Issues, Management and the Critical Role of Information and Communication Technology

Qasim Hamakhurshid Hamamurad[1]*, Dr. Normal Mat Jusoh[2], Uznir Ujang[3]
Universiti Teknologi Malaysia, AHIBS, Information System, Johor Bahru, Johor, Malaysia, 81310[1]
Universiti Teknologi Malaysia, AHIBS, Information System, Kuala Lumpur, Malaysia[2]
Geo-information, Faculty of Built Environment and Surveying[3]
Universiti Teknologi Malaysia, Johor Bahru, Johor, Malaysia, 81310[3]

***Abstract*—Cities are currently dealing with major difficulties that no longer allow slight adjustments to the way cities operate. Instead, local officials must devise creative, transformative solutions. Fortunately, novel approaches to municipal administration and technological advancements are providing city officials with new and beneficial tools. As a result of these improvements, citizens, businesses, and other groups in the city will be able to actively participate in implementing the reforms. In a nutshell, technology can assist cities in becoming smarter. This paper highlighted the implications of the management challenges of cities, the types of cities, and the issues that cities face during epidemics. A coordinated approach that reacts to both COVID-19 and climate change is required to avoid negative outcomes from both. To figure out how Malaysian city administration works and how important information and communication technology is in Malaysian smart cities, it is important to look into technology and data opportunities.**

Keywords*—Cities; modern city; smart city; COVID-19; information communication technology (ICT); challenges; data; geographic information system (GIS)

I. Introduction

Cities today are facing enormous challenges. A small sample of cities from throughout the world deserves special attention [1]. By 2025, 600 cities will account for over 60% of global growth [2]. Currently, the city houses half of the world's population. The globe is seeing tremendous urbanization [3],[4]. The high urban population growth trajectory is not just an intriguing reality; it also causes a demanding urgency for sustainable development and improved livability. The growth of cities is fraught with difficulties. Although cities account for less than 2% of the world's territory, they use almost three-quarters of the world's natural resources and are the primary source of greenhouse gas emissions [5]. Rapid urbanisation has resulted in a loss of essential capabilities that make a place habitable, such as waste management challenges, resource shortages, air quality, human health problems, traffic congestion, and insufficient, degrading, and ageing infrastructure [6].

Social and organisational issues, rather than technical, physical, or material issues, are another group of issues. Multiple different participants, high levels of interconnection, opposing ideals, and political and social complexity are all linked to concerns. Problems get vicious and convoluted in this way [7]. While behind Singapore, Japan, and South Korea, and Taiwan, China, Malaysia is the fifth most urbanised economy in East Asia. Malaysia's three major cities, together with Georgetown City and Johor Bahru, account for over 60% of the country's urban population. For financial services, Kuala Lumpur's economic growth rate per worker is just around 0.46 of Hong Kong's. The cost of living in Kuala Lumpur is 83% more than in New York City [8]. In 2017, over 75% of Malaysia's population lived in cities and metropolitan regions. As a result, Malaysia is one of the most urbanised countries in Southeast Asia. Malaysia's urban population was 77.2% in 2020. Malaysia's urban population expanded from 34.3% in 1971 to 77.2% in 2020, expanding at a 1.67% yearly pace [9].

Malaysia created (Malaysia Urban Information Network [MURNInet]) in 2002 [10]. The plan was renamed MURNInet 2.0 by the (Federal Department of Town and Country Planning Peninsular Malaysia) in 2017 to reflect current urban developments [11]. The (National Urbanization Policy [NUP]) was designed in 2006 by the Federal Department of Town and Country Planning as part of a comprehensive and integrated framework to supply excellent urban services that would assure the establishment of safer, systematic, contemporary, and interesting cities (Federal Department of Town and Country Planning, 2006) [12]. However, because of the many properties of urban planning, many strategies have lived side by side with NUP over time [13]. The smart city goal is not just for medium and large cities; it is equally vital for smaller towns and cities. This research sought to answer the following question: What are the modern city's challenges? What are the challenges confronting Malaysia's urban community? Whether technology and data opportunities have aided in the resolution of these issues is unclear?

The remainder of the paper is summarised as follows: The 2nd section is about the challenges of modern cities; the 3rd section, the challenge of city management, outlines the opportunities associated with technology and data in Section 4, while Section 5, talks about cities' types; the following section, about the benefits of integration; and Section 7, presents the conclusion and goal of the research.

II. Modern Cities' Challenges

And according to the United Nations, urban regions are home to 55% of the world's population [14]. By 2050, the percentage is predicted to climb to 68%. With few exceptions, metropolitan areas are expected to grow and diversify. The issues cities confront today include the following as

*Corresponding Author.

urbanisation grows Table I, particularly in Asian and African countries [15]. The city's challenges are illustrated in "Fig. 1".

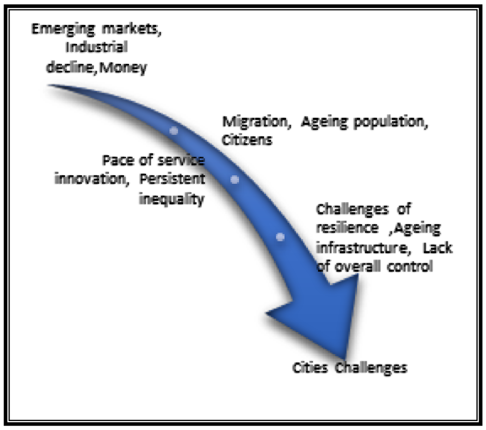


Fig. 1. Cities' Challenges.

TABLE I. CHALLENGES TYPE AND DETAILS

Type of Chanllenges	**Challenges details**
Resilience challenges	Challenges of Resilience Extreme weather[16], food shortages[1], and energy shortages have a significant impact on cities all around the world[17] [18], [19].
Emerging markets	In addition to rising market development, the increasing middle class is predicted to grow from three to six million people by 2050[15]. This will exacerbate economic resource competitiveness. The global race for talent and capital is on[20]. Malaysian cities, such as Kuala Lumpur, must now compete not only with other Malaysian cities but also with global cities[21].
Migration	Many cities are seeing continued inward migration, which is producing traffic jams, pollution, and a rise in service expectations[22]. Even though modern cosmopolitan cities offer a more diverse cultural environment, disparities in social standards might harm social cohesiveness across many populations[23].
Deflation of the industrial sector	In contrast, the population of certain cities is declining while industry is falling [24],[25]
Ageing population	The ageing population is facing an increasingly negative influence on both overall income and government service expenses[26].
Citizens.	As digital technology becomes increasingly widespread in all parts of life [22], citizens demand governmental services and become clients [27].
The pace of service innovation	Because of the quick rate of growth in people's daily lives, long-term planning is becoming increasingly challenging[28].
Ageing infrastructure	Most of the infrastructure in Kuala Lumpur cities [29]will need to be updated or renovated in the next few years to make it suitable[30].
Lack of overall control	Specific organisations decide many critical choices affecting urban life with a limited[31], segmented emphasis, instead of considering the town's overall influence[19]. Many areas of urban life are also influenced by regional or national government or agency actions[1].
Money	It's difficult to get the long-term investment finance required to address municipal problems[32], particularly when business models, innovation strategies[33], or modern tech are involved, and especially when budgets are tight[34].
Ageing infrastructure	Over the next several years, a considerable proportion of the infrastructure in global cities will need to be replaced or retrofitted to try and be fit for its purpose [1], [35]

III. CITY MANAGEMENT CHALLENGES

Cities may be thought of as complicated creatures with a plethora of organizations and infrastructures providing the services required for them all to function correctly Fig. 2. Because the municipal government directly manages only a few important municipal services [36], it should carry out its long-term strategic management role in collaboration with other city stakeholders [37].

The huge number of independent stakeholders who contribute directly or indirectly to the city's overall functioning impeded the city's capacity to function as a cohesive entity [37]. To complicate things worse, independent city systems are becoming more complicated while also becoming more interdependent in ways that are frequently poorly understood [20]. For example, both the electrical and information technology systems are critical components of every other municipal system, and any failure in one of them is likely to have a cascading impact on the others. Smart City Framework Standard [PAS181] examines a city's traditional operating paradigm and identifies the issues that result from a lack of coordination and integration across city systems [38], [30]. This operational model is depicted in Fig. 3.

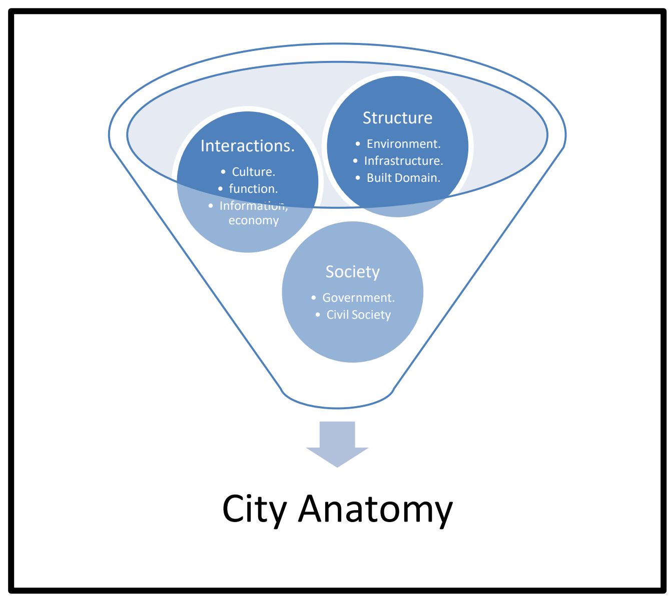


Fig. 2. City Anatomy.

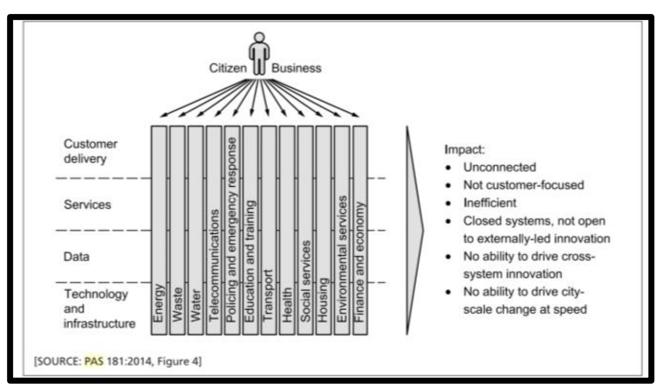


Fig. 3. Citizen Bussiness[39].

Equally difficult is the fact that each city has its own history, set of traits, and geographic location, all of which present distinct difficulties and possibilities for municipal officials to consider. A city must also be run in a way that takes into account its wider social context and the unique opportunities and problems it faces.

IV. Technology and Data Opportunities

Following technological advancements in data utilization Fig. 4 municipal leaders now have new tools and chances for successful transformation that may help them solve these issues.

1) The ability to connect, the digital connection, is rapidly expanding, no longer confined to using computers to access the virtual world; can now use cell phones, tablets, and monitors to access the information, offer feedback, and communicate with one another, anywhere at any time [40].

2) Sensors are a type of device that detects changes in the environment. Sensors, and also smartphones and other digital technology [41], are increasingly being integrated into the built environment, with a rising percentage of people carrying them. This can provide local authorities with a more complete, real-time picture of what's going on in their community [42].

Three more factors are linked to these changes:

- Information. Every two years, the volume of digital data virtually doubles. With the growing usage of (GIS) Geographical Information Systems, data from multiple sources may be connected based on location, providing a much fuller picture of what's going on in the city's many neighbourhoods and districts. It backed this up with open standards from the (Open Geospatial Consortium [OGC]) and the (World Organization for Standardization's Technical Working Group211 [ISO/TC211]) [43]. The capacity to specify a geographic reference, and support access to link data the strong argument for sharing it among departments, agencies, industry, and research organisations [44].

- Trend analysis. The capacity to interpret and deduce meaning from this data is likewise quickly improving [45]. While smartphone applications and devices may already deliver contextualised information to users on a face-to-face basis as they go about their daily lives, quantum computers can scan large amounts of raw data and provide answers to more challenging problems [46].

- Working together. The growing usage of the internet is facilitating a more collaborative connection between customer and provider, resulting in patients, for example, having a more active part in controlling and measuring their health issues [47].

These developments are allowing for the development of tools that will allow people, businesses, and local authorities to respond to issues faster and with more evidence. These are the tools:

- In order to make better decisions for city administration and long-term planning, make data more accessible and affordable, as well as give people the ability to see and analyse it, so that they can do so.

- Encourage collaboration between government entities, citizens, residents, and government agencies.

- Establish innovative techniques for connecting city systems to enable a more holistic approach to city issues and possibilities, assist in city administration and also to enable integration at a much more detailed level, reducing the risk of a major collapse.

Those instances are relevant to current technical developments. They'll be followed by a steady flow of additional technological advancements, such as the utilisation of new materials, robots, new building methods, as well as domestic industry (e.g. 3D printing). The goal for cities is to foster a continual innovation culture, using innovative technology to serve citizens and improve the city's quality of life.

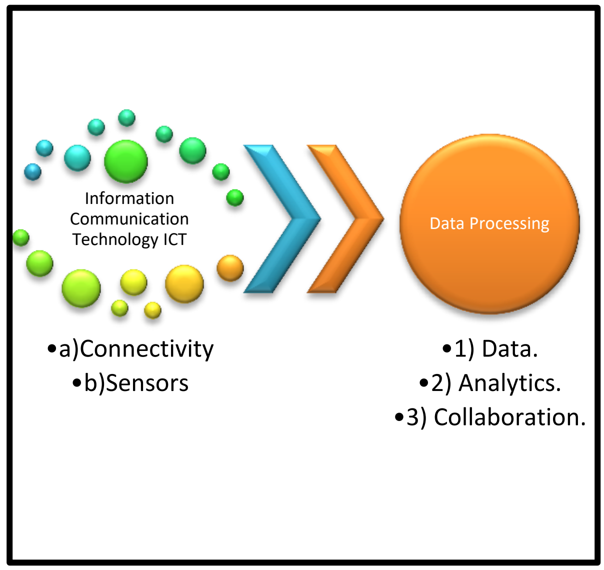


Fig. 4. Technology and Data Processing.

V. Type of Cities

A. Healthy Cities

Healthy cities, smart cities, and sustainable cities have all emerged as major themes in the modern evolution of cities. The epidemic has shown that having healthy cities as a development criterion is equally critical. Preventing future infections and disease outbreaks would require improving urban medical infrastructure to accommodate the influx of patients with new illnesses and interruptions in medical care, especially public health programmes like vaccination [48]. The present urban primary health care infrastructure must be physically and financially enhanced, as well as monitored, surveillance, and accountability systems that are effective and sophisticated. There must also be minimal friction, and all levels of healthcare must collaborate using a pandemic-related operating procedure that is unique and standard [16].

Digital tools can improve healthcare delivery and strengthen health infrastructure. The COVID-19 situation has increased the usage of telehealth services, which residents had been hesitant to use [49]. The rapid surge in the usage of digital health peripherals will provide the necessary proving ground for the evolution of health-tech that complements rather than replaces traditional types of health care [48]. With innovation education and power levelling programmes being implemented in medical, nursing, and paramedical curricula, this transformation in healthcare delivery will also push digitally aware health personnel [50].

B. Green Cities

The absence of open spaces, inefficient resource administration for the afflicted people and interruptions in power consumption are only a few of the fissures caused during the COVID-19 crisis. These are issues that have their foundation in the absence of resilient and sustainable urban development. To avoid detrimental consequences from both COVID-19 and the climate problem during pandemics, a coordinated approach that reacts to both must be developed. Prioritize circular economic structures, sustainable urban transportation, and increased green building and renewable power investment [51].

To make cities robust and sustainable in the face of future pandemics, the transportation industry will need to integrate with the green agenda [55]. Adjustments in transportation preference and user behaviour because of congestion phobia must be balanced against institutional decisions on how to ease bottlenecks. Creating well-ventilated public transportation, enforcing congestion limitations and pricing, and supporting non-motorized transportation are some of the approaches to accomplishing this balance [56].

C. Inclusive Cities

Much before the epidemic, there were discussions about the necessity for cities to consider the needs of their vulnerable populations. This crisis has given the pain of the vulnerable a new depth, and it has sparked a discussion about what might be done to prevent similar pain in the future.

An urban-planning framework that incorporates sex-disaggregated data collection on the number of people, number of homes, pockets and regions of susceptibility, as well as sanitation and hygiene services readily available services, is needed to support pandemic responses [52]. There can be no preparedness or response plans until the percentage of people who live in susceptible locations is known, as well as information on the municipal facilities accessible to them. A concerted effort to collect up-to-date data on formal and informal settlements, as well as labourers in cities, is required to accurately estimate the funds needed to ensure appropriate access to housing, food, sanitation, and services in the financial sector [53].

A gendered study of the pandemic's influence on job losses and employment patterns is also required for post-pandemic rehabilitation efforts. To prevent the harshest effects of future pandemics, improved data management must be included in urban design. This data, together with the efficient coordination that municipal systems can offer, will be critical in the delivery of critical services to the most vulnerable..

D. Rising Cities

In the post-pandemic future, cities' sociological, physical, and economic infrastructure will shift dramatically. Building architecture and workplace layouts will alter as a result of health restrictions put in place during an emergency, with safety being the priority.

At the same time, cities will continue to build skyscrapers and slums, as long as their urbanisation goals are met. However, the most effective urban growth strategy will be one that integrates formal employment generation with the construction of viable housing plans [52].

It may reinvent cities as solitary centres of trade, business, and knowledge as the post-pandemic globe shrinks and the rapid pace of globalisation slows. Commodity and labour movement will be restricted within clearly defined governing zones, and cities may emerge as self-sufficient entities [54].

The COVID-19 problem will force significant changes in how to design, govern and live in cities. The perception of cities as dense hives of continual social contact may need to be rethought, must accept this and guarantee that systems, whether in transportation, housing, work, or financial investments, endure the adjustments that come with it. While it may appear unachievable shortly, parties must work together to implement measures that will help avert future crises and lessen the pandemonium's negative short-term implications [52].

VI. BENEFITS OF INTEGRATION

The streetlamps in Malaysia are turning white. The cause is more complex than you may believe. In Malaysia, seventy-five per cent of street lamps are beyond the age of twenty-five [55]. LEDs are being used to replace incandescent bulbs, which might cut your energy bill in half and save you money on maintenance, with a (return on investment [ROI]) of years [54]. If they were changed, they might also be used as key resources for a Wi-Fi communication network, a hub for a parking system as well as other sensors, a position for air pollution monitoring, and a CCTV stand [56].

Development is one of Kuala Lumpur's major urban redevelopment projects, and it will eventually provide housing, employment, and educational opportunities for 8,419,566 people [57]. The integrated planning and management of water, power, gas, heat, and data services under the control of a single asset owner is a crucial aspect. This integration will allow for significant cost savings and implementation of innovative technology solutions in energy supply, renewable energy, and other areas, all of which will contribute to increased levels of sustainable development and an aimed decrease in CO_2 emissions of over 50% compared to 2005 levels [58].

VII. CONCLUSION

The goal of this research was to assess the obstacles that Malaysia's urban community faces, as well as whether government policies have been effective in resolving these issues. Rising crime rates, a lack of economic prospects, air pollution, and traffic congestion are concerns from the viewpoint of urban residents. The survey found that the government's policies were in line with the public's concerns. However, it would be ideal if the urban population were given the chance to take part in the development of policies on urban livability and contribute comments. The third principle of the City Competitiveness Master Plans (CCMP) framework could not be examined since the data used in this study was obtained from the public domain. As a result, future studies will create particular survey questions that can capture the finer aspects of urban living issues.

The most important message is that cities' prospects would be significantly improved if they could effectively seize opportunities and handle the issues they confront. Below are a few key themes for forthcoming comments on important research and policy activities. When cities are well-managed, they can take advantage of strong tools like innovation and technology, public participation, effective governance, and the ability of communities to bounce back.

ACKNOWLEDGMENT

This research was partially funded by UTM Research University Grant, Vot R.J130000.7752.4J550 and Vot Q.J130000.2452.09G84.

REFERENCES

[1] B. S. I. S. Publication, “Smart cities overview-Guide BSI Standards Publication,” 2015.

[2] “Urban world : Mapping the economic power of cities,” no. March, 2011.

[3] I. B. M. Global, B. Services, and E. Report, “Smarter cities for smarter growth.”

[4] I. B. M. Global, B. Services, and E. Report, “How Smart is your city ?”

[5] J. Marceau, “Management, Policy & Practice,” vol. 9338, no. February, 2016.

[6] H. Innovation, “How Innovation and ICT The Smart City vision :,” pp. 1–9, 2010.

[7] E. P. Weber et al., “Settings All use subject to JSTOR Terms and Conditions Essays on Leadership in Organizations Wicked Problems , Knowledge Capacity Builders in Network Challenges , Settings and Collaborative,” vol. 68, no. 2, pp. 334–349, 2014.

[8] J. L. Baker and M. Lee, “Achieving a system of competitive cities in Malaysia : Main report (English),” 2015.

[9] “Malaysia - Urban population as a share of total population.” [Online]. Available: https://knoema.com/atlas/Malaysia/Urban-population#:~:text =Malaysia - Urban population as a share of total population&text=In 2020%2C urban population for,average annual rate of 1.67%25. [Accessed: 23-Aug-2021].

[10] M. A. Marzukhi, D. Omar, O. L. H. Leh, M. S. Hamir, and M. Barghchi, “Malaysian urban Indicators Network: A sustainable development initiative in Malaysia,” Eur. J. Soc. Sci., vol. 25, no. 1, pp. 77–84, 2011.

[11] “What is Federal Department of Town and Country Planning | IGI Global.” [Online]. Available: https://www.igi-global.com/dictionary/federal-department-of-town-and-country-planning/10983. [Accessed: 25-Mar-2022].

[12] D. B. Omar and O. L. H. Leh, “Malaysian Development Planning System: Kuala Lumpur Structure Plan and Public Participation,” Asian Soc. Sci., vol. 5, no. 3, 2009.

[13] S. Somasundram, M. Sambasivan, R. Rasiah, and T. Pei-Leng, “Ranking the challenges of the urban community in Malaysia,” Institutions Econ., vol. 10, no. 4, pp. 69–89, 2018.

[14] D. Gu, “Exposure and vulnerability to natural disasters for world’s cities* United Nations Population Division Technical Paper No. 2019/4,” United Nations Dep. Econ. Soc. Aff., no. Technical Paper No. 4., pp. 1–43, 2019.

[15] United Nations, Human Development Report 2016: The next frontier - Human development and the Anthropocene. 2016.

[16] V. Alberti et al., The Future of Cities: Opportunities, challenges and the way forward, no. JRC116711. 2019.

[17] K. Archick and D. E. Mix, “The European Parliament,” Congr. Eur. Parliam. Legis. Coop., pp. 41–61, 2012.

[18] K. Koens, F. Melissen, I. Mayer, and C. Aall, “The Smart City Hospitality Framework: Creating a foundation for collaborative reflections on overtourism that support destination design,” J. Destin. Mark. Manag., vol. 19, no. March, p. 100376, 2021.

[19] A. S. de Rosa and T. Mannarini, “Covid-19 as an ‘invisible other’ and socio-spatial distancing within a one-metre individual bubble,” Urban Des. Int., 2021.

[20] Anna Visvizi and Miltiadis D. Lytras, Smart Cities: Issues and Challenges Mapping Political, Social and Economic Risks and Threats. 2019.

[21] Pemandu, “Economic Transformation Programme A Roadmap For Malaysia 123 Chapter 5: Developing Greater Kuala Lumpur/Klang Valley as an Engine of Economic Growth,” ETP A Roadmap Malaysia, pp. 123–162, 2012.

[22] S. Abdullah Almaqashi, S. S. Lomte, S. Almansob, A. Al-Rumaim, and A. A. A. Jalil, “The impact of icts in the development of smart city: Opportunities and challenges,” Int. J. Recent Technol. Eng., vol. 8, no. 3, pp. 1285–1289, 2019.

[23] M. Foth, “Participatory urban informatics: towards citizen-ability,” Smart Sustain. Built Environ., vol. 7, no. 1, pp. 4–19, 2018.

[24] C. Harrison and I. A. Donnelly, “A theory of smart cities,” 55th Annu. Meet. Int. Soc. Syst. Sci. 2011, pp. 521–535, 2011.

[25] K. Axelsson and M. Granath, “Stakeholders’ stake and relation to smartness in smart city development: Insights from a Swedish city planning project,” Gov. Inf. Q., vol. 35, no. 4, pp. 693–702, 2018.

[26] BSI United Kingdom, “BSI Standards Publication BSI Standards Publication Code of practice for Smart city service framework – Guide customer to establishing strategies for smart cities and communities,” 2014.

[27] L. Torres, V. Pina, and B. Acerete, “E-government developments on delivering public services among EU cities,” Gov. Inf. Q., vol. 22, no. 2, pp. 217–238, 2005.

[28] “MSc_thesis_Sander_van_der_Klei.”

[29] S. Eni, M. N. Razali, and N. H. A. Maimun, “Urban regeneration policy framework for city of Kuala Lumpur,” Int. J. Sci. Technol. Res., vol. 9, no. 4, pp. 2872–2876, 2020.

[30] D. Gotlib, T. Kulisiewicz, M. Muraszkiewicz, and R. Olszewski, “SMART GOVERNANCE in the CAPITAL CITY of WARSAW with the USE of ICT and GEOINFORMATION TECHNOLOGIES,” ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci., vol. 4, no. 4/W9, pp. 49–56, 2019.

[31] N. M. Khoi, L. E. Rodriguez-Pupo, and S. Casteleyn, “Citizense-A generic user-oriented participatory sensing framework,” in 2017 International Conference on Selected Topics in Mobile and Wireless Networking, MoWNeT 2017, 2017.

[32] A. Ammar and E. M. Ahmed, “Factors influencing Sudanese microfinance intention to adopt mobile banking,” Cogent Bus. Manag., vol. 3, no. 1, pp. 1–20, 2016.

[33] KPKT, “Smart City Handbook,” Minist. Hous. Local Gov., p. 196, 2020.

[34] Gsma, “Guide to Smart Cities. The Opportunity for Mobile Operators,” no. February, p. 37, 2013.

[35] S. Joshi, S. Saxena, T. Godbole, and Shreya, “Developing Smart Cities: An Integrated Framework,” Procedia Comput. Sci., vol. 93, no. September, pp. 902–909, 2016.

[36] S. Myeong, Y. Jung, and E. Lee, “A study on determinant factors in smart city development: An analytic hierarchy process analysis,” Sustain., vol. 10, no. 8, 2018.

[37] L. Mayangsari and S. Novani, “Multi-stakeholder co-creation Analysis in Smart city Management: An Experience from Bandung, Indonesia,” Procedia Manuf., vol. 4, no. Iess, pp. 315–321, 2015.

[38] L. G. Anthopoulos, Leonidas G. Anthopoulos. 2017.

[39] BSI, “Smart city framework – Guidance for decision-makers in smart cities and communities,” no. SEPTEMBER, pp. 1–60, 2013.

[40] D. Gagliardi, L. Schina, M. L. Sarcinella, G. Mangialardi, F. Niglia, and A. Corallo, “Information and communication technologies and public participation: interactive maps and value added for citizens,” Gov. Inf. Q., vol. 34, no. 1, pp. 153–166, 2017.

[41] C. L. Stimmel, Building Smart Cities. 2015.

[42] Z. Tomor, A. Meijer, A. Michels, and S. Geertman, “Smart Governance For Sustainable Cities: Findings from a Systematic Literature Review,” J. Urban Technol., vol. 26, no. 4, pp. 3–27, 2019.

[43] S. Schade et al., Manual of Digital Earth. Springer Singapore, 2020.

[44] T. J. Kim, “Metadata for geo-spatial data sharing: A comparative analysis,” Ann. Reg. Sci., vol. 33, no. 2, pp. 171–181, 1999.

[45] C. Badii, P. Bellini, P. Nesi, and M. Paolucci, “A smart city development kit for designing Web and mobile Apps,” 2017 IEEE SmartWorld Ubiquitous Intell. Comput. Adv. Trust. Comput. Scalable Comput. Commun. Cloud Big Data Comput. Internet People Smart City Innov. SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI 2017 - , pp. 1–8, 2018.

[46] W. Choi, J. Kim, S. E. Lee, and E. Park, “Smart home and internet of things: A bibliometic study,” J. Clean. Prod., vol. 301, p. 126908, 2021.

[47] F. Afonso and S. Regateiro, “Modelling SM Art Citi ES with Citygml R I T A I N Ê S P R I O R C O S T A P Ar D Al F I L I P E Professor Jorge Manuel Lopes Baptista e Silva Presidente : Professora Maria do Rosário Mauricio Ribeiro Macário Orientador : Professor Francisco Afonso Severin,” 2019.

[48] S. Gao, D. Mioc, X. Yi, F. Anton, E. Oldfield, and D. J. Coleman, “Towards Web-based representation and processing of health information,” Int. J. Health Geogr., vol. 8, no. 1, pp. 1–14, 2009.

[49] T. D. Frank et al., “Global, regional, and national incidence, prevalence, and mortality of HIV, 1980–2017, and forecasts to 2030, for 195 countries and territories: a systematic analysis for the Global Burden of Diseases, Injuries, and Risk Factors Study 2017,” Lancet HIV, vol. 6, no. 12, pp. e831–e859, Dec. 2019.

[50] A. Sharifi and A. R. Khavarian-Garmsir, “The COVID-19 pandemic: Impacts on cities and major lessons for urban planning, design, and management,” Sci. Total Environ., vol. 749, p. 142391, Dec. 2020.

[51] T. Insights, “Post-Pandemic Reflections: Smart Cities.”

[52] R. Aditi and L. J. Preeti, Rethinking Cities in a Post-COVID-19 World. 2020.

[53] A. Rizzo and J. Glasson, “Iskandar Malaysia,” Cities, vol. 29, no. 6, pp. 417–427, 2012.

[54] R. Yesner, “Smart Cities and the Internet of Everything: The Foundation for Delivering Next-Generation Citizen Services,” IDC Gov. Insights, no. October, 2013.

[55] R. M. Ramli, “Application of LED Technology into Public Road Lighting in Malaysia for Replacing The High Pressure Sodium Vapour Lighting,” pp. 76–81, 2015.

[56] P. L. Masterplan and F. Report, “Appendix A : Survey of Existing Road and Street Lighting,” no. September, 2002.

[57] World Population Review, “Kuala Lumpur Population 2020 (Demographics, Maps Graphs),” 2020. [Online]. Available: https://worldpopulationreview.com/world-cities/kuala-lumpur-population/. [Accessed: 25-Mar-2022].

[58] “Southeast Asia Smart Street Lighting Market Report 2020: Focus on Indonesia, Malaysia, and Thailand - ResearchAndMarkets.com | Business Wire.” [Online]. Available: https://www.businesswire.com/news/home/20210623005576/en/Southeast-Asia-Smart-Street-Lighting-Market-Report-2020-Focus-on-Indonesia-Malaysia-and-Thailand---ResearchAndMarkets.com. [Accessed: 25-Mar-2022].

A Novel Framework for Sanskrit-Gujarati Symbolic Machine Translation System

Jaideepsinh K. Raulji[1]
Navrachana University
Vadodara, India

Jatinderkumar R. Saini[2]*
Symbiosis Institute of Computer Studies and Research,
Symbiosis International (Deemed University), Pune, India

Kaushika Pal[3]
Sarvajanik College of Engineering and Technology
Surat, India

Ketan Kotecha[4]
Symbiosis Centre for Applied Artificial Intelligence,
Symbiosis International (Deemed University), Pune, India

***Abstract*—Sanskrit falls under the Indo-European language family category. Gujarati, which has descended from the Sanskrit language, is a widely spoken language particularly in the Indian state of Gujarat. The proposed and realized Machine Translation framework uses a grammatical transfer approach to translatethe written Sanskrit language to Gujarati. Because both languages are morphologically rich, studying the morphology of each item is difficult but necessary to incorporate into implementation. To improve the implementation accuracy and translation clarity, an in-depth research of the creation of Nouns, Verbs, Pronouns, and Indeclinables, as well as their mappings, has been carried out. Tokenization, lemmatization, morphological analysis, Sanskrit-Gujarati bilingual synonym-based dictionary, language synthesis, and transliteration are the proposed framework's primary components. The implementation outcome was tested on 1,000 phrases, using the automated Bilingual Evaluation Understudy (BLEU) scale which yielded a value of 58.04 It was also tested on the ALPAC scale, yielding the Intelligibility score of 69.16 and the Fidelity score of 68.11. The results are encouraging and prove that the proposed system is promising and robust for the implementation in the real world applications.**

Keywords*—Bilingual synonym dictionary; Gujarati; lemmatization; machine translation system (MTS); morphological analyzer; Sanskrit; synthesizer; transliteration

I. Introduction

Aside from computers' incredible processing capacity, researchers have traditionally found it difficult to create and execute Machine Translation Systems (MTS) with great precision. The complexity of natural languages is due to lexical, semantic and contextual aspects, sophisticated morphological nature, and most importantly the pragmatics and discourse, which refers to the speaker's intent. The designing and the implementation of a Machine Translation (MT) system can be done in a variety of ways.

In this paper, a technique for constructing a symbolic MT implementation from Sanskrit to Gujarati is offered due to rare availability of bilingual parallel corpora which form the basis for machine learning techniques. A pure dictionary- based translation system uses no intermediate representation to convert from source to target language.

The Machine Translation (MT) approaches could be classified broadly into four categories, as is depicted diagrammatically in Fig. 1. Notably, two of these four broad categories can be further divided into two sub-categories for each broad category. Historically speaking, the correlation of the categorization of the machine translation approaches existing in the pertinent scientific literature could also be done for the rationalistic, empirical and the hybrid approaches.

For the present research work, a dictionary has been used to accomplish the task, as it will offer a word to word transformation through sub-tasks like morphological analysis supplemented with lemmatizer, grammatical transfer, synthesis. It will later rearrange the words in the sentences of the target language. The method is simple to use, but it is not versatileenough to be applied several other pairs.

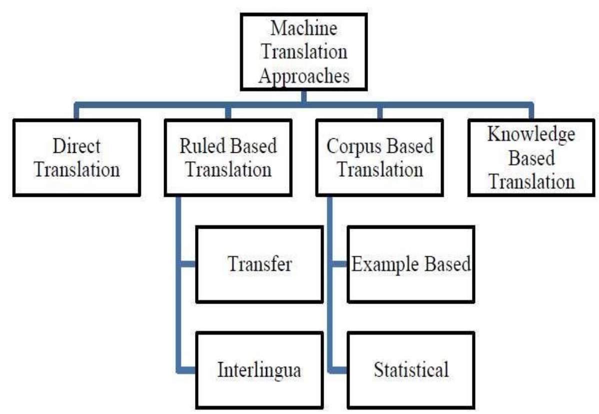


Fig. 1. MT Approaches [2].

The transfer approach is more complicated than the preceding one since it examines properties as lexical, syntactic & semantics and morphological aspects of language. Because it is built to accommodate various languages, the Interlingua approach is still more versatile than transfer. Interlingua is used to construct an intermediate representation of natural language also known

*Corresponding Author

as pivot language which is then transformed to target [1]. The relativeness of Direct, transfer, and interlingua methods are strategically connected, as shown in Fig. 1. If a significant number of labelled, aligned, or parallel corpora are available, the corpus-based technique tends to be accurate enough. Because the grammatical mechanics of a language have no effect on corpus-based models, a single corpus-based MT model can be used to train a model in any language.

II. Literature Review

The amount of study and money invested on the MT system after World War-II is notable. However, after the Automated Language Processing Committee (ALPAC) issued a report in 1966 CE, the funding for the MT system was substantially decreased. After the 1990s, a ray of optimism emerged, thanks to lower computer hardware costs and increased memory and calculation capacity, which led to new techniques. MT-related work used to be limited to languages such as English, Russian, French, and Spanish, but in today's world, MT systems are being developed for a wide range of languages, including Sanskrit.

As shown in Fig. 2, Cancedda et al. [3] presented a diagrammatic representation of the various methods used for machine translation. Many MT systems use Sanskrit and Gujarati in some form or another. Rathod and Sondur presented English-Sanskrit Translator and Synthesizer (ETSTS) which is a combination of rules and example-based MT implementation which transforms sentences to speech [5]. E-Trans is an English to Sanskrit MT tool based on Synchronous CFG proposed by Bahadur et al. The language representation part is implemented through SCFG [6]. Subramaniam [7] built Sanskrit to English rule-based translator. Sandhi Splitter, Translation Generator with Morphological parser are the two important components of the implementation. English to Sanskrit Example-Based MT system is developed by Mishra and Mishra [8] [9]. The main components of the system are Part-of-Speech (POS) tagger, Gender-Number-Person (GNP) detection, as well as Noun, Root Verb, and Adverb detection. A nice piece of work which translates Sanskrit to Hindi has been developed at Jawaharlal Nehru University (JNU). Word sense disambiguation, anaphora resolution, prose order generation, and other modules were studied by the researchers while it was claimed that Yoga and Ayurveda will be added to the system's capabilities [10]. AnglaBharti MT system translates English to Sanskrit. It is based on Paninian Grammar rules also known as PLIL code [11]. Raulji and Saini [4] presented a comparison of the various machine translation systems involving Sanskrit and Gujarati as the language pair.

Sreedeepa and Idicula [12] developed Sanskrit-English MT implementation based on Interlingua. In analysis of language, LFG is used which helps in finding semantic relation between words in a sentence. The semantic analysis was done through Karaka analyzer through Paninian grammar framework. Using interlingua approach, Sanskrit to English MT is developed by Sreedeepa and Idicula [12]. It used Lexical Function Grammar (LFG) build using Paninian Karaka Analysis. The karaka analysis is used to analyse syntactico- semantic relations between words in a sentence. Gupta et al. developed Sanskrit to English MT system. The system is based on grammatical aspect of the language pair [13]. Singh et al. [24] deployed the hybrid usage of Neuro Machine Translation (NMT) and Rule Based Machine Translation (RBMT) to design the MTS for the Sanskrit-Hindi language pair. Akhand et al. [25] while reviewing the MT systems for the Bangla language, found that no MTS exists that involves Bangla-Sanskrit language pair. In addition to the above mentioned MT systems, the researchers have also attempted to evaluate the accuracy of MTS. For instance, Sabtan [26] used the data of social media itself as a language for translation. Ehab et al. [27] investigated the MT using the example based approach for the language pair comprising of Arabic and English languages. Pudaruth et al. [28], similarly, discussed the Rule Based Machine Translation (RBMT) system for the language pair comprising of English and Creole.

Given the richness of the Sanskrit language, there have been several attempts by the researchers involving the analysis of the language. Derivative nouns [29], word segmentation and morphological parsing [30], noun declension and verb conjugation [31], dependency parsing [32], lemmatization [33], and constituency mapper [34] are a few such instances. Similarly, for the Gujarati language, the researchers have explored chunking [35], stemming [36], inflections [37], lexicon-based analysis [38], speech recognition [39], character recognition [40], and spell checking [41]. Based on the detailed literature review till date, we have observed that there is a definite dearth of research on MTS for the Sanskrit-Gujarati language pair. It has also been observed that no formal research works are dedicated to the morphological analysis, comparison and linking of both languages together. The present research work bridges all these gaps and presents not just the theoretical framework but also the working model of the MTS involving these two Indian languages. The results have been found to be encouraging and motivating. Rest of the paper is organized as follows: Section III presents the characteristics of Sanskrit and Gujarati languages while Section IV presents a detailed discussion on the research methodology. This is followed by a section each on results, and conclusions and future work.

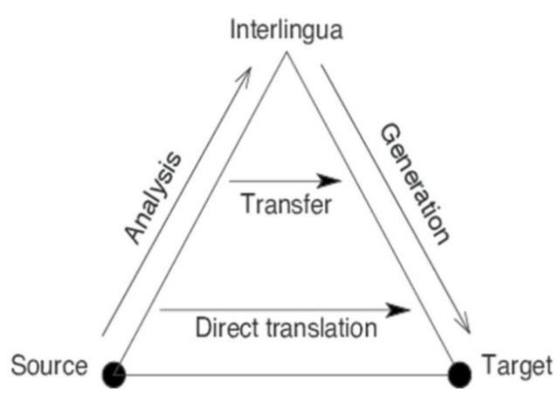


Fig. 2. The Translation Methods [3].

III. Characteristics of Sanskrit and Gujarati Languages

Sanskrit and Gujarati are included in the Indian Constitution as scheduled languages historically belong to Indo-Aryan family of languages. Gujarati is less ordered and regular than Sanskrit. Sanskrit is rich and morphologically structured hence tends to be focused internationally for research in computational linguistics domain. Gujarati is official language of state of Gujarat. Apart from state of Gujarat, it is also spoken in adjoining parts of Rajasthan, Madhya-Pradesh and Maharashtra states of India.

Many Gujarati community are also found in countries viz. UK, USA, Canada, Australia, New Zealand, and few African continent's countries. Sanskrit is an ancient spoken language with tradition dating back to the Vedic period since 2000 BCE. Gujarati is a contemporary language compared to Sanskrit, with a spoken heritage dating back to roughly 1100 CE. [14] [15] [16]. Sanskrit is written in a variety of scripts, the most common of which being Devanagari [17], whereas Gujarati is written in Abugida script, which is a variant of Devanagari. Table I lists a few characteristics of these language pairs [18].

TABLE I. Characteristics of Sanskrit and Gujarati Languages

LanguageElements	Sanskrit	Gujarati
Consonants	33	33
Vowels	12	12
Gender (3 genders ineach)	Masculine	Masculine
	Feminine	Feminine
	Neuter	Neuter
Number (3 numbers in Sanskrit and 2 in Gujarati)	Singular	Singular
	Dual	Plural
	Plural	Plural
Case Markers (8 Cases in each)	Nominative	Nominative
	Accusative	Accusative
	Instrumental	Instrumental
	Dative	Dative
	Ablative	Ablative
	Genitive	Genitive
	Locative	Locative
	Vocative	Vocative
Persons (3 persons ineach)	First	First
	Second	Second
	Third	Third
Tense (6 tenses in Sanskrit and 5 inGujarati)	Present	Present
	Aorist	Past (Simple)
	Past (Imperfect)	Past (Imperfect)
	Past (Perfect)	Past (Perfect)
	Future (First)	Future
	Future (Second)	Future
Moods (4 in Sanskrit and 3 in Gujarati)	Imperative	Imperative
	Potential	Potential
	Conditional	Conditional
	Benedictive	No equivalent

IV. Methodology

The strength of the language analysis performed on the source and target languages determines the success of a rule- based system. Better findings come from a thorough examination of source and target language divergence and similarity mappings. The rule-based paradigm is given here, with an emphasis on grammatical similarities and divergence between Sanskrit and Gujarati, as well as extensive dictionary support. Due of its complexity, the main MT work entails a large number of subs and ancillary tasks. The following sub- sections present the various Natural Language Processing (NLLP) and Computational Linguistic (CL) tasks to finally yield complete MTS. The diagrammatic flow of the working of the proposed system is depicted in Fig. 3. The input text provided in Sanskrit language gets translated to the Gujarati language after passing through stages like tokenization, morphological analysis, lemmatization, translation, synthesis and transliteration.

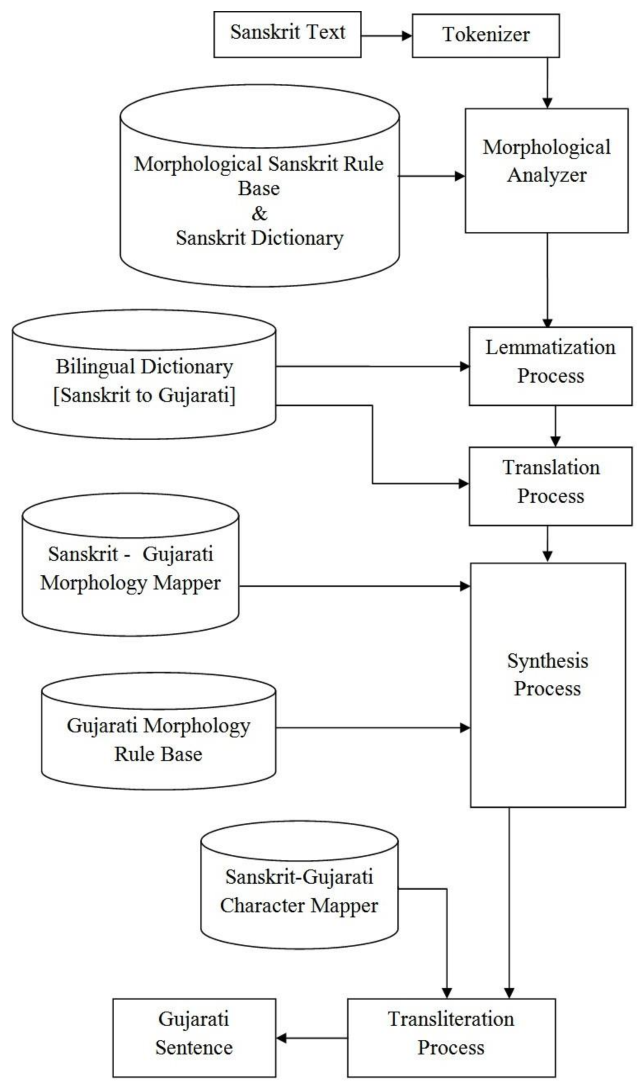


Fig. 3. Framework of Sanskrit-Gujarati MT Implementation.

1) Tokenization phase: Tokenization is the process of breaking down paragraphs into sentences, with each sentence serving as a token. If the sentence is broken down into multiple words, each word serves as a token. Because Sanskrit has a lot of word morphology, the text has to be tokenized into words before it can be properly analyzed. In the language, space separates each word. Fig. 4 depicts the procedure. The single vertical line depicts end of sentence (‘|’) with 2404 asits Unicode and double vertical lines (“||”) depicts end of poetic stanza with 2405 as its Unicode. These two symbols are used to Sanskrit sentence tokenizers. Although the use of '.' (full stop) in modern Sanskrit literature is incorrect, it is nonetheless included in the method for Sentence Boundary Detection (SBD). The space delimiter is used to tokenize Sanskrit words.

2) Morphological-analysis phase: Except for indeclinables, every Sanskrit word can reflect its unique grammatical qualities by adding inflection to the root word. Indeclinables are words that do not possesses any kind of inflectional variants and hence added to dictionary/wordnet. Sanskrit pronouns also have irregular declension patterns; hence they were entered straight into the datastore. The inflectional affixes of the remaining nouns are examined using a grammar rule base and dictionary. The surface grammatical information for the word is provided by the Sanskrit dictionary, such as pronoun, noun, verb, and so on. The G (Gender)-N (Number)-C (Case) labels for nounconstituent and adjective constituents are used to tag a word using deep structure research employing Sanskrit grammatical rules [19]. For verbs, there are Tense-Aspect- Modality (TAM), Person, Number, ‘Parasmaipada’, and ‘Aatmanepada’ labeling modes [19]. Finally, morphological analyzer produces words that have been tagged with grammatical information. To quickly develop the prototype, high-frequency words from corpora of about 75000 words were used to find 75 stop-words, which were then put to the dictionary. This reduces translation time-complexity [20]. The author in [42] presents Sanskrit stop-word analysis while comparison of such analyzers is presented in [43]. The algorithm is shown in Fig. 5 as a logic flow diagram.

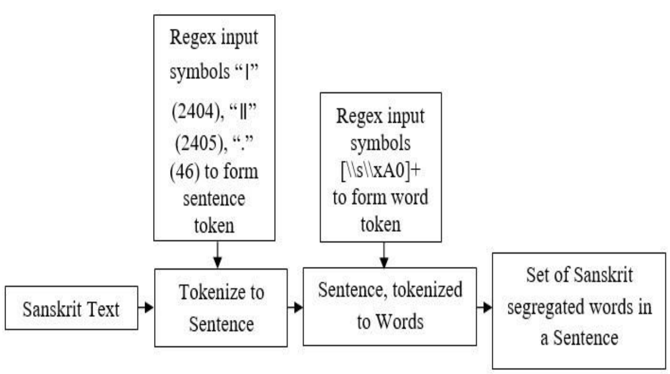


Fig. 4. Tokenizing Sanskrit Text.

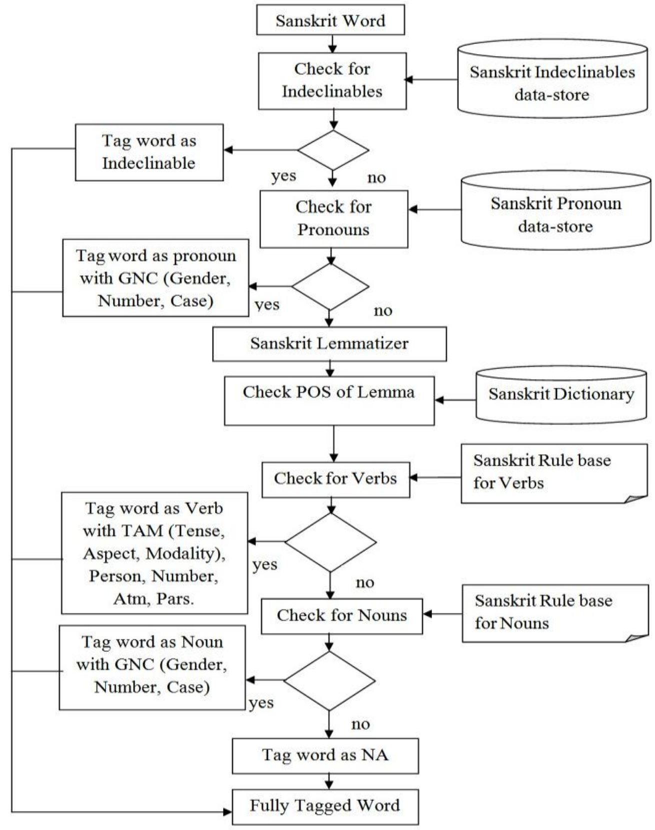


Fig. 5. Morphological Analyzer.

3) Lemmatization phase: A lemma (root word or dictionary form) is derived from an inflected word using this method. Nominal and verbal inflections abound in Sanskrit. If Aatmanepada and Parasmaipada are included, a single Sanskrit noun has 24 variants and 18 verb variants in its inflected forms. As a result, storing all Sanskrit words with such inflection forms necessitates a large number of dictionary entries, and computational retrieval becomes time-consuming. As a result, the dictionary will only contain Sanskrit terms in their basic form. After applying suffix stripping rules, the lemmatizer examines the token and searches the dictionary for the word. Fig. 6 depicts the process diagram.

4) Translation phase: For the translation procedure, the lemma obtained from the Lemmatizer phase is used as the input. The obtained lemma is compared with a bilingual Sanskrit- Gujarati dictionary. It is notable that the output of the lemmatization phase is the root form of the word. It is also noteworthy that we have directly implemented the lemmatizer instead of a stemmer which does not necessarily give the root form. The Sanskrit root word is matched within a bilingual Sanskrit-Gujarati dictionary to get the Gujarati equivalent as mentioned in Fig. 7. To get the Gujarati equivalent, the Sanskrit root word (Sanskrit lemma) is matched in order. The order of matching is as follows: Indeclinables, Pronouns, Verbs, and the remaining Nominals.

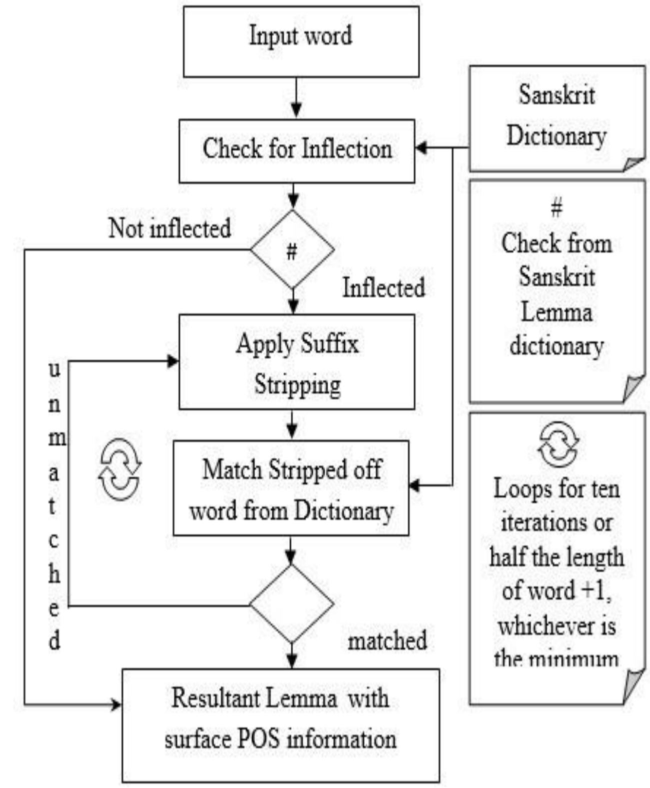


Fig. 6. Sanskrit Lemmatizer.

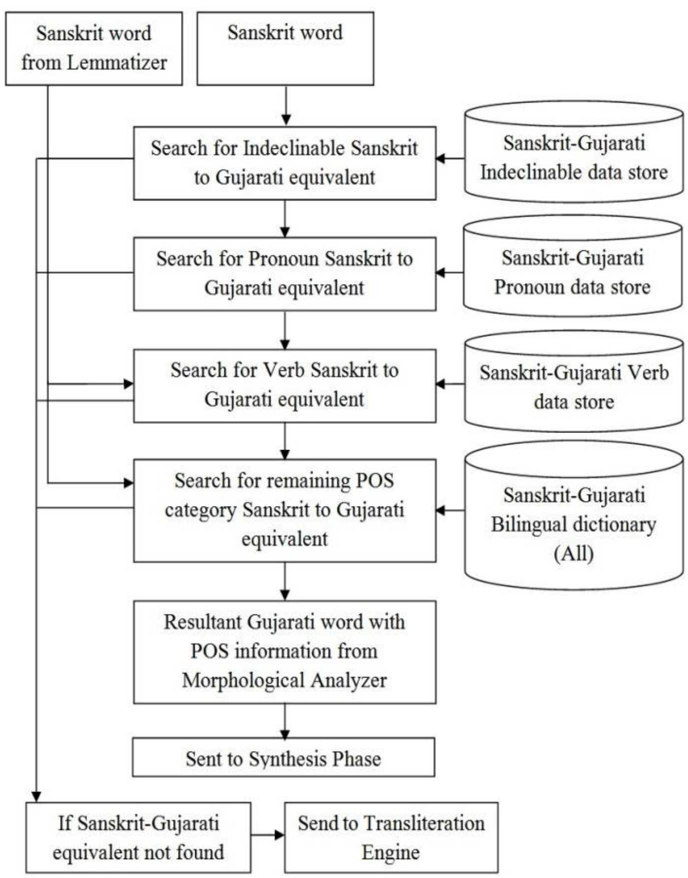


Fig. 7. Translation Phase.

5) Synthesis phase: This phase has mapping repository of morphology of Sanskrit to Gujarati for various Parts of Speech (POS) including nouns, adjectives and verb constituents. Based on the morphological rules derived from the grammar of the source language, it maps to rules of target language and is finally applied on Gujarati lemma to form a meaningful word. Fig. 8 depicts this process diagrammatically.

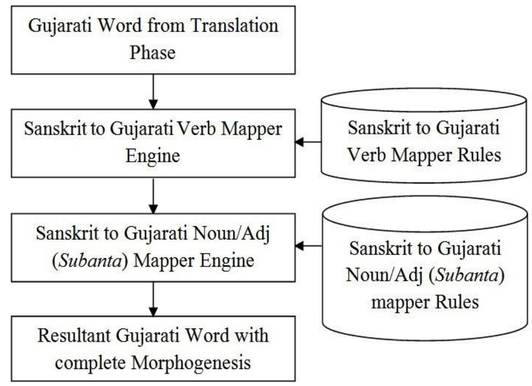


Fig. 8. Synthesis Phase.

6) Transliteration phase: The process of converting language script X to language script Y without harming pronunciation is called as Transliteration. Here the unmatched words from the translation phase are supplied to the transliteration phase, which finally changes Sanskrit (Devanagari) script into Gujarati (Gujarati-Devanagari) equivalents script letters while maintaining their pronunciation. Unmatched terms are mostly seen in the Named Entity class. A Unicode UTF-8 Devanagari scripted font is used to identify the single characters of a Sanskrit word. To generate UTF-8 Gujarati script characters, add 384 to the word, as illustrated in Fig. 9. Because Sanskrit and Gujarati are both free-word order languages, rearranging words in a phrase has little impact on the meaning of the sentence.

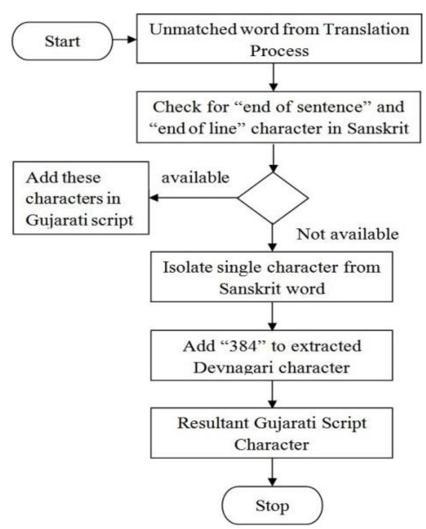


Fig. 9. Transliteration Phase.

V. Results and Findings

Automatic evaluations are significantly more objective because they cover a limited element of the attributes to be examined, whereas human evaluations are too subjective. As a result, it's impossible to compare machine and human results. For morphologically complex language pairs, evaluation by human is considered appropriate, albeit arduous, resource-intensive, and time-intensive task. Despite the fact that BLEU is inappropriate for language with rich morphological characteristics and does not even handle word synonym factor and inflections. The suggested implementational framework is evaluated using the Bi-Lingual Evaluation Understudy (BLEU). However, the general acceptance of BLEU in the MT community is the rationale for its use in evaluation. BLEU was curated and designed by Papinene et al. at IBM [22]. Pn is a modified n- gram precision used by BLEU. Because the BLEU approach is based solely on Precision, it does not use Recall. However, it compensates for recollection by including a Brevity- Penalty feature for short sentences that are translated. The formula can be found below.

$$BLEU = BP * \exp[\sum_{n=1}^{N}\{\left(\frac{1}{n}\right) * \log(Pn)\}]$$

Here,

N = Maximum n-gram (n=1 to N) BP = Brevity Penalty

Here 1000 sentences from varied grammatical categories were chosen to test the system [21], the implemented algorithm received a BLEU score of 58.04. The ALPAC scale was used to assess the same set of sentences manually. The Automated Language Processing Committee's (ALPAC) Fidelity and Intelligibility measure is a two-scale metric [23]. The identical set of sentences used in BLEU were manually assessed using ALPAC's Intelligibility and Fidelity scale by ten language specialists. The Intelligibility score was 69.16, while the Fidelity score was 68.11.

VI. Conclusion and Future Work

To implement a rule-based system is always a challenge due to complexity of rules as well as the number of rules. This is particularly true for morphologically rich languages like Gujarati and Sanskrit. The challenge is to cover each and every grammatical element. Through the proposed implementation, the robust results are obtained due to inclusion of each grammatical feature in details. The unique MT framework on Sanskrit to Gujarati received satisfactory results utilizing the ALPAC’s manual scale on intelligibility and fidelity parameter. It also received good results on the BLEU scale. As the architectural implementation may improve the results by covering a larger range of dictionary words and accounting for any grammatical language exceptions, in future we will consider these additional elements as well. Also on availability of huge bilingual corpus in future, machine and deep learning frameworks can be implemented to make the system more accurate and generic.

References

[1] Poibeau T. (2017), “Machine Translation”, Book – The MIT Press, Massachusetts Institute of Technology, Cambridge, ISBN: 978-0-262-53421-5, Vol-1, PP 296.

[2] Siddiqui T., and Tiwary U.S. (2008), “Natural Language Processing and Information Retrieval”, Oxford University press, ISBN-0195692322.

[3] Cancedda N., Dymetman M., Foster G., and Goutte C. (2008), “A Statistical Machine Translation Primer”. doi: 10.7551/mitpress/9780262072977.003.0001.

[4] Raulji J., and Saini J.R. (2016), “Sanskrit Machine Translation Systems: A Comparative Analysis”, International Journal of Computer Applications, 136(1):1-4. doi: 10.5120/ijca2016908290.

[5] Rathod S., and Sondur S. (2012), “English to Sanskrit Translator and Synthesizer (ETSTS)”, International Journal of Emerging Technology and Advanced Engineering, 2(12):379-383. Available online: https://ijetae.com/files/Volume2Issue12/IJETAE_1212_68.pdf.

[6] Bahadur P., Jain A.K., and Chauhan D.S. (2012), “Etrans – A complete framework for English to Sanskrit Machine Translation”, International Journal of Advanced Computer Science and Applications, Special Issue from ICWETT-2012, 4:52-59. doi: 10.14569/SpecialIssue.2012.020107.

[7] Subramaniam A. (2005), “Sanskrit to English Translator”, Language in India, 5(1). Available online: http://www.languageinindia.com/jan2005/aparnasans kritdissertation1.html.

[8] Mishra V., and Mishra R.B. (2009), “Divergence patterns between English and Sanskrit Machine Translation”, INFOCOMP Journal of Computer Science, 8(3):112-138.

[9] Mishra V., and Mishra R.B. (2010), “ANN and Rule Based Model for English to Sanskrit Machine Translation”, INFOCOMP Journal of Computer Science, 9(1):80-89.

[10] “JNU Sanskrit Hindi MT” (2022), Available online: http://sanskrit.jnu.ac.in/shmt/index.jsp.

[11] Goyal P., and Sinha R.M.K. (2009), “A Study towards Design of an English to Sanskrit Machine Translation System”, proc. of Sanskrit Computational Linguistics: First and Second International Symposia, ACM, pp.287-305. doi: https://doi.org/10.1007/978-3-642- 00155-0_13.

[12] Sreedeepa H.S., and Idicula S.M. (2017), “Interlingua based Sanskrit-English Machine Translation”, proc. of International Conference on Circuits Power and Computing Technologies (ICCPCT-2017), IEEE. doi: 10.1109/ICCPCT.2017.8074251, Vol-16, Issue-9.

[13] Gupta V.K., Tapaswi N., and Jain S. (2013), “Knowledge representation of grammatical constructs of Sanskrit language using rule based Sanskrit Language to English Language machine translation”, proc. of International Conference on Advances in Technology and Engineering (ICATE-2013), IEEE. doi: 10.1109/ICAdTE.2013.6524744.

[14] Kale M.R. (2016), “A Higher Sanskrit Grammar”, Motilal Banarsidass Publishers Pvt. Ltd. 11th Reprint Edition.

[15] Patel P. (2013), “Sanskrit Vyakaran Prabodh”, AB Shah Publications, Saraswati Pustak Bhandar, Ahmedabad, Edition.

[16] Phadke G. (1852), “The Principles of Gujarati Grammar”, digitized by Google. Available online: https://books.google.co.in/books?id=E0IAAAAQAAJ.

[17] Edalji S. (1867), “A Grammar of Gujarati Language”, digitized by Google.

[18] Patel D. (2015), “Gujarati Vyakaran Ane Lekhan”, Gujarati Sahitya Prakashan, Ahmedabad, First Edition.

[19] Raulji J., and Saini J.R. (2019), “Morphological Analyzer for Sanskrit Language”, proc. of International Conference on The Journey of Indian Languages: Perspectives on Culture and Society, 3:198-209.

[20] Raulji J., and Saini J.R. (2017), “Generating Stopword List for Sanskrit Language”, proc. of 7th IEEE International Advance Computing Conference (IACC- 2017), pp. 799-802. doi:

10.1109/IACC.2017.0164.

[21] Automatic BLEU MT Evaluation, Available online: https://www.letsmt.eu/Bleu.aspx, Accessed since 2017.

[22] Papineni K., Roukos S., Ward T., and Zhu W. (2002), "BLEU: a Method for Automatic Evaluation of Machine Translation", proc. of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, pp. 311-318. Available online: https://aclanthology.org/P02- 1040.pdf.

[23] Automatic Language Processing Advisory Committee (ALPAC) (1966), "Language and Machines – Computer in Translation and Linguistics", report by ALPAC, National Academy of Sciences National Research Council, Washington D.C.

[24] Singh M., Kumar R., and Chana I. (2020), "Hybrid machine translation system using deep learning", ASM Science Journal, 13(2):31–45. Available online: https://www.akademisains.gov.my/asmsj/?mdocs- file=4679.

[25] Akhand M.A.H., Roy A., Dhar A.C., and Kamal M.A.S. (2021), "Recent Progress, Emerging Techniques, and Future Research Prospects of Bangla Machine Translation: A Systematic Review", International Journal of Advanced Computer Science and Applications, 12(9):291-307. doi: 10.14569/IJACSA.2021.0120933.

[26] Sabtan Y.M.N.S., Hussein M.S.M., Ethelb H., and Omar A. (2021), "An Evaluation of the Accuracy of the Machine Translation Systems of Social Media Language", International Journal of Advanced Computer Science and Applications, 12(7):406-415. doi: 10.14569/IJACSA.2021.0120746.

[27] Ehab R., Gadallah M., and Amer E. (2019), "English- Arabic Hybrid Machine Translation System using EBMT and Translation Memory", International Journal of Advanced Computer Science and Applications,10(1):195-203. doi: 10.14569/IJACSA.2019.0100126.

[28] Pudaruth S., Sookun L., and Ruchpaul A.K. (2013), "English to Creole and Creole to English Rule Based Machine Translation System", International Journal of Advanced Computer Science and Applications, 10(1):25-29. doi: 10.14569/IJACSA.2013.040805.

[29] Singh A.K., Dave S., Mehta S., and Prathosh A.P. (2022), "A data driven approach for Sanskrit Derivative Nouns analysis", proc. of ACM CODS- COMAD-2022 5th Joint International Conference on Data Science & Management of Data, pp. 288–289. doi: 10.1145/3493700.3493745.

[30] Li J., and Girrbach L. (2022), "Word Segmentation and Morphological Parsing for Sanskrit", Computation and Language arxiv, Cornell University, doi: 10.48550/arXiv.2201.12833.

[31] Biltoo A.K. (2022), "First Steps Towards Sanskrit: Language, Linguistics and Culture", Routledge, Taylor and Francis, First Edition. doi: 10.4324/9780429325434.

[32] Sandhan J., Behera L., and Goyal P. (2022), "Systematic Investigation of Strategies Tailored for Low-Resource Settings for Sanskrit Dependency Parsing", Computation and Language arxiv, Cornell University, doi: 10.48550/arXiv.2201.11374.

[33] Raulji J.K., and Saini J.R. (2019), "Sanskrit Lemmatizer for Improvisation of Morphological Analyzer", Journal of Statistics & Management Systems,22(4):613-625. doi: 10.1080/09720510.2019.1609186.

[34] Raulji J.K., and Saini J.R. (2020), "Sanskrit-Gujarati Constituency Mapper for Machine Translation System", proc. of IEEE Bombay Section Signature Conference (IBSSC-2019), pp.1-8. doi: 10.1109/IBSSC47189.2019.8972989.

[35] Tailor C., and Patel B. (2020), "Chunker for Gujarati Language Using Hybrid Approach", proc. of Rising Threats in Expert Applications and Solutions, Advances in Intelligent Systems and Computing, 1187:77-84. doi: 10.1007/978-981-15-6014-9_10.

[36] Patel C.D., and Patel J.M. (2021), "Influence of GUJarati STEmmeR in Supervised Learning of Web Page Categorization", International Journal of Intelligent Systems and Applications, 13(3):23-34. doi: 10.5815/ijisa.2021.03.03.

[37] Chauhan A., and Shah A. (2021), "Improving Semantic Coherence of Gujarati Text Topic Model Using Inflectional Forms Reduction and Single-letter Words Removal", ACM Transactions on Asian and Low-Resource Language Information Processing, 20(1):1-18. doi: 10.1145/3447760.

[38] Shah P., and Swaminarayan P. (2021), "Lexicon-based sentiment analysis on movie review in the Gujarati language", International Journal of Information Technology, Communications and Convergence, 4(1):63-72. doi: 10.1504/IJITCC.2021.119111.

[39] Raval D., Pathak V., Patel M., and Bhatt B. (2022), "Improving Deep Learning based Automatic Speech Recognition for Gujarati", ACM Transactions on Asian and Low-Resource Language Information Processing, 21(3):1–18. doi: 10.1145/3483446.

[40] Doshi P.D., and Vanjara P.A. (2021), "A Comprehensive Survey on Handwritten Gujarati Character and Its Modifier Recognition Methods", proc. of Information and Communication Technology for Competitive Strategies (ICTCS-2020), 191:841- 850. doi: 10.1007/978-981-16-0739-4_79.

[41] Patel H., Patel B., and Lad K. (2021), "Jodani: A spell checking and suggesting tool for Gujarati language", proc. of 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence- 2021).pp.94-99. doi: 10.1109/Confluence51648.2021.9377072.

[42] Raulji J.K., and Saini J.R. (2020), "Sanskrit Stopword Analysis through Morphological Analyzer and its Gujarati Equivalent for MT System", proc. of the International Conference on ICT for Sustainable Development (ICT4SD-2019), Panaji, India. Vol. 93, pp. 427-433. doi: 10.1007/978-981-15-0630-7_42.

[43] Saini J.R., and Raulji J.K. (2020), "Peer Analysis of "Sanguj" with Other Sanskrit Morphological Analyzers", proc. of 2nd International Conference on Computing Analytics and Networking (ICCAN-2019), Bhubaneshwar, India. Vol. 1119, pp. 65-73. doi: 10.1007/978-981-15-2414-1_7.

Detection and Analysis of Oil Spill using Image Processing

Myssar Jabbar Hammood AL-BATTBOOTTI, Nicolae GOGA, Iuliana MARIN
Faculty of Engineering in Foreign Languages
University POLITEHNICA of Bucharest
UPB, Bucharest
Romania

***Abstract*—Thousands of oil spills occur every year on offshore oil production platforms. Moreover, ships that crosses rivers to reach the harbor cause spills each year. The current study focuses on IRAQI marine and rivers, especially Al-Bakr, Khor al-Amaya, ABOT oil terminal and SHAT AL-Arab river inside Al Başrah oil terminal. In order to mitigate and manage oil spill impacts, an unmanned aerial vehicle has proven to be a valuable tool in mitigating and managing incidents. To achieve high accuracy, the objective of the current research is to analyze captured images for rivers, identify oil pollution and determine its location. The images were taken from the Iraqi Regional Organization for the Protection, General Company for Ports of Iraq, Iraqi Ministry of Environment and online websites. In the current paper is presented a software framework for detecting oil spills, pollution in rivers and other kinds of garbage. The framework based on artificial intelligence is divided into two parts: a training model and an operational model. In the training model part, a machine learning model is applied, which is one of the fastest and most accurate methods, integrated inside PipelineMLML. Thus, the object detection technique used can identify one or more categories of objects in a picture or video. Furthermore, the locations of objects can be identified with the help of neural networks. In the operational mode, models can identify oil spills in images.**

***Keywords*—PipelineMLML; oil spill; artificial intelligence; machine learning**

I. Introduction

Oil spills damage the environment, destroy natural resources, and harm public health. The environment is often exposed to huge pollution following oil spill incidents. As proof of the significant pollution risk associated with offshore oil and shipping crude oil products through these waters, one only needs to consider the British Petroleum Deepwater Horizon incident in 2010 and the huge marine disaster. The information technology era has always been important for different industries, therefore, software systems based on machine learning models have been developed to perform object detection in the past years. The accuracy and efficiency of object detection improved dramatically with the advent of machine learning and deep learning. As a result, these technologies were pivotal for the advancement of computers. The common programming library, PipelineMLML, is designed to enhance the accuracy and reliability of an object detection system.

Deep neural networks (DNN), convolutional neural networks (CNN) have gained fame for their ability to process visual data. Recently, sensor image-based Artificial intelligence has become a key component in a wide range of software applications because of its capability to solve several problems [7]. It can, for example, identify objects and recognize them as one of these issues. A set of applications that use these features are autonomous driving, video surveillance, and healthcare [8, 9, 10].

The current research combines these advances in artificial intelligence, especially in the detection and recognition of objects in the oil and gas industry in order to avoid unwanted incidents based on video monitoring and machine learning. Drones are used to capture photographs or video of the observed regions which oil spills can occur. The triggered files are uploaded on the system's framework to be further analyzed and visualized from a ground station by an oil and gas expert. Machine learning models have been created for two cases, when an oil spill happens in the environment, as well as when it is not present. The models are used to be consumed and to predict the occurrence of oil spill disasters. The overall solution is appropriate spill management and continuous surveillance, speeding up the decision making process. Drones can quickly collect critical data. A big challenge, like reflectance can be solve by adding to the drone's camera a polarizing filter. The problem of oil dissolution is tackled by our solution, as it analyzed four factors, namely a normal spill, areas of water color retreat, leaks in surrounding soil based on their color and crude oil spills.

The limitations of the proposed solution are given by battery life and weather conditions, like extreme wind. Gasoline engines for drones can be an alternative, but they will be noisy and emit exhaust.

Due to there being no effective method of automatic detection for oil pipelines, it forces oil companies to dedicate a lot of human resources to patrol them manually [1]. As a result, operating costs of enterprises are high, while the used detection methods cannot stop oil pollution. This is because manual checks or different technologies that exist are not able to detect oil spills on time.

In the current study the design and development of an intelligent and automated method for detecting oil spills using image merge with deep convolutional neural networks (DCNNs) and a cluster of algorithms. By using these technologies, one of the most effective solutions for identifying and recognizing spills can be created. In summary, this study contributes to: 1. provide a method to ensure a smooth and uninterrupted flow of river environments monitoring; 2. provide full vision and awareness of oil pollution and different garbage.

The paper is organized according to the following structure: Section 2 describes related work regarding object detectors based on ML.NET, Artificial intelligence detection. Section 3 specifies the materials and methods included in the current paper, concerning drone data. The processing model method is outlined in Section 4. Section 5 mentions usage of the detection method and discusses the results obtained based on experimental analysis. Finally, Section 6 summarizes the main conclusions drawn from the current study, while Section 7 details the ideas for future development of the applied detection, recognition, and tracking algorithms.

II. Literature Review

Geographic information systems, spatial statistical methods and computational modeling are used to assess the impacts of oil spills [11, 12, 13, 14, 15]. Special attention must be provided in order to monitor and predict the appearance, existence of oil spills that endanger the marine environment, having a crucial impact over time.

Several closely related works have been analyzed in the context of oil spills in the surrounding environment. Existent state-of-the-art models for oil spill risk assessment include multiple factors and focus on uncertainty. A research regarding the Baltic Sea spills from ship wrecks included into its state-of-the-art model a probabilistic calculation of the sum of factors (construction work, corrosion, diving, earth quakes, military activity, ship traffic, extreme weather and trawling) that can lead to dangerous consequences [21]. The gap which exists in this model, but is considered by our proposed system is that real-time events happen in a short amount of time and drones ease decision making. Other oil spill models [22] do not consider oil dissolution [23, 24]. The paper's model analysis this factor, along with watercolor retreat areas and learks in surrounding soil.

A solution used in Brazil involved the usage of radars in order to determine marine oil rig spills as well as it can determine features and feed based on a logistic regression classifier method [2]. The methods included convolutional neural networks and logistic regression classifiers. The best results were obtained for vertical direction of electromagnetic waves, namely polarized ones. In the case of polarized images, more target information is obtained by the object as the light gets reflected [20]. Polarization ensured accuracy of the acquired image information.

Another research done by Germans and Mongolians focused on oil spill detection in the Caspian Sea, also based on radar images, the water surface images which presented dark slicks on a bright background were analyzed using the EnviView and NEST programs [3]. The outcomes were obtained after doing the comparison between their contamination and navigation maps. The training part of the machine learning algorithm consists in automatically image annotations and matching them with infrared images to generate a dataset, but the training stopped when the neural network has failures when it cannot improve during five consecutive epochs. 7 CNN segmentation architectures and 8 feature extractors are used during the analysis.

A researcher from Belgium [4] studied the Antwerp port environment using drones with thermal infrared cameras. As in paper [2], the used method involved convolutional neural networks. The accuracy reached 89% when involving infrared cameras for oil spill detection in water. A group of researchers from Algeria and France experienced a hybrid swarm UAV monitoring, The UAV swarm plan the trajectory and gather nautical data based on a modified Boustrophedon algorithm, along with an unsupervised natural image classification.

Crude oil spillage is another factor which was analyzed in the current paper. Cases were reported in Southern Iraq, at Rumaila, Basra [18]. Based on the extracted soil samples, it was found that the international standard limits of heavy metals content were exceeded. The contaminated soil can affect the neighboring regions. Our research does not imply physical samples of soil. Real-time image processing coming from drones is sufficient to determine the risk of an oil spill.

A group of researchers studied the impact of crude oil spill using the multispectral satellite Landsat 8 OLI imagery [19]. The analysis was done using machine learning models. 1205 polluted vegetation and wetland hectares were studied during 2015-2018. In the following two years, based on the normalized difference vegetation index and chlorophyll index vegetation, it was found that the region started to recover according to remediation initiatives. The downside of the multispectral Landsat 8-OLI remote sensing imagery is that its' measurements are influenced by irradiance variation, sensor calibration and drift, atmospheric attenuation and radiance path. The drone of our solution can scan the observer region from distance, as well as closely. Moreover, the drone will be trained to navigate in regions with reflectance of the surrounding terrain.

The researcher did not study the influence of unmanned aerial vehicle (UAV) speed variations which have an effect upon the consumed energy [5], reaching a lower level of energy consumption. Another group of researchers from Latvia [6] also included a group strategy, as for paper [5], based on multi remote piloted aircrafts, such that oil spills are analyzed.

The previous mentioned papers are analyzed in Table 1 from the point of view of strengths and weaknesses. Our contribution is also specified as a comparison to the analyzed closely related work (selection made; similar conclusions can be drown for other reported research as well).

The result of the conducted analysis shows that the purpose of the current paper is clearly defined. Three core areas are combined to provide a solution to mitigate oil spills, sense danger at distance, determine geographic information and perform computational modeling.

TABLE I. COMPARISON BETWEEN RELATED WORK AND CURRENT PROPOSED SOLUTION

Closely related work	Comparison		
	Strength	***Weakness***	***Our contribution***
Oil Rig Recognition Using Convolutional Neural Network on Sentinel-1 SAR Images [2]	1.The system works in various weather conditions, as well as during night or day to detect spills by using synthetic aperture radar. 2. False alarms are determined using a convolutional neural network based on several layers which are named VGG-16 and VGG-19.	1. Average accuracy of fifty sampling tests varied between 86.4% and 84.1% using the VGG-16 and VGG-19 architectures. 2.The used images were of average resolution and big swaths as result that will affect the accuracy.	1. Our mean global accuracy based on the test of 127 situations reached 82.80%, while the precision was 87.55%. 2. The images were of small dimensions, 275 x 183 pixels, such that they do not occupy a lot of space.
Oil Spill Detection in the Kazakhstan Sector of the Caspian Sea with the Help of ENVISAT ASAR Data [3]	1. According to the visual interpretation of ASAR ENVISAT radar images to determine oil pollution sources. 2. Boundaries, bathymetry, coastline, infrastructure were determined using the ArcGIS program.	1. Radar images of sea surface captured during the period April-October can be used, because for the rest of the time, the area is covered with ice, making oil pollution identification impossible. 2. For some cases, oil slicks detection based on radar screens can be difficult due to winds and streams which appear as dark spots.	1. Our proposed solution took into consideration 4 factors, namely normal oil spill, watercolor retreat areas, leaks in surrounding soil and crude oil spills. 2. Different weather conditions were part of the analysed 127 scenarios.
Oil Spill Detection Using Machine Learning and Infrared Images [4]	1. The study uses unmanned aerial vehicles and a thermal infrared camera to also be able to offer surveillance during nighttime. 2. The created dataset is part of the convolutional neural network (CNN) training process.	1. Images are classified based on pixel color value, determining the need of optimal weather conditions. Direct sunlight and other disturbances need to be avoided.	1. The training model included a machine learning model. The operational model was based on PipelineMLML.NET. 2. The images are classified based on a model that uses images for all four studied factors which are monitored in the current paper. Different whether conditions were taken into account.
Oil Spill Detection Using Multi Remote Piloted Aircraft for the Environmental Monitoring of Sea Aquatorium [6]	1. Analysis of oil spill detection based on multi remote piloted aircrafts (RPA), namely UAV swarms. 2. A Centralized multi-RPA algorithm is used to detect oil spills by applying a decision tree algorithm and ant colony algorithm.	1. Failure can appear in the case of a monitoring mission even if just a single RPA is not functional. 2. The ant colony algorithm takes a long computational time.	1. Monitoring will be done independently by every drone. 2. The computational times will depend on the server side capabilities. The used computer for the detection of oil spills was a Macbook Pro laptop with an i9 processor, 2.30 GHz, and 16 GB RAM.
Spatio-Temporal Analysis of Oil Spill Impact and Recovery Pattern of Coastal Vegetation and Wetland Using Multispectral Satellite Landsat 8-OLI Imagery and Machine Learning Models [19]	1. Crude oil detection based on Support Vector Machine and Random Forest machine learning models to classify polluted and nonpolluted vegetation and wetlands. 2. Images were captured using a Landsat 8 OLI multispectral satellite due to low cloud cover.	1. Clouds can make the observed regions to become unreachable. 2. Sun angle atmospheric correction was performed in order to acquire better quality images.	1. Drone monitoring can be done even during cloudy weather based on a cost-effective solution, compared to multispectral satellites. 2. No sun angle atmospheric correction is needed.

III. FRAMEWORK OVERVIEW

A. Materials and Work Component

This research focus to determine oil spills relies upon images. At present, deep learning can be used to analyze images so that UAVs in the future work, are able to detect oil spills automatically and accurately.

The current paper describes a framework for object detection and recognition based on drones equipped with a high-resolution camera and gimbal path controller. The used computer for the simulation of the proposed framework was a Macbook Pro laptop with an i9 processor, 2.30 GHz, and 16 GB RAM. The framework uses computer neural networks to locate and recognize objects in images or videos in real-time. Object detection and recognition can be performed in many programming languages. In this paper, the algorithm is built using ML.NET, an open-source, cross-platform, and machine learning framework designed by Microsoft and publicly released in 2018 to offer the power of machine learning (ML) to .NET applications to cover various scenarios, like sentiment analysis, oil price prediction, recommendations, and classification of images.

The system identifies the class of objects (oil pollution and illegal bilge water dumping), and their precise location. Classification can predict objects in an image. Object recognition and detection modules were designed for use with oil and gas pipelines. The solution can also be adapted to different environments by adjusting these frameworks according to the same structure.

B. Study Area

As part of Phase 2, part of the images with which the model is trained are taken from the deep sea south of Iraq (for a more detailed description of all areas from where images are taken for training see next section) the main oil harbor at Al Başrah [16], known as ABOT, the marine loading terminal for oil being located at almost 50 kilometers south-east the Al-Faw Peninsula, in the Persian Guf. Together with its other terminal, the Khawr al ʻAmīyah Oil Terminal, KAAOT, both terminals offer the main export source for more than 80% of Iraq's gross domestic product. The SHAT AL-Arab river is used to reach the Al Başrah oil terminal and its fauna and flora are put in danger in case of a disaster, like an oil spill.

The exported crude oil which is produced in the southern Iraqi oil fields, is carried via three 1.2 m diameter pipelines to the southern part of the Al-Faw peninsula, as well as undersea to the ABOT platform [16]. One 1.2 m and two 0.81 m pipelines supply the KAAOT platform.

The facilities of ABOT are able to transfer daily up to 3 million barrels (Mbbl) of oil when all four supertanker berths operate at maximum capacity and have a maximum draft of 21 m [17]. Three single-point mooring systems (SPM) were included in the year 2012, each having a design rating of 800 thousand barrels of oil daily, and two additional SPMs to increase the total loading capacity to 6.4–6.6 Mbbl of oil per day [17]. The KAAOT facility possesses a shallower depth and its two berths can be used for Suezmax oil tankers with capacities up to 1 Mbbl and has a transfer capacity of about 240 kbbl of oil daily [17].

C. Oil Spill Dataset

In this research to identify oil spills, the dataset gathered with help of: 1) http://ropme.org (Regional Organization for the Protection - ROPME) was initially established in 1979, was convened in Kuwait, 1978. The Conference adopted in 1978 included the Action Plan for the Protection and Development of the Marine Environment and the Coastal Areas of Bahrain, Iran, Iraq, Kuwait, Oman, Qatar, Saudi Arabia and the United Arab Emirates the Kuwait Regional Convention for Co-operation on the Protection of the Marine Environment from Pollution. ROPME Information regards the geographic coordinates and timestamps of the pollution provided by the General Company for Ports of Iraq through the pollution department. 2) General Company for Ports of Iraq / pollution department; 3) Ministry of environment / Southern office; 4) online websites.

IV. Processing Model Method

The current processing model used images taken from the Iraqi Regional Organization for the Protection, General Company for Ports of Iraq, Iraqi Ministry of Environment and online websites.

The future research will focus on how the raw dataset will be gathered from several aerial images. The trained model which was used in the current paper will analyze and identify an oil spill based on the main algorithm framework explained in the present article. The analysis process will be in real-time or after transferring data to the ground station. Moreover, the operator will make sure to double-check the spill detection alert.

Every image will be resized and segmented on board. When oil is detected, the system will gather all related information. The information will regard two major factors. The first factor, onboard cameras, will capture raw aerial images that provide information about spill size and locations. The second factor, UAV hardware will provide altitude, flight speed, flight path and GPS coordinates to help identify both the UAV and spill locations during data collection. Moreover, the triggered data will be sent to the ground station, where actions can be taken.

The current paper application developed with ML.NET contains several steps (Fig. 1):

- Load data – Raw data is loaded into memory.
- Create a pipeline – The pipeline contains the steps that transform data or train a machine learning algorithm. ML.NET offers different transformational steps, like one-hot encoding and several machine learning algorithms.
- Train a machine learning model – When the pipeline is created, training can begin. This is done based on the Fit() method.
- Evaluate – The evaluation of the model can be done at any point and supplementary decisions can be taken according to evaluations.
- Save – When the model is trained, it is saved into a file. The resulting application is created based on one microservice which trains and evaluates the machine learning model, and the other microservice uses it.
- Load – The created machine learning model can be loaded and applied for further predictions.

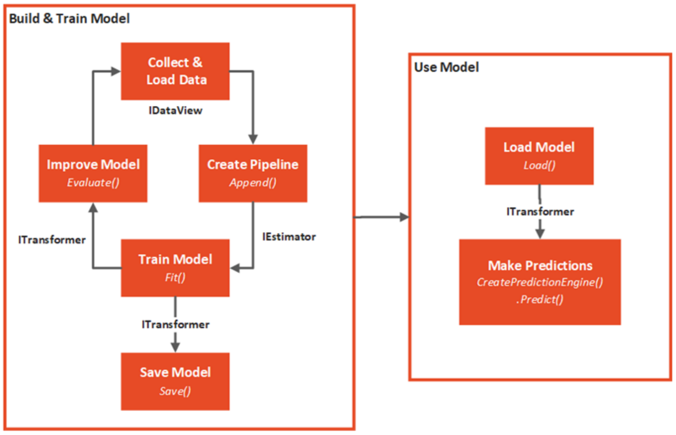


Fig. 1. Machine Learning Part Building Model Steps.

V. Detection Method and Results

The system is designed to capture pictures and video along pipelines that extend in or out of fields and through the trained model, the algorithm begins to analyze collected data either in real-time or after transmission to the ground station where it starts to analyze data.

The following factors are taken into consideration to detect possible leaks:

- A normal spill occurred.
- Areas of water color retreat were identified.
- Leaks in surrounding soil were detected by observing their color.
- A crude oil spill occurred.

UAVs also have the capability to detect anomalies along with pipeline networks normal category spills, such as construction or roadwork that might compromise pipeline integrity. If any unusual situations occur, namely leaks, explosions, attacks, or any other unusual events, they provide real-time video and alert notifications. In order to assist the teams to value different situations and provide clear information to emergency crews, the detection framework is depicted in Fig. 2.

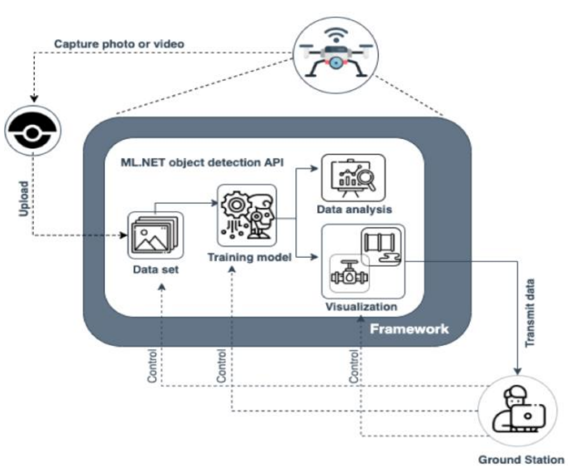


Fig. 2. Detection Framework.

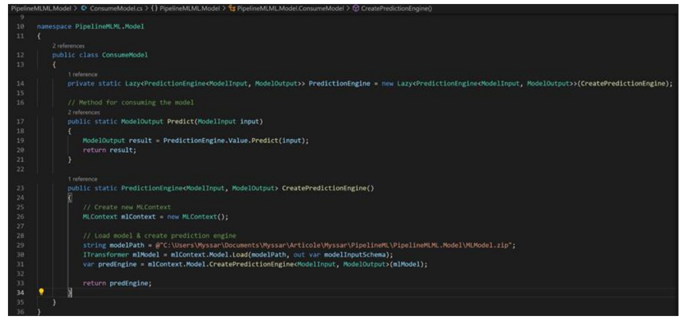


Fig. 3. ML.NET Machine Learning Model.

The training model of the proposed framework included a ML.NET machine learning model, as illustrated in Fig. 3. The operational model was implemented using ML.NET.

127 different scenarios were tested, namely, 1) normal spill for 40 images, 2) watercolor retreat areas for 50 images, 3) larks in surrounding soil for 20 images, 4) crude oil spill for 23 images. The previous mentioned factors were included in the study, as illustrated in Fig. 4.

1. Normal Spill.
2. Watercolor Retreat Areas.
3. Leaks in Surrounding Soil.
4. Crude Oil Spill.

Fig. 4. Leak Factors.

The probabilities of the studied leak factors based on 127 different images are displayed in Fig. 5. On the X axis are placed the studied situations, namely normal spill, watercolor retreat areas, leaks in surrounding soil and crude oil spill. On the Y axis are the true and false probabilities for every studied case, the values varying between 0 and 1. The rectangles represent the most frequent range of points belonging to every category. All the dots are probability values.

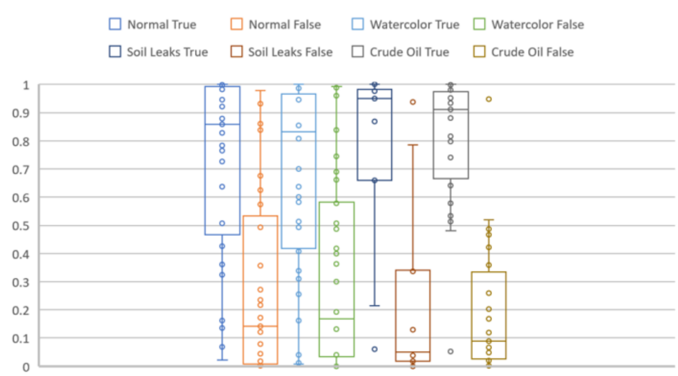


Fig. 5. Probability Variations based on Studied Leak Factors.

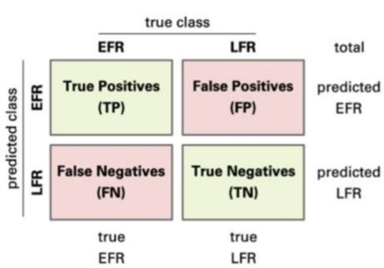
TABLE II. CLASSIFICATION TABLE BASED ON THE FOUR STUDIED SITUATIONS

Situation	**Probability values**			
	True Positive	***False Positive***	***False Negative***	***True Negative***
Normal spill	0.8777745	0.122255	0.233689	0.766311
Watercolor retreat areas	0.858818	0.121182	0.265185	0.734815
Leaks in surrounding soil	0.900071	0.099929	0.112232	0.887758
Crude oil spill	0.847981	0.152019	0.266434	0.733566

The true positive, false positive, false negative and true negative values for all the analyzed cases are displayed in Table 2. The determined accuracies were 71.67% for normal spills, 70.88% for watercolor, 73.13% for soil leaks and 79.74% for crude oil spills.

According to the overall results, it was clear that in 99% of the cases, the triggered situation was a broken ship / pipe, while for 28 cases, the situation was good. The following table shows the probabilities for the four analyzed situations.

The highest certainty value was obtained for leaks in surrounding soil, namely 90%, followed by normal spills, with a probability of 87.77%. Close by were the water retreat areas with 85.88% and crude oil spill with 84.79%. The overall true positive rate in case of an oil spill was 87.11%, with a false positive value equal to 12.38%. The false negative value was determined to be 21.93% and true negative of 78.06%.



$$\mathrm{PR} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}$$

$$\mathrm{RE} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$$

$$\mathrm{CA} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}$$

$$\mathrm{F}_1 = \frac{2\mathrm{TP}}{2\mathrm{TP}+\mathrm{FP}+\mathrm{FN}}$$

Fig. 6. Confusion Matrix.

Precision, recall, accuracy and F1-measure formulas as illustrated in Fig. 6. Precision (PR) indicates how accurate the proposed oil detection model is and it is equal to 87.55%. Recall (RE) computes how many actual positives are captured in the model and the value for it is equal to 79.88%. Accuracy (CA) is influenced by true negative values and it is equal to 82.80%. A better measure is represented by F1-measure which is a means to determine a balance between precision and recall. The value for it is equal to 83.54%.

VI. CONCLUSION

Oil pollution prevention and response in marine accidents consists of a series of steps and procedures. In this study, it was demonstrated how deep learning and artificial intelligence can be used to accomplish this goal.

The framework proposes a clearly useful method despite its condition being dependent on hypothetical parameters of any given event (normal spill, watercolor retreat areas, leaks in surrounding soil, crude oil spills). 127 cases were analyzed and the results provide good outcomes. Dangerous situations have been correctly determined based on the precision, recall, accuracy and F1-measure values which have been computed, proving that the solution can be further used in real-case scenarios. The system proved that it filled the gap of oil dissolution detection.

VII. FUTURE WORK

The proposed system will be useful especially in waterways near populated areas. The reason is caused by the fact that densely populated settlements have narrow passageways, canals and straits, leading to extensive oil transportation. Accurate decisions and speed are imperative in such sensitive areas. For the responsible coastal authority, accident response management is crucial and the costs for cleaning oil spills need to be decreased.

Future work will include an oil spill system modeling which will help to integrate breakthroughs. The demand for oil will not decrease soon and the production of spill accidents will also endure. Models will be created to understand the emergence of oil spills, spatiotemporal dynamics and to enhance the response to mitigate such situations. Moreover, different datasets will be tested to prove the scalability of the currently proposed work.

REFERENCES

[1] Z. Jiao, G. Jia, and Y. Cai, "A New Approach to Oil Spill Detection that Combines Deep Learning with Unmanned Aerial Vehicles," Computers & Industrial Engineering, vol. 135, pp. 1300-1311, 2019.

[2] L. E. Falqueto, J. A. S. Sa, R. L. Paes, and A. Passaro, "Oil Rig Recognition Using Convolutional Neural Network on Sentinel-1 SAR Images," IEEE Geoscience and Remote Sensing Letters, pp. 1–5, 2019.

[3] A. Holstein, M. Kappas, P. Propastin, and T. Renchin, "Oil Spill Detection in the Kazakhstan Sector of the Caspian Sea with the Help of ENVISAT ASAR Data," Environmental Earth Sciences, vol. 77(198), pp.1-11, 2018.

[4] T. De Kerf, J. Gladines, S. Sels, and S. Vanlanduit, "Oil Spill Detection Using Machine Learning and Infrared Images," Remote Sensing, vol. 12(24), 2020.

[5] S. Bella, G. Belalem, A. Belbachir, and H. Benfriha, "HMDCS-UV: A Concept Study of Hybrid Monitoring, Detection and Cleaning System for Unmanned Vehicles," Journal of Intelligent and Robotic Systems, vol. 102(2), 2021.

[6] A. Urbahs, and V. Zavtkevics, "Oil Spill Detection Using Multi Remote Piloted Aircraft for the Environmental Monitoring of Sea Aquatorium," Environmental and Climate Technologies, vol. 24(1), pp. 1-22, 2020.

[7] K. Yetilmezsoy, B. Ozkaya, and M. Cakmakci, "Artificial Intelligence-Based Prediction Models for Environmental Engineering," Neural Network World, vol. 21(3), pp. 193-218, 2011.

[8] L. Li, "The UAV Intelligent Inspection of Transmission Lines," Proceedings of the 2015 International Conference on Advances in Mechanical Engineering and Industrial Informatics, pp. 1542–1545, 2015.

[9] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39(6), pp. 1137–1149, 2015.

[10] V. Raimondi, L. Palombi, D. Lognoli, A. Masini, and E. Simeone, "Experimental Tests and Radiometric Calculations for the Feasibility of Fluorescence Lidar-Based Dis- crimination of Oil Spills from UAV," International Journal of Applied Earth Observation and Geoinformation, vol. 61, pp. 46–54, 2017.

[11] A. Zeggada, F. Melgani, and Y. Bazi, "A Deep Learning Approach to UAV Image Multilabeling," IEEE Geoscience and Remote Sensing Letters, vol. 14(5), pp. 694–698, 2017.

[12] C. Yuan, Y. Zhang, and Z. Liu, "A Survey on Technologies for Automatic Forest Fire Monitoring, Detection, and Fighting using Unmanned Aerial Vehicles and Remote Sensing Techniques," Canadian Journal of Forest Research, vol. 45(7), pp. 783–792, 2015.

[13] J. R. Nelson, and T. H. Grubesic, "Oil Spill Modeling: Computational Tools, Analytical Frameworks, and Emerging Technologies," Progress in Physical Geography: Earth and Environment, vol. 43(1), pp. 129-143, 2018.

[14] S. Abdikan, Ç. Bayık, and F. Balık Şanlı, "The Role of Sar Remote Sensing To Detect Oil Pollution And Emergency Intervention," Turkish Marine Research Foundation (TUDAV), pp. 157-175, 2018.

[15] D. Aydın, and M. S. Solmaz, "Modeling of the Oil Spill in M/V Lady Tuna Accident and the Evaluation of the Response Operation in Simulated Condition with PISCES II," International Journal of Environment and Geoinformatics, vol. 6(2), pp. 213-224, 2019.

[16] M. Kaschubsky, "Offshore Petroleum Security: Analysis of Offshore Security Threats, Target Attractiveness, and the International Legal Framework for the Protection and Security of Offshore Petroleum Installations," Australian National Centre for Ocean Resources and Security, Faculty of Law, University of Wollongong, 2011.

[17] A. Johnston, K. O'Connor and Y. Rawal, "Al Basrah Oil Terminal. Basrah, Iraq," Office of the Special Inspector General for Iraq Reconstruction, pp. 5-7, 2007.

[18] K. Khwedim, "Crude Oil Spillage and the Impact of Drilling Processes on the Soil at Rumaila Oil Field- Southern Iraq," Iraqi Journal of Science, vol. 57(2A), pp. 918-929, 2016.

[19] A.-L. Balogun, S. T. Yekeen, B. Pradhan, and O. F. Althuwaynee, "Spatio-Temporal Analysis of Oil Spill Impact and Recovery Pattern of Coastal Vegetation and Wetland Using Multispectral Satellite Landsat 8-OLI Imagery and Machine Learning Models," Remote Sensing, vol. 12(7), pp. 1-25, 2020.

[20] R. Li, Q. Chu, K. Zhao, and J. Piao, "Foggy Image–Sharpening Method with Multi-channel Polarization Information System," Advances in Mechanical Engineering, 11(3), 1-12, 2019.

[21] P. Amir-Heidari, L. Arneborg, J. F. Lindgren, A. Lindhe, L. Rosen, M. Raie, L. Axell and I.-M. Hassellov, "A State-of-the-Art Model for Spatial and Stochastic Oil Spill Risk Assessment: A Case Study of Oil Spill from a Shipwreck," Environment International, vol. 126, pp. 309-320, 2019.

[22] P. Keramea, K. Spanoudaki, G. Zodiatis, G. Gikas and G. Sylaios, "Oil Spill Modeling: A Critical Review on Current Trends, Perspectives, and Challenges," Journal of Marine Science and Engineering, vol. 9, no. 181, pp. 1-38, 2021.

[23] P. D. Yapa, and F. Chen, "Behavior of Oil and Gas from Deepwater Blowouts," Journal of Hydraulic Engineering, vol. 130, no. 6, pp. 540–553, 2004.

[24] L. H. Sim, "Blowout and Spill Occurrence Model," Oregon State University: Corvallis, USA, 2013.

Smart Information Retrieval using Query Transformation based on Ontology and Semantic-Association

Ram Kumar*, S. C. Sharma
Indian Institute of Technology Roorkee, India

***Abstract*—A notable problem with current information retrieval systems is that the input queries cannot express user information needs properly. This imprecise representation of the query hampers the effectiveness of the retrieval system. One method to solve this problem is to transform the original query into a more meaningful form. This paper proposes an ontology-based retrieval system that transforms initial user queries using domain ontologies and applies semantic association during the indexing process. The proposed system performs a semantic matching between an ontologically enhanced query and index to capture query-related terms. To show the performance of the proposed system, it is evaluated using standard parameters like precision, recall, and NDCG. In addition, the authors presented a comparison between the proposed and existing retrieval systems on three test datasets. Experimental results on these datasets indicate that the use of ontology and semantics has significantly increased the retrieval efficiency obtained by baseline. This work highlights the importance of ontology and semantics in information retrieval.**

Keywords*—*Ontology; semantics; information retrieval; query transformation; indexing

I. Introduction

In this digital age, the fast rate of data generation created a massive collection of data. It becomes a challenging task for computer users to extract relevant information from this large data collection. World Wide Web [1] is one such collection with billions of such web pages. Information Retrieval System (IRS) enables web users to find the desired information from such extensive resources. Search engines are commercially available IRS, which play a vital role in finding information online. Nowadays, the search engine has become the primary means to find websites. The simple Web search engine retrieves information from Web pages using keyword matching between queries and documents. The Web is rapidly growing, and different users seek focused information; now, it has become a must for search engines to utilize semantic techniques to satisfy users' information needs [2]. Using Semantic search, the machine can also understand the user's interest and the context in which the search is issued; this will help in providing the most relevant information to the user's search need. Traditional search engines need to adopt new changes to find exact information from the Web.

Ontology-based information retrieval is a small step in this direction. This can meet those challenges that were not met by traditional retrieval systems. The authors propose an ontology-based retrieval system for finding the most relevant search results.

The Internet is a collection of interlinked documents (billions in numbers) distributed over the most extensive network. In the last two decades, the web has grown exponentially. The large scale of the Web made it almost impossible to retrieve desired information without any tool. That's why internet user searches the web for the topic of their interest using a search engine. A search engine [3] is a software program or tool to find information from many web pages distributed over the internet. A search engine searches the document for entered set of keywords and returns a list of results in links to relevant resources. These links are Uniform Resource Locators (URLs) of those documents where any or all of the searched keywords can be found. A search engine provides results quickly using high-speed systems working globally known as index servers. The searched URLs are accessed using a program called a web browser.

Sir Tim Berners-Lee defines the Semantic Web [4] as an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation. Semantic search is a search where humans and machines try to find concepts behind terms used by different users. The Semantic Web supports more efficient discovery, automation, integration, and data reuse. Semantic Web languages have been developed to describe knowledge using a new W3C standard. These are RDF(S) (Resource Description Framework/Schema), OWL (Web Ontology Languages), and OIL, DAML+OIL. Most of these standards relied on Ontology for releasing the dream of the semantic web. Gruber [5] defines ontology as a specification of a shared conceptualization.

Ontology always includes a vocabulary of representational concept labels to describe a shared domain. These concept labels are usually called terms (lexical references) associated with entities. Ontology is one of the essential concepts used in the semantic web infrastructure. These days, researchers use the Semantic Web and Ontology to manage data in information retrieval systems. The Semantic Web simplifies and improves knowledge-intensive applications through ontology by addressing weaknesses in information retrieval, matching data, and data integration on the current Web. The semantic web aims to provide an extra machine-understandable layer, simplifying programming and maintenance efforts for knowledge-based web services.

*Corresponding Author.

This paper proposes an ontology-based retrieval system that transforms initial user queries using domain ontologies and applies semantic association during the indexing process. The main contributions of this paper are:

- Firstly, this paper proposes an ontology-based information retrieval system.
- Secondly, this paper implements the semantic approach in the indexing of documents.
- And finally, it also demonstrates how ontology-based search systems outperform the baseline.

The rest of the paper is organized as, in Section II, the authors present related work, Section III shows the proposed methodology, architecture, and algorithm. The experimental setup is given in Section IV. Section V presents the results and its discussion; we conclude in Section VI.

II. Related Work

By establishing successful approaches to overcome the difficulties in providing a more specific description of a user's information need, Ontology and Semantics took a giant step forward. Furthermore, it has also outperformed traditional IRS in terms of retrieval results. Researchers have adopted various techniques to transform query terms for performing retrieval. Some of them are given below.

Maxat Kulmanov et al. [6] employed fuzzy ontology in this study to aid with query transformation for an IRS. This procedure generates a dictionary of concepts from a given domain and an exterior ontology and then assigns fuzzy memberships via ConceptNet. They have used ConceptNet Global Ontology to determine fuzzy membership. By adding semantic weights to every one of the Concept-Net Ontology's semantic relations, the researchers created fuzzy membership for all these relations.

The researcher of this study [7] provides an overview of the methods for computing similarity using ontologies and incorporating them into machine learning techniques; they discuss how ontology embeddings and semantic similarity measures can leverage the background knowledge contained in ontologies, as well as how ontologies could provide restrictions that enhance machine learning methods.

The reference [8] established a mechanism for semantic query expansion based on domain ontologies. It expanded on synonyms, hypernyms, and similar words. Adding a similarity function to a system improved the quality of the formed query and the search engine results. Using a programming language domain enabled the system to be evaluated manually by impartial and experienced personnel.

Prilipsky et al. [9] proposed a hybrid Personal Knowledge Management (PKM) system to extract and store helpful information. The software-assisted PKM is the most effective one. Many software tools are used to retrieve concept mapping, tagging, flashcards, and hyperlinking. These tools are added to the traditional PKM to make it flexible and extendable. The advantage of the state-of-art work is a single solution using PKM instead of integrating all the functionalities from different tools.

Kim et al. [10] proposed an i-Dataquest prototype for a graph-based information retrieval system. I-Dataquest prototype contains three steps, data pre-processing, query pre-processing, relevance evaluation, and feedback. The graph data integrates the query with all syntactic and semantic extensions to retrieve the required answer. The PAINT'R dataset is selected, which is the same as the data of the manufacturing company.

Esposito et al. [11] proposed a hybrid query transformation approach. This approach is used in Information retrieval-based QA systems. It is based on lexical resources and word embeddings. In the query expansion for the question-answer system, the answer for the user-defined question is retrieved from an already formulated database having a particular required domain. This method modifies the natural language questions to enhance proper semantics to get the required sentences. The author used WordNet to produce appropriate nouns and verbs for the user questions. Then as per the sense of the question, the answer is retrieved using the Word2Vec model, which is created using a semantic similarity metric. The main drawback is that this approach uses only nouns and verbs for evaluating the query; it does not consider other syntactic categories, such as adjectives and adverbs.

III. Methodology

The problem with current IR systems is that the input queries are generally too short and too ambiguous to express the user's information needs. Such imprecise representation of users' information needs directly affects retrieval performance. In other words, it can be said that a simple query can't satisfy users' information needs.

A. Problem Formulation

Vocabulary mismatch between the query and documents. Let's consider a document collection with D relevant documents for a user query Q to understand this problem. For representing a single concept, Q and D may use different vocabulary. A traditional retrieval system performs only keyword matching between query and document. It does not detect similarity between Q and D. The authors address this problem using semantic and ontologies, which provide shared meaning for two different terms.

B. Proposed Information Retrieval System

To solve the above-stated IR problem, the researchers propose a novel ontology-based retrieving system as shown in Fig. 1. The steps of the proposed system are as follows:

1) The user's information need is specified by a user query (typically made of keywords) entered via the user interface.

2) The initial query is processed using domain ontology and query processing operations. Same operations are applied to document collection by semantically association module for Indexing purposes.

3) Index building from the document source is an offline process performed by the indexing module.

4) The transformed query is a semantic representation of user information needs.

5) This query is executed by searching using a semantics module to retrieve a set of relevant documents. Fast matching between query keywords and documents terms is done by the index structure.

6) The set of retrieved documents is then ranked according to the estimated relevance with respect to the term matching score.

7) The user then examines the set of ranked documents; he might point to a subset of the documents as useful and thus provide feedback to the system.

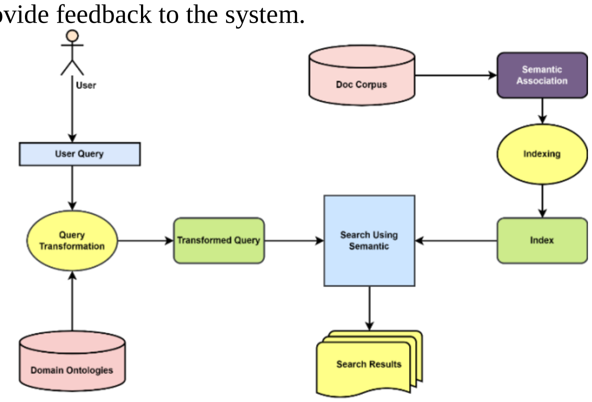


Fig. 1. An Ontology-based Semantic Information Retrieval System.

1) Query processing: To discover the documents that meet the user's information requirements, the query must first pass through a pre-processing module, which converts it into a more precise or machine-readable format. The matching module receives this type of question, known as a processed query. A typical search engine query consists of several terms. A list of terms with weights can also represent such a query. A standard information retrieval system will provide a high percentage of pages relevant to the user's query after matching. If the sender of the inquiry finds a page relevant to the topic for which the query was submitted, it is considered relevant.

2) Indexing: The IRS indexes millions of web pages containing a comparable number of distinct terms. Indexing can be defined as a process that collects, parse and store data to facilitate fast and accurate information retrieval. During indexing, a search engine records the words and phrases from downloaded pages; then, it prepares an index based on this data. It stores terms in an inverted file structure known as an inverted index. An inverted index stores the positions of text for each occurrence of a term. Another reason for indexing web pages by search engines is that it carried out processing like lexical analysis similar to the query processing phase, which improves the performance of search engines. In full-text indexing, virtually every word in the document is employed as an index term.

Indexing is an integral part of every search engine because it optimizes the query performance by improving the response times considerably. Along with Indexing, search engines also perform ranking, which is an attempt to see how good an approximation to "importance" can be obtained from just the link structure of the web. The index is built from text documents by the indexing module. Preparing an index is an offline process that parses text documents into tokens. Various text operations are performed on these tokens, transforming them into indexing terms.

3) Search using semantics: In this module, query keywords are matched with index terms. It retrieves those documents from the index that contain given query terms. It is typically a standard search for processed query terms in an index of documents. The degree of matching between a page and a query, called the similarity, can be measured by the number of terms they share. A similarity score between query and index terms is calculated to rank returned documents. A simple approach is to match query keywords with index terms and return the URLs of those documents that contain matching terms. This keyword or syntax matching between query and document can be improved with semantic matching.

The computational processing required for an NLP-based query having a probabilistic weighted model is more than an unweighted, Boolean matching model. Ranking scores all retrieved documents according to a relevance metric. One of the fundamental difficulties in information retrieval, the scientific/engineering subject that underpins search engines, is ranking query results. The task is to rank or sort the documents in D according to some criterion such that the "best" results appear first in the result list displayed to the user, given a query q and a collection D containing documents that fit the query. Traditionally, ranking criteria have been described in documents relevance to a query's expressed information demand. These graded documents are returned to the user using a user interface.

Algorithm 1: Pseudo Code for The Proposed OBS-IRS

Input	:	User Information need
Output	:	Retrieved relevant documents
		start
Step 1	:	Formulates the initial query *q* and submits it via the user's interface
Step 2	:	Query transformation module process *q* using domain ontology
Step 3	:	*q* is transformed to a more machine-readable form *q*'
Step 4	:	QT utilizes domain knowledge from ontology dataset
Step 5	:	Build Index *I*;
Step 6	:	Process documents from Doc Corpus using semantic association
Step 7	:	Apply Index preparing method Inverted Index on fetched pages
Step 8	:	for terms extracted from doc with semantic data
Step 9	:	Semantic Search
Step 10	:	Match *q*' with index terms
Step 11	:	Start for
Step 12	:	each term in *q*'
Step 13	:	if (*q*' term and index term have the same meaning)
Step 14	:	Retrieve document according to the matching function
Step 15	:	End for
Step 16	:	Apply Ranking on documents
Step 17	:	Return search results
		end

C. Query Transformation

The initial user query is not able to represent his information need properly. Our system uses three main query transformation techniques.

1) Query expansion: The IR use of Query Expansion (QE) has been the subject of research [12]. As seen in manual search, it can be seen that the user reformulates their search query. They do it because they didn't get the exact result of their original query. In the QE, the IRS improves the user's query by automatically expanding it. This can be done in several manners, like providing suggestions by guessing the user's intention according to the user's past behaviour.

This technique adds additional terms to the user's initial query based on local and global information resources analysis. This analysis focuses on finding semantically related terms to the original query. These target resources can be the whole document collection, the initially retrieved documents set, or documents from the computer. Expansion of queries [13] with matching terms improves performance in terms of recall. However, any method must find similar terms carefully during this process because it may lose gain in terms of precision.

2) Query refinement: Researcher use this refinement technique to improve the matching between queries and documents. The process of reflecting user needs with high accuracy is called query refinement [14]. In this process, the feedback by the user plays an important role. It generates a new query after applying the refinement process. Research on query refinement is not as dominant as query expansion.

3) Query suggestion: Query suggestion is also a part of the query transformation module. The most common form is spell checking during query processing by any search engine. The user is offered replacements to the initial query. These alternatives are more specific to the user's information need. Query suggestions provide more detailed descriptions of the search concept. This technique uses the extensive query history collected by web applications [15]. To implement query suggestion, the retrieval system generates a new query. It is different from the QE because it does not always add new terms to the initial query. Table I shows the query transformation approaches concerning their behavior on user feedback.

TABLE I. SUMMARY OF QUERY TRANSFORMATION METHODS

	Query expansion	Query refinement	Query suggestion
Generate new query		Yes	Yes
Expand original query	Yes		
User feedback before modification	Yes	Yes	
User feedback for final modification			Yes

IV. EXPERIMENTAL SETUP

For the experiment, the authors used Terrier [16] retrieval system. It is an open-source tool for IR experiments developed by Glasgow University, UK. Three standard datasets are also used for comparing the results of both approaches.

A. Datasets

The first dataset, ANTIQUE [17], is a non-factoid community question answering dataset. It is a collection of 2,626 open-domain non-factoid questions from a diverse set of categories. The Cranfield [18] is a small curated dataset extensively used in information retrieval experiments. There are 226 queries (search terms), 1400 documents, and 1837 (evaluations) in the dataset. The TREC MRT dataset [19] helps fill evaluation gap issues in IR. Due to the sensitive nature of medical records, data constraints are the overarching factor for Medical Records.

B. Evaluation Metrics

The authors used three standard evaluation parameters to evaluate our results. The detailed information about these measures is given below.

1) Precision (P): It is the fraction of retrieved documents that are relevant. It measures the quality of the results [20]. It is also known as positive predictive value. It can be calculated at different values, denoted as Precision at k (shortened as P@k). It can be calculated using Eq (1).

$$\text{Precision} = \frac{\#(\text{relevant items retrieved})}{\#(\text{retrieved items})} \tag{1}$$

2) Recall (R): It is the ratio of relevant documents that are retrieved. It can be seen as a measure of the quantity of documents corresponding to an information need [21]. It is also known as sensitivity. Being ratio, it has a value between 0 and 1. It can be calculated using both equations given below as Eq. (2).

$$Recall = \frac{\#(relevant\ items\ retrieved)}{\#(relevant\ items)} \tag{2}$$

3) Normalized Discounted Cumulative Gain (NDCG): To represent non-binary relevance, the use of cumulative gain or specifically normalized discounted cumulative has been increasing. It is a widely accepted evaluation parameter in the IR community [22]. It can also be calculated at a given rank cutoff (e.g. ndcg_cut_10). It can be calculated for k, top search results. For query, j form a set of queries Q, consider R(j,d) as relevance score is given to document d, it is mathematically calculated by Eq. (3).

$$NDCG(Q,k) = \frac{1}{|Q|}\sum_{j=1}^{|Q|} Z_{kj} \sum_{m=1}^{k} \frac{2^{R(j,m)}-1}{log_2(1+m)} \tag{3}$$

V. RESULT AND DISCUSSION

To show the performance of the proposed and baseline system, the authors evaluated both on three datasets. The parameter used to judge the performance are precision, recall, and NDCG. The results were taken at different levels such as 10, 20, 50, and 100 top documents. The detailed discussion for each parameter is given in the following subsection.

A. Precision Analysis

Fig. 2 shows results for the precision on the ANTIQUE dataset. Here you can see that the proposed combination of semantics and ontology outperformed the baseline results. For p@10, the value for the ANTIQUE dataset was 0.2356 in the baseline model, whereas it was 0.2452 for the proposed system. For P@20, values of baseline and proposed were 0.2322 and 0.2396. The value of @50 was 0.2145 and 0.2292. At the 100 documents, the precision values of baseline and proposed were 0.2012 and 0.2086. This shows that the proposed approach performed better for all values of precision. Fig. 3 shows results for precision on the Cranfield dataset. For this dataset also, the proposed retrieval outperformed bassline retrieval. Similarly, in Fig. 4, you can see the precision results for the TREC MRT dataset; our system was also better in terms of P@10, P@20, P@30, and P@100 results.

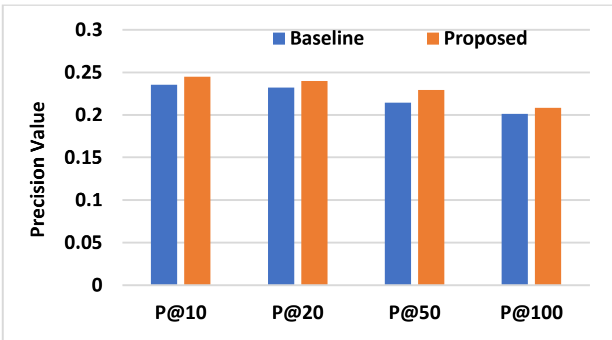


Fig. 2. Precision Comparison between Baseline and Proposed on ANTIQUE.

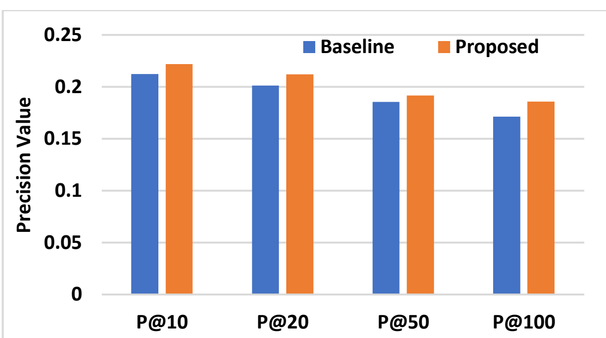


Fig. 3. Precision Comparison for Baseline and Proposed System on Cranfield.

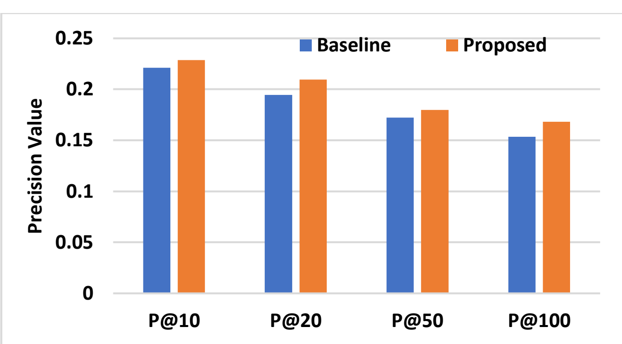


Fig. 4. Precision Comparison for Baseline and Proposed System on TREC MRT.

B. Recall Analysis

The results for recall measures on the ANTIQUE dataset are shown in Fig. 5. Here you can see that the proposed system outperformed the baseline. For R@10, the baseline achieved a value of 0.3234, and the proposed system achieved 0.4321. The values for baseline and proposed at R@20 were 0.3745 and 0.4023. The best value of R@50 is 0.4421 given by the proposed approach. For R@100, baseline achieved 0.5122, but our semantics method gave 0.5622. In Fig. 6 comparison between baseline and proposed on Cranfield is presented. The graphical representation in Fig. 7 shows that the proposed methods have beaten the baseline for TREC MRT results.

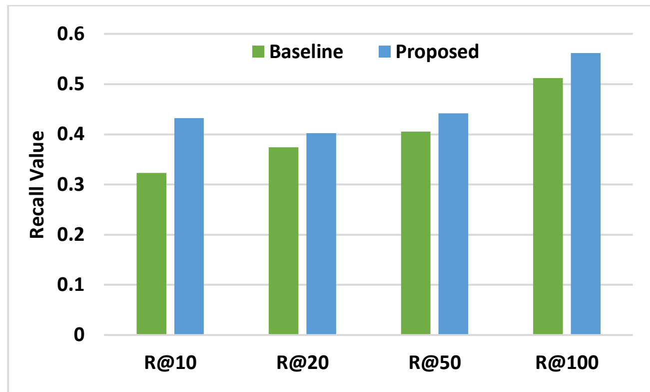


Fig. 5. Recall Comparison between Baseline and Proposed on ANTIQUE.

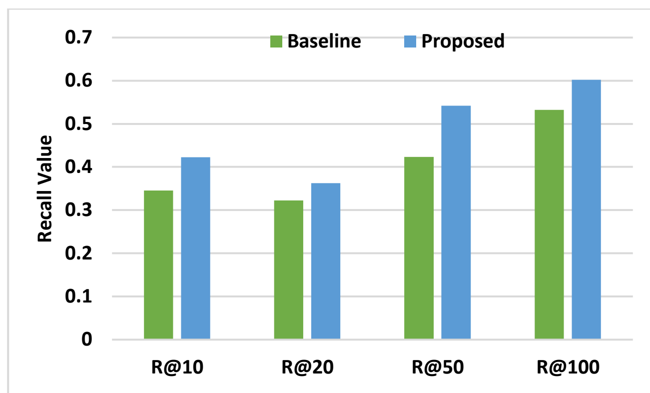


Fig. 6. Recall Comparison between Baseline and Proposed on Cranfield.

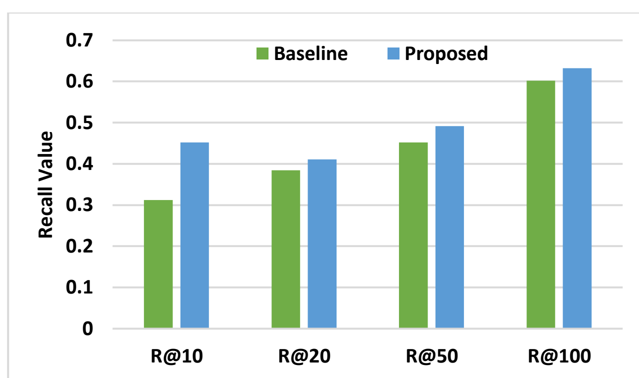


Fig. 7. Recall Comparison between Baseline and Proposed on TREC MRT.

C. NDCG Analysis

From Fig. 8, you can see NDCG results for the ANTIQUE dataset, for cut10 proposed approach gave 0.2168, whereas for baseline value was 0.2101. In comparing cut20 results, the authors found that the difference between baseline and proposed is less than 0.0028. For cut50, the value for the ANTIQUE dataset was 0.2045 in the baseline model, whereas in the proposed system corresponding value was 0.2128. For cut100 value for the proposed was 0.2162 higher than the baseline. Fig. 9 and Fig. 10 show NDCG results for the Cranfield and TREC MRT dataset, respectively; from these comparisons, the authors found that the ontology method outperformed the baseline.

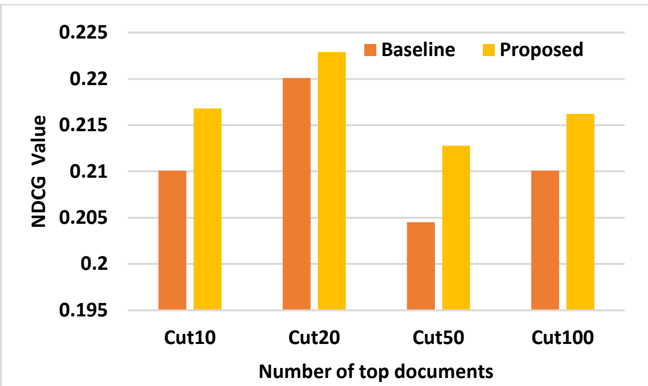


Fig. 8. NDCG Comparison between Baseline and Proposed on ANTIQUE.

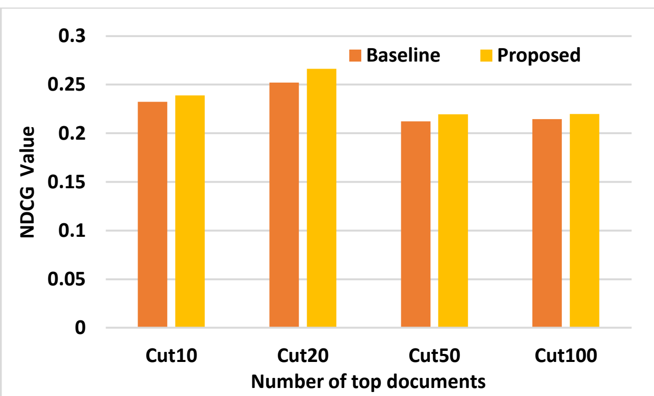


Fig. 9. NDCG Comparison between Baseline and Proposed on Cranfield.

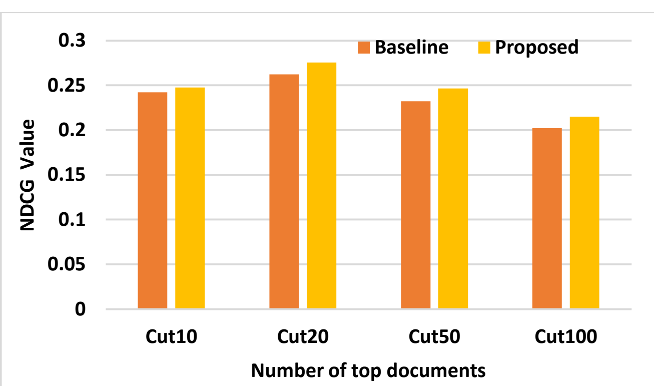


Fig. 10. NDCG Comparison between Baseline and Proposed on TREC MRT.

VI. Conclusion

The results of ontology-based query transformation have shown that if any IRS uses domain ontologies for query processing, its retrieval performance improves. Similarly, semantic association in the indexing process helped to capture the related terms from documents. Hence both modules of the proposed system abetted to retrieve relevant documents as per the user's information need. The different comparisons between the proposed and the baseline results proved that the author's hypothesis is correct. The use of ontology and semantics techniques delivered better results on all three datasets. The proposed system achieved 4% high precision at 100 top documents for ANTIQUE dataset. The recall values on Cranfield dataset at top documents increases by 13%. The NDCG parameters values at top 100 retrieved documents is 6% higher than baseline results. So, it is undeniably an efficient system for retrieving relevant documents from the extensive collection of unstructured data.

The authors conclude that query transformation using ontology has a high impact on retrieval performance. Use of semantic matching and domain-specific knowledge helped IR users to find documents that satisfy their information need. The authors hope that commercial search engines will utilize the full benefit of domain ontology and semantics in near future.

References

[1] M.J.H. Mughal, "Data mining: Web data mining techniques, tools, and algorithms: An overview," Information Retrieval, 9(6), 2018.

[2] R. Kumar and S. C. Sharma, "Information retrieval system: An overview, issues, and challenges," International Journal of Technology Diffusion (IJTD), 9(1), pp. 1–10, 2018.

[3] F. Liang, C. Qian, W. G. Hatcher, and W. Yu, "Search engine for the internet of things: Lessons from web search, vision, and opportunities," IEEE Access, 7, 104673-104691, 2019.

[4] Tim Berners-Lee, James Hendler, and Ora Lassila. "The semantic web," Scientific American, Issue 284, Vol. no. 5, pp. 34-43, 2001.

[5] Thomas R. Gruber, "A translation approach to portable ontology specifications," Knowledge Acquisition, 5(2), pp. 199-220, 1993.

[6] Maxat Kulmanov, Fatima Zohra Smaili, Xin Gao, and Robert Hoehndorf, "Semantic similarity and machine learning with ontologies," Briefings in bioinformatics, 22(4), pp. bbaa199, 2021.

[7] Shivani Jain, K. R. Seeja, and Rajni Jindal. "A fuzzy ontology framework in information retrieval using semantic query expansion." International Journal of Information Management Data Insights, 1(1) pp. 10009, 2021.

[8] J. Singh, M. Prasad, K. O., E. M. Joo, A.K. Saxena, and C.T. Lin, "A novel fuzzy logic model for pseudo-relevance feedback-based query expansion." International Journal of Fuzzy Systems, 18(6), pp. 980–989, 2016.

[9] R.E. Prilipsky and M.A. Zaeva, "A Hybrid System for building a Personal Knowledge Base," Procedia Computer Science, Vol. no.169, pp.96-99. 2020.

[10] L. Kim, E. Yahia, F. Segonds, P. Véron, and A. Mallet, "i-Dataquest: A heterogeneous information retrieval tool using data graph for the manufacturing industry," Computers in Industry, Vol. no.132, pp.103527, 2021.

[11] M. Esposito, E. Damiano, A. Minutolo, G. De Pietro, and H. Fujita, "Hybrid query expansion using lexical resources and word embeddings for sentence retrieval in question answering," Information Sciences, Vol. no. 514, pp.88-105, 2020.

[12] Claudio Carpineto, and Romano Giovanni, "A survey of automatic query expansion in information retrieval." ACM Computing Surveys (CSUR), 44(1) pp. 1-50, 2012.

[13] D. K. Sharma, R. Pamula, D. S. Chauhan, "A contemporary combined approach for query expansion," Multimedia Tools and Applications, 2020.

[14] H. Scells, G. Zuccon, and B. Koopman, "Automatic Boolean query refinement for systematic review literature search," In The World Wide Web conference, pp. 1646-1656, 2019.

[15] Makoto P. Kato, Tetsuya Sakai, and Katsumi Tanaka, "When do people use query suggestion? A query suggestion log analysis," Information Retrieval, 16(6), pp. 725-746, 2013.

[16] C. Macdonald, R. McCreadie, R.L. Santos, I. Ounis, "From puppy to maturity: Experiences in eveloping terrier," Proc. of OSIR at SIGIR, pp.60– 63, 2012.

[17] Helia Hashemi, Mohammad Aliannejadi, Hamed Zamani, and W. Bruce Croft, "ANTIQUE: A non-factoid question answering benchmark," In European Conference on Information Retrieval, pp. 166-173. 2020.

[18] Cyril Cleverdon, "The Cranfield tests on index language devices," In Readings in information retrieval, pp. 47-59. 1997.

[19] Nicola Ferro and Peters Carol, "Information Retrieval Evaluation in a Changing World: Lessons Learned from 20 Years of CLEF." Vol. 41. Springer, 2019.

[20] L. Azzopardi, P. Thomas, and N. Craswell, "Measuring the utility of search engine result pages: an information foraging based measure," In The 41st international ACM SIGIR Conference on research & development in information retrieval, pp. 605-614. 2018.

[21] N. Thakur, N. Reimers, A. Rücklé, A. Srivastava, and I. Gurevych, "BEIR: A heterogenous benchmark for zero-shot evaluation of information retrieval models." arXiv preprint arXiv:2104.08663. 2021.

[22] T. Sakai, "On the instability of diminishing return IR measures," In European Conference on Information Retrieval, pp. 572-586, Springer, Cham. 2021.

Characters Segmentation from Arabic Handwritten Document Images: Hybrid Approach

Omar Ali Boraik[1], M. Ravikumar[2]
Department of Computer Science
Kuvempu University
Shankaraghatta, Shimoga 577451
Karnataka, INDIA

Mufeed Ahmed Naji Saif[3]
Department of Computer Applications
Sri Jayachamarajendra College of Engineering
Visvesvaraya Technological University (VTU)
Mysore, Karnataka, INDIA

***Abstract*—Character segmentation in Unconstrained Arabic handwriting is a complex and challenging task due to the overlapping and touching of words or letters. Such issues have not been widely investigated in the literature. Addressing these issues in the segmentation stage reduces errors in the segmentation process, which plays a significant role in enhancing the accuracy of the Arabic optical character recognition. Therefore, this paper proposes a hybrid approach to improve the accuracy for interconnection, overlapping or touching character segmentation. The proposed method includes several stages: removing extra shapes such as signatures from the document. Using morphological operations, connected components and bounding box detection, detect and extract individual words directly from the document. Finally, the touching characters segmentation is achieved based on background thinning and computational analysis of the word's region. The proposed method has been tested on KHATT, IFN/ENIT database and our own collected dataset. The experimental results showed that the proposed method obtained high performance and improved the accuracy compared to other methods.**

Keywords—Arabic handwritten character recognition; connected components; word segmentation; character segmentation; morphological operators; overlapping and touching characters

I. Introduction

Recently, electronic devices and modern technology have become an important and essential in human daily life. A lot of efforts and time were spent to protect and maintain valuable historical documents, letters, and books into digital images for scientific, service, and future uses. Optical recognition systems appeared as significant tools to avoid the loss of such valuable documents which convert text images into editable digital texts. There are various uses of machine learning techniques in Optical Character Recognition (OCR) systems; the gap is still largely due to unlimited obstacles in Arabic handwriting. Comparing OCR to Latin and other languages with the recognition of Arabic, the Arabic recognition system is still incomplete and unsatisfactory. In today's world, interested parties in the field of documenting are required to save digital images possible to modify, i.e., repairing deterioration in historical books, correcting errors, using a text part of it in other applications.

Converting text images into editable digital forms is called Optical Character Segmentation (OCR) [1]. The text images are either printed or handwritten. The deficiencies of Arabic handwriting OCR systems are more complex than printed and incomplete. Because handwriting does not abide by the font's criteria, specific size, different font and size style writing for a word repeated several times in the same document itself, other problems of interconnection, overlapping and touching, and difference gaps among word/sub-words increase the complexity of Arabic handwriting. There are common factors that make Arabic handwritten as well as printed text complex such as Arabic nature cursive, writing from right to left, connecting the letters, and so on.

Arabic handwritten character recognition has the same situation as other languages. In some cases, it seems to be more complex depending on the language, the challenges it faces, and difficulties for line, words [2]-[4] and the challenges in character segmentation of the input document images. These challenges, such as interconnection letters word, cursive overlapping, touching existence of ligatures, diacritical marks and the position and number of dots above or under some letters. These challenges may lead to misclassification and unsatisfactory results. However, some of these challenges were recognized by many researchers in the OCR field using machine learning techniques [2].

The deficiencies are based on the failure of the Arabic character segmentation stage. Cursive, overlapping, and unrestricted writing challenges are the most long-term barriers to correct segmentation. A study [1] presented a projection profile technique for Arabic characters segmentation; which was tested and evaluated successfully in Arabic database with various sizes, styles, and font types. But it is limited to the Arabic printed document. The proposed system fails to deal with handwriting documents.

Another study [4] presented a solution for overlapping and touching Arabic characters segmentation by overlapping set theory and contour tracing. It is low accurate when segmenting the multiple touching letters. While [6] suggested a hybrid method focusing on the middle point of the word area. This study was focused on handwriting documents in the Hindi language. This method succeeded in fragmenting the multi-touching letters, and to apply it to the Arabic language; it needs to be developed because the Arabic calligraphy starts from right to left.

The segmentation solutions for single/ multiple touching, overlapping challenges, and interconnected characters problems still need to be expanded to include more Arabic handwritten documents. Achieving progress in respect to Arabic text recognition is hindered by such obstacles and challenges. Moreover, the complexity of finding a solution to the segmentation of overlapping and touching Arabic handwritten words or characters made few researchers interested in addressing these two problems and developing techniques to address these complex problems. These two challenges created gaps in attempts to process them, i.e., there is a gap of at least two years between every two research works (Ouwayed et al., 2009, Belaïd & Ouwayed, 2011, Aouadi, N. et al., 2013, Aouadi & Kacem, 2017, Ullah et al., 2019, AbdAllah, N., & Viriri, S. 2021, Ahmed et al., 2022 as a review survey). These are some of the motivations that encouraged us to propose a hybrid approach to reduce these complexities and improve segmentation techniques.

This article proposed hybrid approach to segment the Arabic handwritten document into direct words with an objective to enhance and increase the accuracy ratio of Arabic handwritten character segmentation while dealing with overlapping, interconnected, and touching Arabic handwritten documents. The work was divided into eliminating non-textual appendages in documents, segmentation, and extracting words from the image using connected components and thinning techniques. A hybrid method is used with computation analysis of the word's region to segment a word into characters. The middle point is detected to extract the structure features for dealing with the input, which contains the touching overlapping character and distinguishes from isolated character based on calculating the vacant space index value. The hybrid method has been proven to be a flexible and efficient method to deal with various renewed database. The contributions of this study are listed below:

1) Proposing a hybrid approach for character segmentation in Arabic handwriting that addresses the challenges of segmentation of touching characters eliminates the signatures from input document images and segments words directly from the input images, including wavy lines.

2) Creating a new database for Arabic handwriting to evaluate the proposed hybrid approach.

This article is organized as follows: Section 2 reviews the existing studies related to Arabic characters, Section 3 describes the methodology for character segmentation, Section 4 discusses the experimental results, and Section 5 concludes the study.

II. Related Work

The challenges of touching and overlapping lead the researchers to focus and work on line/ word/ character segmentation. The OCR systems depend on hybrid techniques, which are considered better than other systems in accuracy, speed, and flexibility in dealing with renewable databases. It is appropriate for dealing with handwritten documents as indicated in [7]. The researchers suggested a method in which they used a technique based on connected components. Then, it selects the estimating and the alignment transformation of these connected components, stored in segmented models by templates for two most similar Connected Components (CCs) of the touching words. Finally, it separated the CCs into two regions using the Centre Points of these regions.

The authors [4] introduced a solution for touching Arabic characters by using a hybrid approach for character segmentation. The proposed method to split through letters is by selecting the touching point by overlap set of theory and endpoints of Arabic word and then tracing the boundaries of the touching letters to segment them individually. The segmentation is good, just dealing with single touching characters. Due to the differences in positions of words in different documents, the hybrid methods need to be improved and developed to give better results in the future. The authors [8] suggested a method for touching Arabic characters segmentation based on template segmentation. This method creates a dictionary file. This file contains a template for each touchpoint with its necessary details. Then it is compared with the input images to identify a sample from the dictionary file. As mentioned earlier, the template method is difficult to apply in handwritten character segmentation. It saps computer hardware to save each touchpoint in a template for detecting similar touching points. The runtime is very long and tedious. Fatal flaws in the handwritten character segmentation.

In the study [9], the authors used a segmentation method that extracts topographic features. These features identify potential segmentation points of the characters block connected. The segment of the possible points is based on the average width of the character. The study achieved approximately 70% of character segmentation. Moreover, many errors in segmenting, handwritten Arabic characters, mainly the letters which connect from two sides. While in [10] suggested a method to segment touching handwritten Arabic characters. It first detects the intersection points and the beginning ligature's pixel near the upper line from the baseline. The process starts from this point (ligature's pixel) to pursue the descending character to the intersection point and respect a different angular corresponding to the descending character curvature. The proposed method was tested on 100 Arabic document images containing 256 touching lines. However, the success ratio reached 94% of the segmentation. But it is inappropriate for the segmentation of multiple touching components. Three methods from the literature were compiled and already developed. So, it is a mixture method. The hybrid method is suggested by [11] to segment the touched printed and handwritten Latin characters (obviously, the success of segmentation for handwritten characters. will be the highest efficiency and accuracy in printed characters). The disadvantage of this method is that it was tested on a few images of documents. Therefore, the success rate is insufficient and unsatisfactory.

The problems of overlapping, interconnection, and touching in handwritten Arabic characters make the segmentation process complicated task. From the aforesaid literature review, it was noted that these issues had been extremely slightly discussed in previous studies. In addition, period intervals between those studies were absolutely longer. However, there are shortcomings in solving these problems, which involves limitation in the studies related to Arabic databases and lack in performance of segmentation of the multiple-touching Arabic characters. The proposed methods are suitable for the segmentation process but are inflexible in dealing with morphological differences. Therefore, this study aims at developing a hybrid method to address the above-mentioned problems.

III. Proposed Method

This section describes the methodology of the proposed work as shown in Fig. 1. Initially, the pre-processing is performed for the input image to improve low quality and prepare for the next stage. Then, it segments the whole document into individual words, especially if the input image contains wavy lines. If the input image includes shapes or signatures, the approximate polygon methods remove these shapes. In contrast, the signatures are extracted and removed from the document based on the connected component analysis. Finally, divide the words into isolated letters.

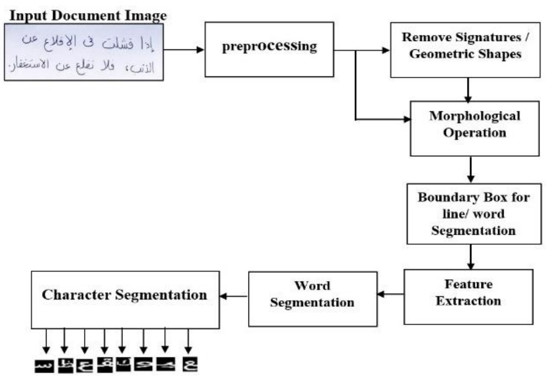


Fig. 1. Block Diagram of the Proposed Method.

A. Pre-processing

In this work, a new database is created which contains old and modern handwritten documents. The quality of the collected images is good, medium, and many of them lack lighting. The problem of poor and unstable lighting in the whole document is big obstacle to make the pre-processing stage complicated as shown in Fig. 2, where the different spots of lighting in same (one) document, which lead to a lack in performing the segmentation process. Using some filters such as Gaussian Filter, Median Filter, Low pass Filter, Custom Filter, etc. which give enhancement in these images, but the results are poor because the most challenge was the unstable lighting in the image documents in which one image has different contrast lighting. The results, on the other hand, were good when the input images with stable lighting were applied which is similar to applying IFN/ENIT or KHTT database, because these databases have Binarization images.

One of the challenges faced when colleting the images, some of the images have poor quality and resolution as shown in Fig. 2. The image (a) is darker than others. Its histogram in (b) placed at 0 - 200 to represent (the) pixel value in x-axis takes the image (a) pixel value and the value placed at 0 to 200 that indicated the values grouped at black values, this represents the image (a) consisting of more black pixels compared to other gray level value. The image (c) is a low contrast image, its histogram (d) is placed at almost the center towards white. The image (e) is a bright image where its histogram (f) values are placed at 170 towards white 255, which indicated the image (e) has more white pixels. Although the preprocessing approaches were applied to these images the results were not good.

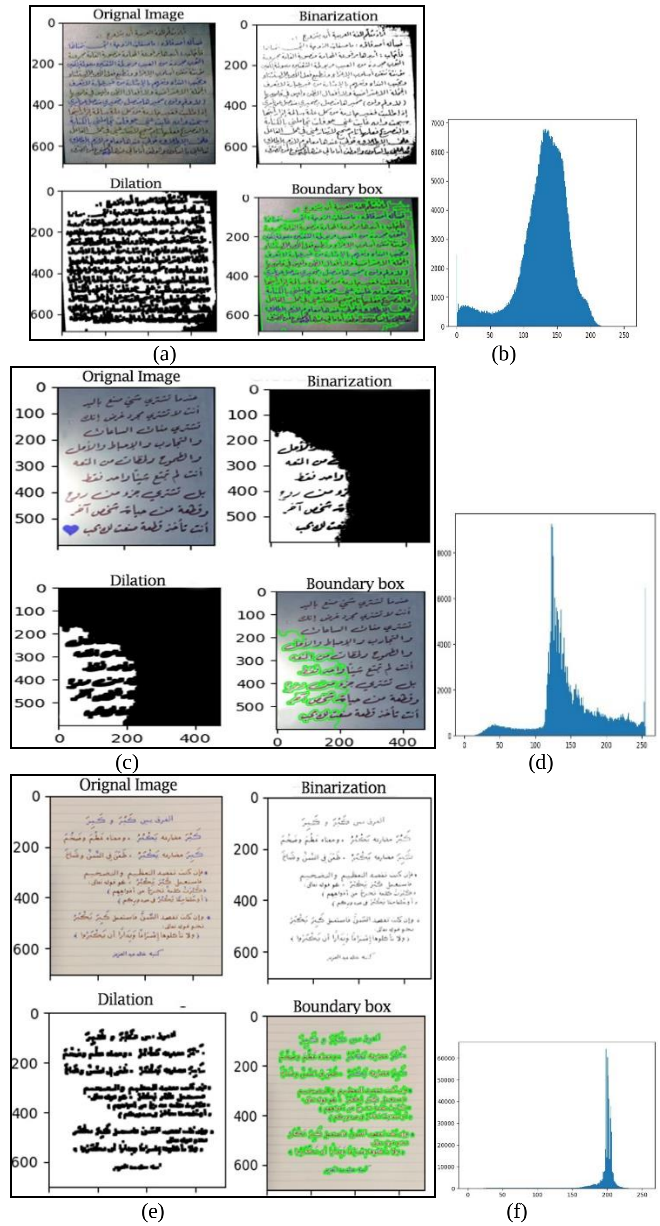


Fig. 2. Different Lighting and Histogram of Intensity Distribution. (a), (c) and (e) are the Input Document Images, While (b), (d) and (f) are the Original Images' Intensity Distribution Histogram for showing the Intuition about the Contrast Brightness Intensity Distribution.

B. *Removing the Signatures*

Many images of documents contain signatures, so before shifting to the next stages of Arabic OCR stages, it should eliminate these signatures to help to word and character segmentation. The method is used to remove these signatures depending on Connected Components and find out the thresholding average as shown in Fig. 2. By using image processing, the regions of connected pixels were recognized by this algorithm. It commonly gives the same result. In other words, the given input image is scanned by these connected components along with this attached signature. The next step is gathering the pixels into new components connectivity i.e. the elements of the image connected to the same intensity values of each pixel and showing the link with other values. Therefore, the components will be recognized, and every pixel will be highlighted with a specific colour (color labelling). Each pixel may highlight with a grayscale according to the located component.

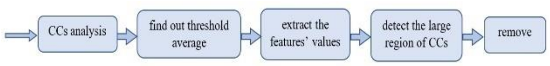


Fig. 3. Block Diagram for removing the Signatures.

Now-a-days, classification of each connected component along with the assorted dissociates are essential to many analysis applications of image's machine-driven. In this process, the whole image, from left to right and top to bottom, is scanned to recognize the region of each pixel which is connected to the image's component. In other words, it can be said the adjacent pixels of each component share constant value V. it can be applied to binary or grayscale images, and it measures connectivity differently. Before applying the mask technique move over, the input image should be a binary, 8-connectivity where a mask created and each pixel and its surroundings is are checked using this 8-connectivity. The operator moves over the image to scan rows individually until it arrives at ‘p’ point, It examines the remaining eight neighbours of the labelled pixel (at any stage, ‘p’ is the labelled pixel) for which V = [1]. Then it examines the four neighboured labelled pixels from right to left and from the upper diagonal direction, which were already encountered in the scan. According to previous details of the scan, the term ‘p’ is classified when the process finds an adjacent value equal to ‘1’. Then the label is assigned to ‘p’. At the same time, if other neighbours have the same value, all of the labels will be assigned. The equivalencies will be noted; if all pixels are 0, a new label will be given to ‘p’. This process is followed by the initial scan of the label pair area units and sorted into equivalent categories (classes and distinct labels). The next step is the second scan of the image in which every label is replaced by the equivalent category, even though the labels may not be identical.

A Scikit image library provides an exciting feature to identify and label the connected components. This library is used to check the scanned input image documents and find these connected components with their labels in addition to grey and color labels. It turns out that the largest connected component is the signature compared with word components. Therefore, if it is possible to extract the largest component of the whole document, the signature can be recognized. However, using large connected components can extract unwanted words, lines, or other shapes. Therefore, a threshold value is used to solve this problem, it is used to detect outliers, i.e., any lines, structures, and texts that do not belong to signatures are calculated after a series of experiments which have been performed. In terms of a mathematical formula (1), which is obtained based on experimental results. It gives quite effective recognition of signature's regions in dealing with most A4 size scanned documents. Table I shows the obtained values representing the signature region's characteristics. The place of the signature is determined based on the biggest of the connected component's value and then compared with the average of this region to extract the signature to delete. Fig. 4 shows the identification of the signature's region from a sample set of images.

$$\mathrm{const\_A4} = \left[100 + \left(250 * \left(\mathrm{average}/84\right)\right)\right] \tag{1}$$

TABLE I. FEATURES OF SIGNATURE’ REGIONS

Sample Images	Biggest Components	Average	Small Components	Big Components
A	924	162.508772	583.657059	10505.82707
B	7682	793.734940	2462.30634	44321.51463
C	1158	128.257353	481.718312	8670.929621
D	1029	38.2051282	213.705739	3846.703298
E	1156	77.5617021	330.838392	5955.091185
F	5888	123.899471	468.748425	8437.471655
G	566	85.7945946	355.341055	6396.138996

Fig. 4. Extraction of Signatures from Sample of Images.

C. *Line Segmentation*

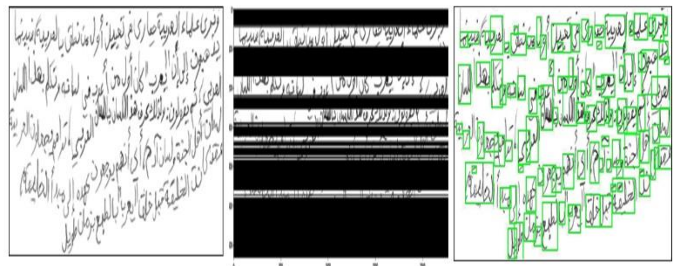
Separating the input images into isolated lines is successful with high accuracy in documents which may contain printed, historical or handwritten texts, these documents' lines have gaps between them that are orderly spaced [12], and the handwriting is on lined sheets. These gaps are ordered and these gaps may be almost empty of noises or unimportant objects. Because the feature of the gaps between the lines is the most important for the success of lines segmentation and extraction individually for using another application, in skew correction with multiple slanting or copy particular lines to other programs. Many handwritten documents lack clear orderly gaps between the lines, which are close, touching, and sometimes wavy. These gaps may be spaced at one side and touching or overlapped at the other. These challenges make the line segmentation process difficult and complex. Many studies have proposed techniques to solve these challenges and achieve satisfying success despite the convergence, contact and overlap between the lines. A study [13] proposed an A* path planning algorithm to line segmentation directly. The localization of each line is detected by two steps: Binarization method and projection profile analysis. The author's experiments tested handwritten historical documents from the MONK and the Saint Gall dataset. The study [14] presented a method of line segmentation based oriented anisotropic Gaussian Kernel, in which the authors divided the input text image into connected components achieved by Boundary Boxes. The method was tested on a sample of English Handwritten documents. In [15], the authors used midpoint detection technique for every two lines or words. The technique was tested on Gurmukhi printed and handwritten (Handwritten scripts are orderly distances among the lines) scripts (Punjabi). Their results achieved 95% of success. Study [5] also used a technique and a database containing English historical documents. The line regions were detected using rough detection throughout black/white transition map which is used to extract the lines through corresponding lines axes and skeletonization. PROHIST database was used in this study and achieved 82.18% of accuracy.

After it converted the input image into binarization in preprocessing phase, the input image locates the text (object) areas of the white background where the text is black. The process Uses Dilation operations to make the pixels' value of each line. This pixels' value has a single value so the line is considered one-component [16]. The starting and ending four points of each component determine the bounding box location among these points around the line component [17]. This boarded rectangle is segmented for each line, returned the segmented line to its original size using the erosion operation and the thinning method to give a better distance between words in preparation for their segmentation. Then the matrix of the rectangle border is saved into an image.

The overlapping and touching challenges between lines, make the dilation operation more extended, two or more lines are considered one component [18]. These challenges are solved by tracing the contour points of the touching area horizontally and overlapping vertically, using the method (direction contour tracing) in [18]. It traces the overlap path with calculation operations. This method is not good in separating lines in Arabic Handwritten documents because of the stretching of strokes and some letters are written vertically and extendly, such as these letters ‘أ‘,‘ک‘,‘ظ‘,‘ط‘,‘م‘. Also, most line segmentation techniques fail to segment wavy lines.

D. Word Segmentation

As it was mentioned earlier, many of the input images contain lines skewness, touching, overlapping, wavy and more closely spaced lines, where the line extraction fails, so the lines are separated into two or more lines as shown in Fig. 4. In the study [13], Hidden Markov models were used for word segmentation from the entire document directly. Current study, the words or sub-words are segmented by applying the Connected Components method for overcoming and solving many challenges such as the distance between words is less than that between sub-words and overlapping in one word. The Boundary Box function is used to extract them. These two methods achieved a higher success rate and are better than dividing the lines before the words, as shown in Table III.



(a) Original Image. (b) Bad line segmentation. (c) Words segmentation.

Fig. 5. Segment the Document Image into Words.

Fig. 5(a) shows a sample of a document image which contains wavy and touching lines. Many methods were applied to segment the lines in such image. The line segmentation methods such as projection profile method, a method based on tracing, another method based on contours, third method based on baseline, forth method based on morphological operation or other methods. These methods did not succeed in extracting the lines as shown in Fig. 4(b), where the first line only was successfully extracted. The rest of the lines were considered one component, or extracted a part of a particular line with previous or next line in the same process. Fig. 4(c) shows the words were extracted directly from the document with better accuracy and success.

E. Character Segmentation

This step is crucial to the segmentation stage. After overcoming the previously mentioned, challenges as much as possible, such as different light spots in the input image and existing shapes or signatures. After the success of words/sub-words segmentation, the character segmentation follows.

First, detecting the touching of characters: The Connected Components method is used to solve the overlapping problem between characters. It is also used to measure the weight of the character. A threshold value has been fixed to evaluate the weight. If the obtained value is less than the threshold, it can be considered a single character and split automatically. If the value is greater, it can be regarded as a touching character using the variable:

Tc (2) is the aspect ratio of the touching character is greater than the character is automatically split. In order to determine the touching characters using the variable (Tc), this aspect needs to be improved due to the similarity between interconnected and touching of Arabic characters. The touching characters are defined by.

$$T_c = \frac{e^g}{1+e^g} \qquad (2)$$

Where $g = \frac{w}{h}$, w is width of the character, and h is height. After identifying the touching character, it is classified as either horizontal, vertical or multiple touching.

By comparing the two values of height and width, the type of touching is determined by $g = h1 > w1$ the touching is vertical. $g = w1 > h1$ the touching is horizontal. While the multiple touching is defined as $g = w2h2 > w1h1$.

Second, using hybrid approach by following these steps:

1) Find the area of a word/ sub-word throughout detect the start point and endpoint of this area, allocate the area on width (w) and height (h). (w, h = contour area(word)).

2) Using thinning, closing and opening operation

3) Using midpoint steps to separate word/ sub-word isolated.

- The vacant space index value is calculated on the word/ sub-word's constrained such as width and height.
- Check the previous pixel and the next one (i+1, i+2, i-1, i-2) to save the column values and check if broken character appears.

4) Saved the vacant space between the characters in array_index.

- Detect the centre value of every vacant space between the characters until the end of the word.

M_point = (start index + last index)/2.

- This mid-point as a centre value is considered as the detection points to split the isolated character.

5) To determine the existence of joining (touching) characters, the total number of characters in a word is calculated.

6) Total value of characters is detected by the ratio of width and height.

Total = $\frac{width}{height}$

7) Total value of characters is compared with the segmentation point.

Value of M_point = total + 1.

8) If joining character is exist then number of segmentation points does not exceed the total no of characters in a single word. Otherwise go to f.

9) f- Calculate the distance sequentially between the middle values, if the distance above 110% of height, there must be a joining character present, which could be single or multiple joining characters.

10) Using clustering method to find the cluster in identified area of importance of the character in the middle part.

11) Discover the region of importance cluster between M_point1 + 10-(M_point2 +10) to obtain the heap of pixel.

- Scan every column to determine the cluster, if pixel calculate is found to be 10 then it is considered as joining point of the character.
- By leaving three columns in a row, you can segment the joining character.
- The new segmentation points should be extracted.
- Split the word from all the segmentation points.
- Show the results as in Fig. 6.

Fig. 6. Sample Images Show the Character Segmentation Isolated, Touching and Joined Characters.

IV. Evaluation Metrics

Evaluate the performance of the proposed hybrid approach for character segmentation. The performance evaluation metrics used by [19] are followed; the same evaluation strategy was also applied in this work, which uses five factors for evaluation: successful segmentation rate (SR), precision (P), recall (Re), correct segmentation rate (CS), and F-measure (Fm). The authors are illustrated in Equations 3-7. These factors are figured out by counting the number of matches between the resultant segmented words and then characters by the algorithm and ground truth characters in text word segments.

$$SR = \frac{NCc}{NCr} * 100 \qquad (3)$$

$$P = \frac{(NCc+NCo)}{NCr} \quad (4)$$

$$Re = \frac{(NCc+NCo)}{NCg} \quad (5)$$

$$CS = \frac{NCr-(NCi+NCo)}{NCg} \quad (6)$$

$$F_m = \frac{(Re*P)}{(Re+P)} * 2 \quad (7)$$

Where NCg: Number of ground truth words, characters respectively.

NCc: Number of segment correct words, character isolated.

NCr: Number of segmentation results (words, characters).

NCi: Number of incorrect segmentations for words, characters.

NCo: Number of over segmentation of touching in words, characters.

V. Experimental Result and Discussion

This section describes the implementation and evaluation of the proposed method. The proposed method was implemented using Python 3.8.8, Open CV environment (Spyder4[MSC V.1916 64 bits]), Win 11 pro 64_bit OS, with Intel(R) Core (TM) i5-9300HF CPU 2.40 GHz, RAM 8 GB.

The proposed method is tested on three Arabic databases. The images of the first database that have been collected were 2,300 handwritten text images, 20156 lines, 302,348 words; the ground-truth value of lines and words were calculated manually and estimated. The collected images are obtained from a scanner and also through social media. These images contain shapes, graphics and signatures. Some of these images are poor in quality and lighting. In such cases, the pre-processing methods such as filters, binarization morphological operation, normalization, skew correction were applied for image purification. The shapes were removed based on the size of the shape region while the signatures were removed from the document images. 140 images contain signatures as shown in Fig. 3. Removing the signatures process was successful in 122 images, the process failed in 7 images, over-segmentation of signatures was noticed in 11 images. However, features of any word's region are similar to features of signature's region results in being removed as shown in earlier Fig. 3.

The proposed method failed in 170 images with a failure rate of 7% due to the poor quality of the images sent through social media. 15,721 lines were segmented correctly. Table IV shows the results of line, word, sub-words or characters segmentation. Because lines are touched, closed, wavy, or slant, the words, and sub-words were directly extracted from the input image. Arabic word's letters may be interconnected overlapped, so the segmentation process of a word depends on the gaps between every two words or sub-word. Table II and Fig. 6 show the experiment results.

KHATT database contains 1000 images of Arabic handwritten forms written by 1000 writers from different countries, 9327 lines, 165890 words [20]. And IFN/ENIT database involves 26459 words which are names of Tunisian cities [21]. Testing the proposed model on three databases was in stages: First, constant documents were segmented into lines, while inconstant documents contain overlap, wavy, or touching problems which are mostly segmented into direct words. The hybrid approach achieved high success rates, as shown in Table II which figures out the distribution of success rates over the three databases. The proposed method achieved a lower accuracy rate in our database for the following reasons:

1) Poor lighting and quality problems.

2) Degradations in the document images.

3) Colored images, while the images in the other two databases are binary.

4) Our database involves more documents than the other two databases.

Table II and Fig. 7 show the words and character segmentation for three databases. In addition, 1450 Arabic handwritten document images were taken from the three mentioned databases. This study focused on images containing overlap and touch problems of words or characters and wavy lines. Table III illustrates the results of the proposed method.

TABLE II. Segmentation Accuracy of Words and Characters

Databases	Number of:		Correct Segmentation		Incorrect Segmentation		Accuracy Rate	
	Words	**Characters**	**Words**	**Characters**	**Words**	**Characters**	**Words**	**Characters**
KHATT	165890	589924	161411	548629	4479	41294	97.3%	93%
IFN/ENIT	26459	212211	24342	188867	2910	12732	92%	89%
Our DB	302348	1257191	275136	1111356	27211	150863	91%	88%

TABLE III. Segmentation Accuracy of Touching Words and Characters

Inputs	NCg	NCr	NCo	NCi	NCc	SR	P	Re	Fm	CR
words	26468	25409	985	833	25357	99	1.00	0.99	1	89.130
characters	79404	74639	2896	2328	74071	99.23	1.00	0.96	0.99	87.4

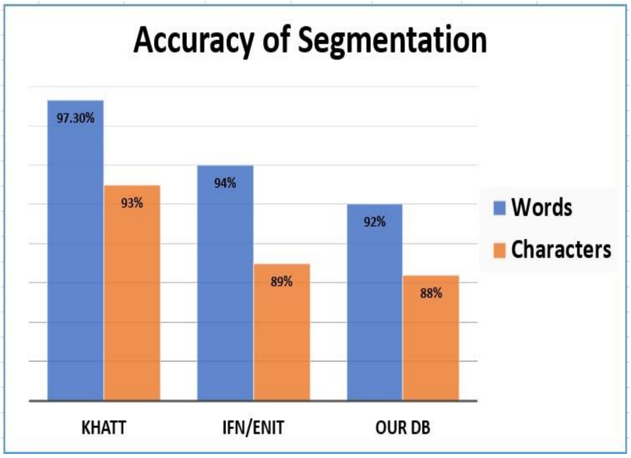


Fig. 7. Segmentation Accuracy of Words and Characters for each Database.

Second: 26468 Arabic handwritten words were taken from the three databases, which contain complex touching and overlapping problems. The ground-truth value of lines and words was calculated manually in our database. The statistics of the performance measure for words, as shown in Table III, indicate that the segmentation rate was 99% successful (SR); this segmentation rate is divided into over-segmentation, correct and incorrect segmentation. The accurate segmentation rate (CR) for words is 89%; Recall (Re), precision (P) and F-measure (Fm) obtained approximately 1.00, respectively. The performance measure indicates that the incorrect segmentation rate is approximately 11%; it means that the number of over-segmentations is 985 words (NCo), and 833 words are missed or incorrect segmentation (NCi). As for segmentation, the number of touching Arabic letters is 79404. These letters are either isolated from the word's origin or have single or multiple touching. Table III also indicates that 99% of the total segmented Arabic letters, 87% of the correct segmentation (CSR), and 13% of them was wrong segmentation. 79404 of touching characters were segmented. (P) was 0.96. (Fm) and (Re) were 1.00. Table IV shows the types of challenges in Arabic handwriting obtained from Table III.

Table V shows the comparison results between the outputs of this work with previous works that discussed such problems. On the other hand, the comparison results are difficult because of the different documents collected by researchers from various Arabic handwritten sources that involve these relevant challenges. The primary purpose is to reach satisfactory and accurate solutions even though the collected documents differ, but the essence of the problem is common. It should be noted that the comparison results are based on the findings of the researchers' works.

TABLE IV. SHOWS THE ACCURACY OF THE HYBRID METHOD IN EACH TYPE OF INPUTS

Type of Characters	**No of characters**	**Correct Segmentation**	**Accuracy**
Isolated	28401	26696	94%
Single Touching	32573	28012	86%
Multiple Touching	18430	15186	82%
Total	79404	69452	87.4%

TABLE V. COMPARISON ANALYSIS

Reference	**Technique**	**No. of images**	**Types of Input**	**Segmentation ratio**
(Aouadi & Kacem, 2017) [22]	Nearest model selection and detect the model's parts centres	820	Segment touching lines and words.	94%
(Ullah et al., 2019) [4]	Overlapping Set Theory and Contour Tracing	220	Segment a single touching character	97.2 %
Proposed method	Hybrid approach	huge	Segment a overlapping and single/ multiple touching characters	90%

VI. CONCLUSION

Characters segmentation in Arabic handwritten is a complex task. However, to improve the efficiency of the recognition system, it is desirable to enhance the character segmentation process in Arabic handwritten. To this extent, this article proposed a hybrid approach for Arabic handwritten character segmentation considering the overlapping or multiple touching issues. The proposed approach aims at improving the efficiency of the segmentation stage. The evaluation is conducted over Arabic handwritten databases which contain more complex challenges than the previously existing ones. The results showed that the proposed approach was highly efficient in word segmentation. And it is an effective, feasible and flexible approach in the segmentation of interconnected, overlapping or multi-touching Arabic characters. But errors or over-segmentation may occur in multi-touched characters. Therefore, in future work, the problem of over-segmentation of multi-touch characters will be considered. Pre-processing operations have also to be improved in formatting very poor documents.

REFERENCES

[1] Qaroush, Aziz, et al. "An efficient, font independent word and character segmentation algorithm for printed Arabic text." Journal of King Saud University-Computer and Information Sciences 34.1 (2022): 1330-1344.

[2] Jindal, Payal, and Balkrishan Jindal. "Line and word segmentation of handwritten text documents written in Gurmukhi script using mid point detection technique." 2015 2nd International Conference on Recent Advances in Engineering & Computational Sciences (RAECS). IEEE, 2015.

[3] Louloudis, Georgios, et al. "Text line and word segmentation of handwritten documents." Pattern recognition 42.12 (2009): 3169-3183.

[4] Naeem-Ullah, Unsar, et al. "First authentic report of Spodoptera frugiperda (JE Smith)(Noctuidae: Lepidoptera) an alien invasive species from Pakistan." Applied Sciences and Business Economics 6.1 (2019): 1-3.

[5] Louloudis, Georgios, et al. "Text line and word segmentation of handwritten documents." Pattern recognition 42.12 (2009): 3169-3183.

[6] Shamsan, Ehab A., et al. "Off line Arabic handwritten character using neural network." 2017 IEEE 4th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA). IEEE, 2017.

[7] Zeki, Ahmed M. "The segmentation problem in arabic character recognition the state of the art." 2005 International Conference on Information and Communication Technologies. IEEE, 2005.

[8] Kang, Le, and David Doermann. "Template based segmentation of touching components in handwritten text lines." 2011 International Conference on Document Analysis and Recognition. IEEE, 2011.

[9] Hamid, A. and R. Haraty. A neuro-heuristic approach for segmenting handwritten Arabic text. in Proceedings ACS/IEEE international conference on computer systems and applications. 2001. IEEE.

[10] Belaïd, A. and N. Ouwayed, Segmentation of Ancient Arabic Documents, in" Guide to OCR for Arabic Scripts," Eds. Volker Märgner and Haikal El Abed. 2011, Springer-Verlag, London.

[11] Farulla, G.A., N. Murru, and R. Rossini, *A fuzzy approach to segment touching characters.* Expert Systems with Applications, 2017. 88: p. 1-13.

[12] Radaideh, A.A. and M.S.M. Rahim, *Existing techniques in Arabic characters recognition (ACR).* Journal of Informatics and Mathematical Sciences, 2016. 8(5): p. 347-360.

[13] Ahmad, I. and G.A. Fink. Class-based contextual modeling for handwritten Arabic text recognition. in 2016 15th international conference on frontiers in handwriting recognition (ICFHR). 2016. IEEE.

[14] Brodic, D. and Z.N. Milivojevic, Text line segmentation with the algorithm based on the oriented anisotropic Gaussian kernel. Journal of Electrical Engineering, 2013. 64(4): p. 238.

[15] Surinta, O., et al. A path planning for line segmentation of handwritten documents. in 2014 14th International Conference on Frontiers in Handwriting Recognition. 2014. IEEE.

[16] Sanchez, A., et al. Text line segmentation in images of handwritten historical documents. in 2008 First Workshops on Image Processing Theory, Tools and Applications. 2008. IEEE.

[17] Hashrin, C., et al. Segmenting Characters from Malayalam Handwritten Documents. in 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT). 2019. IEEE.

[18] Alaei, A., U. Pal, and P. Nagabhushan, *A new scheme for unconstrained handwritten text-line segmentation.* Pattern Recognition, 2011. **44**(4): p. 917-928.

[19] S Deshmukh, Manisha, and Satish R Kolhe. "A hybrid character segmentation approach for cursive unconstrained handwritten historical Modi script documents." Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM), Amity University Rajasthan, Jaipur-India. 2019.

[20] Mahmoud, Sabri A., et al. "KHATT: An open Arabic offline handwritten text database." Pattern Recognition 47.3 (2014): 1096-1112.

[21] Pechwitz, M., et al. IFN/ENIT-database of handwritten Arabic words. in Proc. of CIFED. 2002. Citeseer.

[22] Aouadi, N. and A. Kacem, *A proposal for touching component segmentation in Arabic manuscripts.* Pattern Analysis and Applications, 2017. 20(4): p. 1005-1027.

Machine Learning for Exhibition Recommendation in a Museum's Virtual Tour Application

Shinta Puspasari[1]
Department of Informatics
Universitas Indo Global Mandiri
Palembang, Indonesia

Ermatita[2]*
Department of Computer Science
Universitas Sriwijaya, Indralaya
Indonesia

Zulkardi[3]
Department of Mathematics
Education, Universitas Sriwijaya
Indralaya, Indonesia

***Abstract*—The museum visit is having a crisis during the COVID-19 pandemic. SMBII Museum in Palembang has a remarkable decrease of visitors up to 90%. A strategy is needed to increase museum visits and enable educational and tourism roles in a pandemic situation. This paper evaluates the machine learning model for exhibition recommendations given to visitors through virtual tour applications. Exploring unfamiliar museum exhibitions to visitors through virtual museum applications will be tedious. If virtual collections are ancient and do not display any interest, they will quickly lead to boredom and reluctance to explore virtual museums. For this reason, an effective method is needed to provide suggestions or recommendations that meet the interests of visitors based on the profiles of museum visitors, making it easier for visitors to find exciting exhibition rooms for learning and tourism. Machine learning has proven its effectiveness for predictions and recommendations. This study evaluates several machine learning classifiers for exhibition recommendations and development of virtual tour applications that applied machine learning classifiers with the best performance based on the model evaluation. The experimental results show that the KNN model performs best for exhibition recommendations with cross-validation accuracy = 89.09% and F-Measure = 90.91%. The SUS usability evaluation on the exhibition recommender feature in the virtual tour application of SMBII museum shows average score of 85.83. The machine learning-based recommender feature usability is acceptable, making it easy and attractive for visitors to find an exhibition that might match their interests.**

Keywords—Machine learning; recommender system; museum exhibition; virtual tour; pandemic

I. Introduction

As institutions with educational and tourism purposes, museums must maintain their existence during the pandemic that impacts museum visits. The museum must transform into a modern museum based on digital technology to maintain its roles in pandemic situations. Utilizing the internet of things technology will benefit museums. Visitors can interact with museum collections without having physical contact with the collections, including through applications based on Augmented or virtual reality, robotics, and games [1]. Visitors can explore the museum through a virtual museum application [2]. With this application, visitors can explore every corner of the museum and the collections displayed as if they were exploring the real physical museum [3].

The problem arises when exploring the museum through a virtual tour without a guide. Visitors will be confused and need extra time to explore each exhibition room with many collections. As a result, it can lead to boredom due to the lack of clarity in the exhibition that attracts visitors and spends more time visiting [4]. Visitors will immediately leave the boring virtual space to look for other exciting spaces for educational and tourism purposes [5], especially during the pandemic where physical activity is limited. It needs a method that intelligently understands visitors' interest in the museum's exhibition and presents exciting and interactive information that satisfies visitors to explore the virtual museum.

Machine learning has been proven its effectiveness in understanding user profile and providing recommendations [6]. Based on the user profile, this algorithm learns the information extracted from past data to provide prediction of tourist destinations [7] and recommendations [8] or museum experience improvement [9]. Machine learning does not require formulating a mathematical model to predict or provide recommendations to users. The model is built based on the collected data for training the model [10]. Good training data will produce good recommendations as expected.

This study evaluates several machine learning classifier models to provide exhibition recommendations that match the user profile of a virtual museum. The boredom because of too long browsing the virtual museum displaying collections that do not match the user's interests will make the user exit the virtual tour application. Machine learning is expected to predict and recommend exciting exhibitions and match user requirements or profiles, reducing browsing time or finding interesting collections for virtual museum visitors [11]. Visitors can quickly enjoy exciting presentations and interact virtually with the collection through a virtual tour application based on machine learning recommendations, reducing boredom due to monotonous information presentations [12]. Visitors' interest will be indicated in the performance of the museum's educational and tourism roles. Visitors' knowledge about museums is expected to increase, and visitors' interest in museums through the information presented virtually [13]. The visitor will visit the physical museum when possible during a pandemic after being interested in virtual exploration and might have an economic impact on physical museums to improve the museum visit.

II. Literature Review

Traditional museums have been transformed into modern museums based on digital technology and have begun to

*Corresponding Author.

develop based on artificial intelligence and Internet of Things technology. During the pandemic, digital museums enabled museums to play their role in tourism and education by conducting virtual museum exhibitions. Like a tour guide, the exhibition recommender feature in the virtual tour application makes more accessible exploration when visitors are unfamiliar with the museum.

A. *Use of Museum's Exhibition Virtual Tour Application*

The museum's exhibition display museum's collections that attract visitors. The visitors' experience and satisfaction depend highly on the exhibition environment [14]. Visitors can visit for a long in the museum space or even experience boredom and immediately leave negative sentiments. It impacts return visits to the museum or will affect the interest of new visitors to visit the museum.

Modern museums are transformed based on digital technology to survive in the digital era and meet the millennial's needs for speed of access to information based on internet technology with an exciting and interactive presentation [15]. The museum, identical to ancient and boring collections, presents a monotonous exhibition system, so it needs a touch of digital technology for educational and tourism purposes. The current pandemic hinders visits to physical museums and geographical constraints, such as in Indonesia, where museums with historical and cultural collections are spread across the archipelago, became barriers for visitors. Internet-based technology enables to remove the barriers. Computer technology that can present museum collections and exhibits and allow visitors to explore the museum as if they were in the real environment of a physical museum is a 360-degree virtual tour application. This technology allows visitors to interact with the museum, that presents an image of the real museum environment and interacts with 360-degree views [16]. Visitors can explore independently without the limitations of distance, space and time like a physical museum.

Some museums have integrated virtual tour applications into their virtual museums. However, it will not be easy to explore each static museum exhibition without a guide who plays a vital tourism role [17]. Suppose in a physical museum there is a guide who can suggest and direct visitors to collections or exhibitions that might interest visitors. In that case, a virtual museum should also be equipped with a virtual guide feature to reduce boredom while exploring the museum through a virtual tour application. Some museums in Indonesia [18] have integrated the virtual tour 360-degree feature in their virtual museums (Table I). However, none integrates a feature as a recommender that gives suggestions like a tour guide. The virtual museum was designed as an information system without interactive features for visitors' experience enhancement and engagement. The virtual tour feature adds value to attract physical museums because visitors will still come to physical museums to interact and experience the real museum environment, which is not obtained through virtual museums [19]. Virtual tours will engage visitors to the physical museum for tourism or educational purposes. More features need to be developed and integrated into the virtual museum for the museum's role optimization.

TABLE I. REVIEW OF INDONESIAN VIRTUAL MUSEUMS

Museum Object	Virtual Museum	Virtual Tour	RF*
Museum Nasional	https://www.museumnasional.or.id	✓	x
Museum Bank Mandiri	https://museummandiri.wixsite.com/mbmcorner	x	x
Museum Tsunami	https://museumtsunami.id/	x	x
Museum Balai Kirti	https://balaikirti.kemdikbud.go.id/	✓	x
Museum Sumpah Pemuda	http://museumsumpahpemuda.kemdikbud.go.id/	✓	x
Museum Perumusan Naskah Proklamasi	https://kebudayaan.kemdikbud.go.id/mpnp/	x	x
Galeri Nasional	http://galeri-nasional.or.id/	x	x
Museum Basoeki Abdullah	https://museumbasoekiabdullah.id/virtual/	x	x
Benteng Vredeburg	https://vredeburg.id	x	x
Museum Tekstil Jakarta	https://www.mitramuseumjakarta.org/tekstil	x	x

*RF = Recommender Feature

B. *Machine Learning for Recommender System*

Machine learning methods have been applied to various fields such as education, economics, tourism, and cultural heritage. Machine learning models that have proven their effectiveness in the cultural field, especially museums, include K-Nearest Neighbor (KNN), Decision Tree, Random Forest, Neural Network, and Support Vector Machine [11]. Machine learning analyzes user profiles for various application purposes to improve performance, including education [20], economics and industry [21], and socio-culture [11]. An application of machine learning for museums is a recommender system. It is software providing suggestions related to decision making [22]. A recommender system is built to attract tourist visits for the tourism sector. It enables visitors to find what they need in a large museum [23]. Machine learning is applied by analyzing tourist profiles to find destinations that match the profile based on historical data. This method is effective for tourists to find travel destinations easier, especially for unfamiliar tourists with the destinations, products, or tourism services provided. ML classifier used for filtering and recommending a cultural item to explore, such as museums [23], was implemented in a mobile app but did not discuss the ML classifier selection phase. ML was also proven to give suggestions for museum curators regarding the preventive actions to be taken for historical buildings conservation [24]. However, it has no deep discussion on data processing and ML performance analysis. The app has not been built yet for user feedback evaluation, especially for the pandemic situation when users have limited access to the physical, cultural heritage such as the museum, and the climate changes during the pandemic might affect the building.

The museum is one of the exciting spaces for tourism and learning. The museum collections can be thousands, such as in the Sultan Mahmud Badaruddin II (SMBII) Palembang museum, making it is not easy for visitors to find collections that interest them. Monotonous exhibition rooms, ancient and many collections can make visitors feel bored and not excited exploring the museum virtually by just staring at the screen. The role of the tour guide is needed to provide suggestions for visits to the exhibition rooms that meet the interests of visitors so they continue to explore the museum. They are expected to provide positive sentiment for repeat visits or influence other people to visit the museum. A recommender system can act as a tour guide in the virtual museum. It gives guidance to collections or exhibitions that might interest visitors. The recommender system has been proven effective in improving students' learning achievement [25], [26]. Moreover, the recommender system is expected to enhance visitors' experience in culture education and tourism at the museum.

III. Research Methodology

This research methodology is divided into two stages. The first is evaluating the machine learning model performance, consisting of data preparation, data processing, and performance analysis. The second stage is developing a virtual tour application implementing a machine learning model based on the performance analysis results in the first stage of this research. Fig. 1 illustrates the research methodology.

A. Data Preparation

The dataset was developed based on measurement of the research variables related to visitor's profile. Research variables consist of independent variables and a dependent variable. The museum exhibitions interest dependent variable was influenced by 6 (six) independent variables: age, gender, origin, education, occupation, and motivation [27]. This study uses data of visitors' visits to the SMBII museum in Palembang, South Sumatra, during the COVID-19 pandemic. SMBII museum visitors remarkable decrease up to 90% during the pandemic and impact on collected data. This study assumes that museum visitors are only Indonesian citizens. Several methods are used for data preparation. Data collection was conducted using a questionnaire at the SMBII museum.

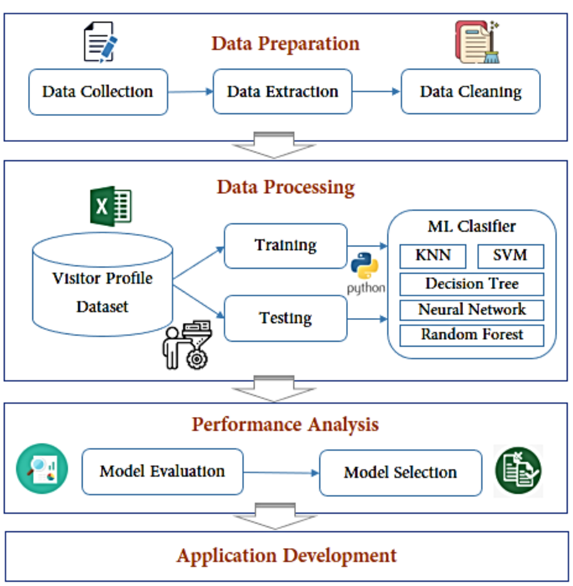
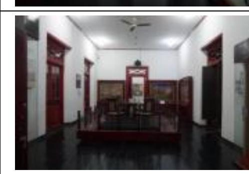
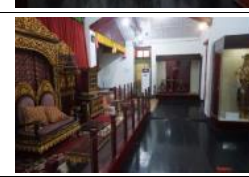


Fig. 1. Research Methodology.

1) Instrument: Data were collected utilizing a questionnaire containing several questions related to the visitor's profile. The questionnaire was divided into two parts. The first part contains questions related to the identity of the respondent and the second part relates to the respondent's interest in visiting the SMBII museum. After exploring the physical museum, respondents were asked to choose their favourite or most interesting exhibition room in the SMBII museum. Table II describes exhibition rooms at the SMBII museum.

TABLE II. Exhibition Rooms in SMBII Museum

Exhibition	**Room View**
Srivijaya Kingdom: Display artefacts from the Srivijaya Kingdom, the largest kingdom in Asia	
Pre-Palembang Sultanate: Display collection on Pre-Palembang Sultanate, a transition era after the Srivijaya Kingdom	
Palembang Sultanate: Display artefacts collection related to Palembang Sultanate, such as SMBII painting and suit	
Colonial: Display artefacts related to colonial era such as weapon	
Life Cycle: Display collection of life equipment used by Palembang society in the past	
Art: Display collection related to arts and cultures that exist since the Srivijaya Kingdom	
Craft: Display artefacts of traditional craft such as Songket and the tool to produce it	

2) Respondent: Respondents are visitors who visited the SMBII museum during the COVID-19 pandemic. This study uses n = 550 samples with various respondents' profile to build the dataset of museum visitors during the COVID-19 pandemic. Table III describes the demographics of the respondents to develop this research dataset. After the data collection, the next step is data extraction by performing the tabulation and coding process of the data to be processed quantitatively with the machine learning (ML) model that will be tested in this study. To optimize the results of the ML classification, then the data cleaning was conducted. The possibility of incomplete data, duplication and various noises interfering with the classification process is carried out on the extracted data. The clean data is saved in .csv format with the Excel tool and ready for data processing.

TABLE III. RESPONDENT'S DEMOGRAPHY

Factor	Criteria	Frequency	Percentage
Gender	Male	200	36%
	Female	350	64%
Age	<22	330	60%
	22-30	151	27%
	>30	69	13%
Education	High School	276	50%
	Undergraduate	254	46%
	Post Graduate	20	4%
Origin	Palembang	414	75%
	South Sumatra	46	9%
	Sumatra	34	6%
	Others	56	10%
Occupation	Student	415	76%
	Teacher/ Lecturer	83	15%
	Employee	41	7%
	Others	11	2%

B. Data Processing

The dataset will be processed using some machine learning classifier models to analyze the profile of SMBII museum visitors who visit the museum in the COVID-19 pandemic situation and provide recommendations for exhibitions room that might match the visitor's interests. The dataset is divided into 80% training and 20% testing dataset. The process began with training machine learning classifier models using the training dataset. The classifier is used to classify the visitor's exhibition interest. The classifiers are Decision Tree, SVM, NN, Random forest, and KNN. The testing dataset tests the ML classifier in classifying visitor profiles and predicting interest in visiting museum exhibitions. Furthermore, the performance of the machine learning classifier model is analyzed and evaluated based on indicators of classification effectiveness accuracy, including cross-validation (CV) accuracy and F-Measure [20], [26]. A confusion matrix is also used to visualize the classification performance of ML classifier. Fig. 2 presents the confusion matrix used for the SMBII exhibition rooms interests classification, where P, Q, R, S, T, U, V represent the exhibition rooms in SMBII museum. The machine learning classifier performance indicators for exhibition classification are formulated based on the confusion matrix. Accuracy and the F-Measure formula are shown in equations (1) and (2). Based on the evaluation results of the classification effectiveness indicators obtained, selecting the best machine learning model is carried out. The selection result will be implemented for exhibition recommendation in the SMBII museum's virtual tour application. The recommendation feature is integrated into the virtual tour application and acts as a tour guide to the user who visits the virtual museum. The effective recommendation will shorten the user's time exploring museum exhibitions minimize the boredom of staring at the screen browsing for collections in museum's exhibitions.

True Label

	A	B	C	D	E	F	G
A	AA	AB	AC	AD	AE	AF	AG
B	BA	BB	BC	BD	BE	BF	BG
C	CA	CB	CC	CD	CE	CF	CG
D	DA	DB	DC	DD	DE	DF	DG
E	EA	EB	EC	ED	EE	EF	EG
F	FA	FB	FC	FD	FE	FF	FG
G	GA	GB	GC	GD	GE	GF	GG

Predicted Label

Fig. 2. Confusion Matrix for SMBII Museum Exhibition Rooms Classification.

$$Accuracy = \frac{(AA+BB+CC+DD+EE+FF+GG)}{Number\ of\ Samples} \quad (1)$$

$$F - Measure = 2\ x\ \frac{Precision\ x\ Recall}{Precision+\ Recall} \quad (2)$$

where Precision and Recall are measured by using formula (3) and (4),

$$Precision = \frac{1}{7}\Big(\frac{AA}{AA+BA+CA+DA+EA+FA+GA} + \frac{BB}{AB+BB+CB+DB+EB+FB+GB} + \frac{CC}{AC+BC+CC+DC+EC+FC+GC} + \frac{DD}{AD+BD+CD+DD+ED+FD+GD} + \frac{EE}{AE+BE+CE+DE+EE+FE+GE} + \frac{FF}{AF+BF+CF+DF+EF+FF+GF} + \frac{GG}{AG+BG+CG+DG+EG+FG+GG}\Big) \quad (3)$$

$$Recall = \frac{1}{7}\Big(\frac{AA}{AA+AB+AC+AD+AE+AF+AG} + \frac{BB}{BA+BB+BC+BD+BE+BF+BG} + \frac{CC}{CA+CB+CC+CD+CE+CF+CG} + \frac{DD}{DA+DB+DC+DD+DE+DF+DG} + \frac{EE}{EA+EB+EC+ED+EE+EF+EG} + \frac{FF}{FA+FB+FC+FD+FE+FF+FG} + \frac{GG}{GA+GB+GC+GD+GE+GF+GG}\Big) \quad (4)$$

C. Museum's Virtual Tour Application Development

The SMBII museum virtual tour application was developed in this study. The application development applied the Multimedia Development Life Cycle (MDLC) methodology, which consists of 6 stages (Fig. 3) [28],

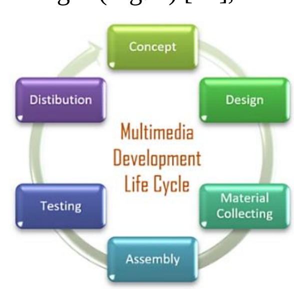


Fig. 3. MDLC Methodology.

1) Concept: The SMBII museum's virtual tour application concept is formulated at this stage by concerning the implementation of n a machine learning model selected based on machine learning model performance analysis results of visitor profiles for museum exhibition recommendations.

2) Design: The design is conducted in line with the formulated concept of the SMBII museum virtual tour application. The design of the interface specifications and the functionality of the application using UML modelling is carried out at this stage. Fig. 4 shows the design of the virtual tour application's use case model diagram. The design is also carried out on the virtual tour asset content in videos, images, text, audio that will be part of the use case: Gallery, Museum History, and Virtual Tour.

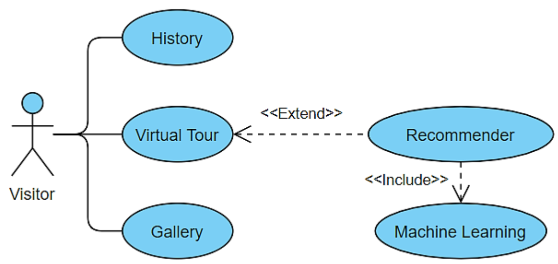


Fig. 4. Use Case Diagram of SMBII Museum's Virtual Tour App.

The Gallery use case will display information on SMBII museum collections and events. The History use case delivers the historical background of the SMBII museum as a cultural heritage protected by the Indonesian government, and the Virtual Tour use case allows users to explore the museum virtually from 360-degrees views. The Recommender use case is extended in the virtual tour use case. This use case includes machine learning to classify the exhibition room that interests visitors as a basis for recommendations given to visitors to quickly find the exhibition room that might match their interests to explore.

3) Material collecting: Multimedia has been designed and constructed utilizing various appropriate tools then collected for integration in the SMBII museum's virtual tour application environment. Table IV describes specific tools that must be prepared for data acquisition of asset content of the virtual tour app.

4) Assembly: The assembly is carried out on each virtual tour's asset in various file formats collected in the previous stage into a virtual tour application that users can run and access. The virtual tour app in this study is designed to be web-based, which makes users can use a browser to run a virtual tour of the SMBII museum without doing the application installation process.

5) Testing: Testing is carried out on the SMBII museum's virtual tour application to ensure each feature successfully carried out its functions as expected, including the machine learning-based exhibition recommender feature. The System Usability Scale (SUS) test was carried out to measure the level of usability of the recommender feature applied to the SMBII museum's virtual tour application by taking $n = 15$ respondents. The samples n were proven effective in generalizing the test results. The samples n were proven effective in generalizing the test results [29]. Fig. 5 presents the rating of usability based on the SUS score.

6) Distribution: After passing the testing stage, the application is uploaded to the server so that the public can widely use it via the internet to explore Palembang's cultural heritage virtually through the SMBII museum's virtual tour application, which provides machine learning-based exhibition recommendations. This recommender feature provides an alternative tour guide for visitors to explore the exhibition room that matches the interest of the SBMII museum's visitors.

TABLE IV. MATERIAL COLLECTING TOOLS

Asset	Data Format	Acquisition Tool
Audio	.mp3	Recorder
Video	.mp4	Video Recorder
Narrative Text	.txt	Text Processor
Image	.jpg	Camera 360

SUS Scoring		
0-64	65-84	85-100
Not Acceptable	Acceptable	Excellent

Fig. 5. Rating of SUS Score.

IV. RESULT AND DISCUSSION

The final output of this research methodology is the SMBII museum's virtual tour application that effectively applies a machine learning model for museum exhibition recommendations. The selection of the ML model was based on the evaluation results of several ML classifier models' performance that were tested using the SMBII museum's visitor profile dataset built in this study.

A. Dataset

The dataset contains 550 SMBII museum visitor's records consisting of data fields: age, gender, origin, education, occupation, motivation, and exhibition, representing research variables. Each variable can be related that affects the other. This study designed the exhibition variable as the dependent variable and the other six variables as independent variables. A heatmap diagram in Fig. 6 illustrates the correlation value between variables in the data set. The correlation score (r) is in the range of $-1< r <1$, which indicates the strength of the relation between the two variables. The closer the r score is to 1, the stronger the relationship with a positive or negative impact [10].

In this study, the positive and negative signs are ignored for the numbers from the data do not indicate the numeric level but rather the label of the data coding results in the data extraction process for a quantitative approach to data processing. From the illustration in Fig. 6, it is known that the variables that have a high impact on visitors' interest in the exhibition space are gender, origin, and occupation variables, with the strongest correlation being the education variable where $r = 0.16$. The educational background of visitors gives the most substantial impact in determining the interest in the exhibition room, which is also influenced by the motivation variable where r = 0.14. The visitor's education is high school and undergraduate in average came to visit a museum with the motivation of educational purposes for study or doing assignments, which is 74% (Fig. 7) with interest in the Srivijaya Kingdom exhibition room as much as 24% and the Palembang Sultanate exhibition room 34% of visits (Fig. 8). The two exhibition rooms present a collection of culture and history regarding the city of Palembang as the oldest city in Indonesia.

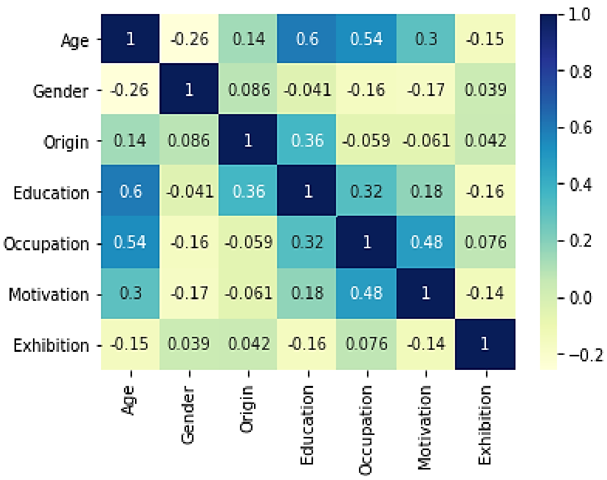


Fig. 6. Data Correlation Heatmap.

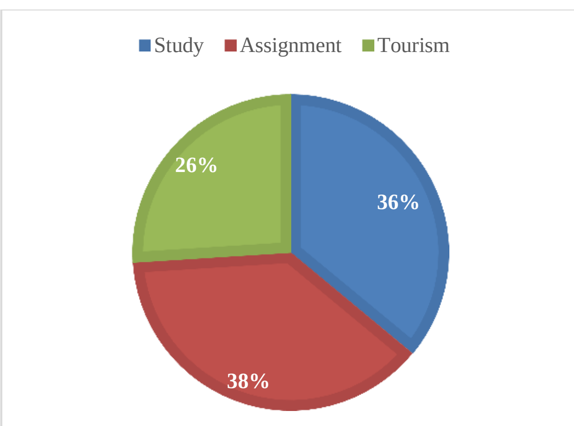


Fig. 7. SMBII Museum Visitors' Motivations.

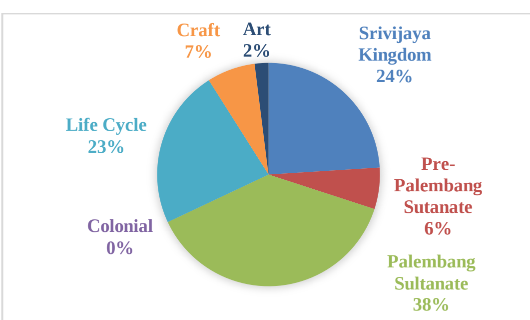


Fig. 8. Visitors' Exhibition Rooms Interests in SMBII Museum.

B. Machine Learning Evaluation for Exhibition Recommendation

Evaluation of the machine learning model for the classification of visitors' interest exploring exhibition rooms in the SMBII museum is used to select the best model applied to the exhibition recommender feature for museum visitors through the virtual tour application developed in this study. Evaluation is carried out on the value of cross-validation accuracy and F-Measure of each machine learning model as an indicator of model performance using a dataset divided into training and testing data. Table V presents the evaluation results based on the cross-validation accuracy and F-measure indicators. Fig. 9 illustrates a graphical comparison of the performance levels of the ML classifier model evaluated in this study. Based on these results, it is known that the best model for classifying exhibition room interest in the SMBII museum based on visitor profiles is KNN with CV accuracy = 89.09% and F-Measure = 90.91%. The RF method also has high accuracy with a difference of 0.18% but a lower F-measure of 0.91%. F-measure scores were obtained based on precision and recall values using different testing data.

Cross-validation accuracy uses data that allows the same training at testing. Therefore, F-measure can show the level of accuracy when the model is used to predict or provide exhibition recommendations that match the interests of visitors to the SMBII museum. The confusion matrix shows the prediction accuracy by calculating the results of true and false predictions. It is illustrated in Fig. 10. It shows a hundred samples of testing data were true predictions. Ten samples were false predictions where the labeled-3 is the most correctly predicted, namely the Palembang Sultanate exhibition room, the favorite exhibition for visits in SMBII museum. The final selection stage determines the KNN model with the best performance for the exhibition recommender feature included in the SMBII museum's virtual tour application.

TABLE V. MACHINE LEARNING MODELS PERFORMANCE FOR CLASSIFICATION OF VISITORS' EXHIBITION ROOM INTERESTS

Classifier	**Accuracy**	**F-Measure**
KNN	89,09%	90,91%
DT	86.55%	90,00%
NN	77.64%	73.64%
SVM	53.09%	53.64%
RF	89.27%	90.00%

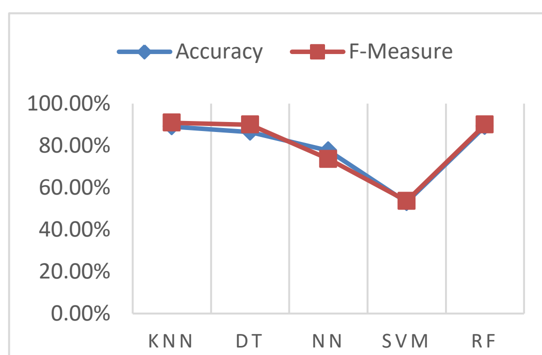


Fig. 9. Machine Learning Models Performance Comparison Visualization.

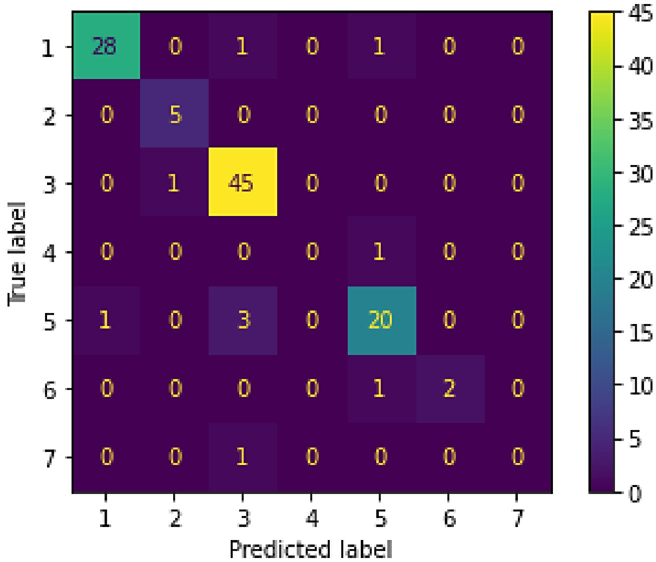


Fig. 10. Confusion Matrix of Data Testing Results.

C. Exhibition Recommendation in Virtual Tour Application

The SMBII museum's virtual tour application was constructed utilizing 3D Vista software. The app enables users to explore the virtual museum like the physical museum with 360-degree views. Fig. 11 is the home page design of the SMBII museum's virtual tour app, which is run on a browser,

There is a tour recommender feature for visitors unfamiliar with the SMBII museum exhibition rooms on the home page. It is Pemandu Pintar (or Smart Guide) feature menu. This feature aims to help visitors find exhibition rooms and collections that are expected to match the interests of users as museum visitors who come with various demographic and motivational backgrounds, including studying, doing assignments, or tourism. Fig. 12 illustrates the interface design of an exhibition recommender feature that applies machine learning to suggest an exhibition room that might match user profiles. The user can input their profile, and then the app processes it by performing the KNN classifier. The machine learning result is presented to the user as a recommendation for starting exhibition room touring from the most interest for visitors predicted by machine learning model (Fig. 12). Users may decide to follow the given exhibition recommendation or explore the museum from the virtual museum entrance. By following the exhibition recommendation given by the machine learning classifier, users will be directed to the recommended exhibition room and then independently explore each corner of the room in 360-degrees views (Fig. 13).

An evaluation was conducted on the usability of the machine learning-based SMBII museum exhibition recommender feature in a virtual tour application. The respondents for SUS testing in this study consider variables with a strong correlation score, $r > 0.1$, on exhibition where there are independent variables: age, education, and motivation. Therefore, some samples are taken in portions based on the variable's value with the highest r score, namely the education variable, while the other variables are taken randomly. The number of samples n =15, the minimum size of samples is obtained for SUS testing and contains all education variable values. Table VI represents the SUS scoring results. Based on these results, it is known that the average SUS score for the exhibition recommender feature is 85.83, which can be categorized as excellent acceptable [30]. This score shows that the exhibition recommender feature implementing the best machine learning model for museum visitor profile analysis, KNN, has a good usability value to meet user needs for exhibition recommendation through SMBII museum's virtual museum application.

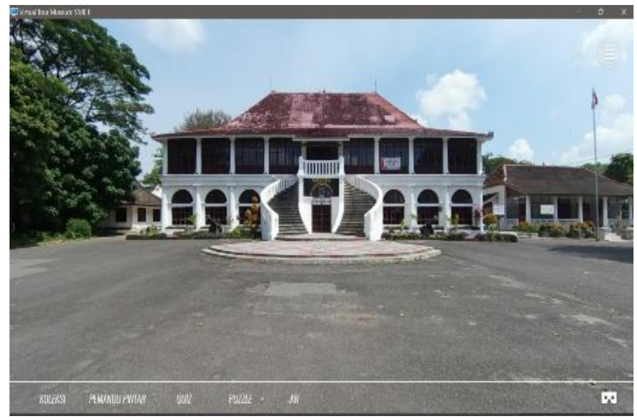


Fig. 11. User Interface of SMBII Museum’s Virtual Tour Home Page.

Fig. 12. User Interface of SMBII Museum’s Virtual Tour Recommender Feature.

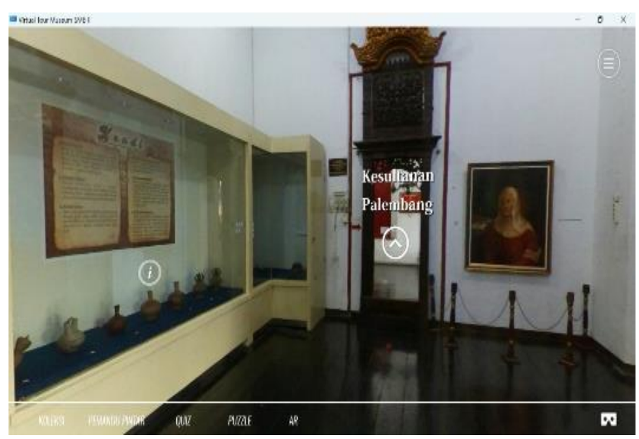


Fig. 13. User Interface of Palembang Sultanate Exhibition Room Virtual Tour.

TABLE VI. SUS SCORING RESULTS

Respondent ID	Education	Age (years)	Motivation	SUS Score
1	High School	13	Assignment	95
2	High School	13	Assignment	95
3	High School	14	Assignment	67.5
4	High School	14	Assignment	67.5
5	High School	14	Assignment	95
6	High School	19	Study	95
7	High School	19	Study	100
8	High School	20	Study	87.5
9	Under Grad	25	Study	72.5
10	Under Grad	27	Tourism	67.5
11	Under Grad	28	Tourism	80
12	Under Grad	33	Tourism	70
13	Under Grad	37	Tourism	100
14	Post Grad	39	Study	95
15	Post Grad	41	Tourism	90

D. Limitations and Contributions

The developed SMBII museum virtual tour is assumed to be only intended for visitors to Indonesian citizens, so the dataset used has limitations, namely on regional origin and occupation factors. The profiles of the respondents of the developed dataset are dominated by museum visitors from Palembang city who were students. The situation of the COVID-19 pandemic is a reason for the small number of visits to the SMBII museum, so not much data has been collected for variations of visitor profiles. The dataset will affect ML's ability to predict visitor interest in the museum exhibition room. The more varied the data used for training, the better ML's ability in testing, especially for SMBII museum exhibition recommendation. It is necessary to add a variety of datasets to improve ML performance for predictions and recommendations of museum exhibition in the SMBII virtual museum tour application. In addition, it needs to test the ML classifier model for different datasets. The SMBII Museum is a cultural heritage that stores various collections of cultural objects. It is necessary to test whether the KNN method is still the best to recommend a visit to an exhibition if the museum presents collections other than cultural objects, such as history, art, and others.

Hopefully, a recommendation that matches visitors' interests in the museum's exhibition can provide convenience in exploring and enjoying the presentation of the museum's collections for learning and tourism purposes while reducing boredom in exploring the monotonous virtual exhibition room and making it less attractive to visitors. Visitors can quickly find the exhibition offerings that match their interests. The findings of this study can be a baseline for determining an effective ML method to be applied in an intelligent virtual tour application of a museum as a strategy to increase museum visits and enable educational and tourism roles in the COVID-19 pandemic. The recommendation feature that implements ML has proven effective in making it easier for visitors and is expected to reduce boredom in exploring the virtual museum.

V. CONCLUSION

This study evaluates the machine learning model for exhibition recommendations given to visitors through virtual tour applications based on user or visitor profiles. The machine learning-based recommender feature in the virtual tour application will act as a tour guide that gives suggestions to the exhibition room, which might interest the museum's visitors. The best machine learning model was selected based on its performance assessed based on the cross-validation accuracy and F-measure scores. The dataset of visitor profiles consists of the exhibition variable as the dependent variable, and the independent variables are age, gender, origin, education, occupation, and motivation. The results of the correlation analysis show that the motivational variable has the strongest correlation to exhibition interest with a score of r = 0.16. The dataset is then processed using the machine learning model to classify the exhibition room interest based on the profile of visitors to the SMBII museum, namely the KNN, DT, NN, SVM, and RF models. Based on testing results, it is known that the best machine learning model for classifying interest in visiting the SMBII museum's exhibition room is the KNN model with cross-validation accuracy= 89.09% and F-Measure = 90.91%. These results indicate that the machine learning model effectively predicts the interest in the SMBII museum's exhibitions room and can be applied to provide recommendations for virtual museum visitors.

The SMBII museum's virtual tour application applied the KNN method for the recommender feature of the museum's exhibition was developed using MDLC methodology. For usability evaluation, SUS testing was conducted to evaluate the usefulness of the exhibition recommendation features in the application. The testing results obtained the average SUS score = 85.83, which indicates that the recommender feature in the virtual tour application is acceptable for the user. This feature makes users easily find exhibition rooms that match their interests and explore interactive museum spaces, reducing boredom in exploring museum exhibitions with lots of rooms and thousands of historical and ancient collections, especially for millennial students. Based on the SUS scores can also be concluded that the machine learning based exhibition recommendation is easy to be used and is expected to attract public interest in visiting museums that increase museum visits for educational or tourism purposes, especially during the COVID-19 pandemic.

REFERENCES

[1] Y. Luo, "On the Information Application of Multimedia Technology in Museum," in Proceedings - 2020 5th International Conference on Communication, Image and Signal Processing, CCISP 2020, 2020, pp. 299–303, doi: 10.1109/CCISP51026.2020.9273468.

[2] E. Bran, E. Bautu, and D. M. Popovici, "Towards a sustainable future: Ubiquitous knowledge mixed reality museum," Procedia Comput. Sci., vol. 176, pp. 2878–2885, 2020, doi: 10.1016/j.procs.2020.09.266.

[3] D. A. Loaiza Carvajal, M. M. Morita, and G. M. Bilmes, "Virtual museums. Captured reality and 3D modeling," J. Cult. Herit., vol. 45, pp. 234–239, 2020, doi: 10.1016/j.culher.2020.04.013.

[4] C. Antón, C. Camarero, and M. J. Garrido, "A journey through the museum: Visit factors that prevent or further visitor satiation," Ann. Tour. Res., vol. 73, no. September, pp. 48–61, 2018, doi: 10.1016/j.annals.2018.08.002.

[5] J. Pallud and D. W. Straub, "Effective website design for experience-influenced environments: The case of high culture museums," Inf.

Manag., vol. 51, no. 3, pp. 359–373, 2014, doi: 10.1016/j.im.2014.02.010.

[6] S. Puspasari, E. Ermatita, and Z. Zulkardi, "Constructing Smart Digital Media for Museum Education Post Pandemic Recovery : A Review and Recommendation," 2021.

[7] B. Sohrabi, I. Raeesi Vanani, N. Nasiri, and A. Ghasemi Rud, "A predictive model of tourist destinations based on tourists' comments and interests using text analytics," Tour. Manag. Perspect., vol. 35, no. July 2019, p. 100710, 2020, doi: 10.1016/j.tmp.2020.100710.

[8] C. Huda, A. Ramadhan, A. Trisetyarso, E. Abdurachman, and Y. Heryadi, "Smart Tourism Recommendation Model : A Systematic Literature Review," Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 12, pp. 167–174, 2021.

[9] M. Majd and R. Safabakhsh, "Impact of machine learning on improvement of user experience in museums," in Artificial Intelligence and Signal Processing Conference (AISP), 2017, pp. 195–200.

[10] R. K. Mishra, "Machine Learning based Forecasting Systems for Worldwide International Tourists Arrival," vol. 12, no. 11, 2021.

[11] M. Fiorucci, M. Khoroshiltseva, M. Pontil, A. Traviglia, A. Del Bue, and S. James, "Machine Learning for Cultural Heritage: A Survey," Pattern Recognit. Lett., vol. 133, pp. 102–108, 2020, doi: 10.1016/j.patrec.2020.02.017.

[12] S. Surapaneni, "Exploring Themes and Bias in Art using Machine Learning Image Analysis," 2020 Syst. Inf. Eng. Des. Symp. SIEDS 2020, 2020, doi: 10.1109/SIEDS49339.2020.9106656.

[13] S. Puspasari, N. Suhandi, and J. N. Iman, "Enhancing The Visitors Learning Experience in SMB II Museum Using Augmented Reality Technology," in International Conference on Electrical Engineering and Informatics (ICEEI), 2019, pp. 296–300.

[14] V. Kamariotou, M. Kamariotou, and F. Kitsios, "Strategic planning for virtual exhibitions and visitors' experience: A multidisciplinary approach for museums in the digital age," Digit. Appl. Archaeol. Cult. Herit., vol. 21, no. December 2020, p. e00183, 2021, doi: 10.1016/j.daach.2021.e00183.

[15] A. Leopardi et al., "X-reality technologies for museums: a comparative evaluation based on presence and visitors experience through user studies," J. Cult. Herit., vol. 47, pp. 188–198, 2021, doi: 10.1016/j.culher.2020.10.005.

[16] L. Argyriou, D. Economou, and "360-degree interactive video application for Cultural Heritage Education," … Learning …. westminsterresearch.westminster.ac …, 2017, [Online]. Available: https://westminsterresearch.westminster.ac.uk/item/q28q0/360-degree-interactive-video-application-for-cultural-heritage-education.

[17] H. Chen, B. Weiler, and R. Black, "Exploring knowledge-building in tour guiding research: A content analysis of empirical papers on tour guiding, 1980–2015," J. Hosp. Tour. Manag., vol. 37, no. October, pp. 59–67, 2018, doi: 10.1016/j.jhtm.2018.09.005.

[18] L. Banun, "Exploring Museum Through Virtual Tour," Indonesian Ministry of Education and Culture, 2021. https://kebudayaan.kemdikbud.go.id/seru-di-rumah-lewat-virtual-tur-museum/ (accessed Aug. 16, 2021).

[19] R. Quattrini, R. Pierdicca, M. Paolanti, P. Clini, R. Nespeca, and E. Frontoni, "Digital interaction with 3D archaeological artefacts: evaluating user's behaviours at different representation scales," Digit. Appl. Archaeol. Cult. Herit., vol. 18, no. April, p. e00148, 2020, doi: 10.1016/j.daach.2020.e00148.

[20] S. D. A. Bujang et al., "Multiclass Prediction Model for Student Grade Prediction Using Machine Learning," IEEE Access, vol. 9, pp. 95608–95621, 2021, doi: 10.1109/ACCESS.2021.3093563.

[21] R. Sindhwani, D. Nagar, S. Raghav, A. Bhardwaj, R. Kumar, and P. L. Singh, "Machine learning : Best way to sustain the supply chain in the era of industry 4 . 0," 2021, doi: 10.1016/j.matpr.2021.01.267.

[22] F. Ricci, L. Rokach, and B. Shapira, Recommender Systems Handbook, no. October\. Springer Science Business Media, 2011.

[23] X. Su, G. Sperli, V. Moscato, A. Picariello, C. Esposito, and C. Choi, "An Edge Intelligence Empowered Recommender System Enabling Cultural Heritage Applications," IEEE Trans. Ind. Informatics, vol. 15, no. 7, pp. 4266–4275, 2019, doi: 10.1109/TII.2019.2908056.

[24] F. M. La Russa and C. Santagati, "An AI-based DSS for preventive conservation of museum collections in historic buildings," J. Archaeol. Sci. Reports, vol. 35, no. November 2020, p. 102735, 2021, doi: 10.1016/j.jasrep.2020.102735.

[25] P. Lin and S. Chen, "Design and Evaluation of a Deep Learning Recommendation Based Augmented Reality System for Teaching Programming and Computational Thinking," IEEE Access, vol. 8, pp. 45689–45699, 2020, doi: 10.1109/ACCESS.2020.2977679.

[26] N. Yanes, A. M. Mostafa, M. Ezz, and S. N. Almuayqil, "A Machine Learning-Based Recommender System for Improving Students Learning Experiences," IEEE Access, vol. 8, no. ii, pp. 201218–201235, 2020, doi: 10.1109/ACCESS.2020.3036336.

[27] F. Z. Kaghat, A. Azough, M. Fakhour, and M. Meknassi, "A new audio augmented reality interaction and adaptation model for museum visits," Comput. Electr. Eng., vol. 84, 2020, doi: 10.1016/j.compeleceng.2020.106606.

[28] S. L. Rahayu, Fujiati, and R. Dewi, "Educational Games as A learning media of Character Education by Using Multimedia Development Life Cycle (MDLC)," in 2018 6th International Conference on Cyber and IT Service Management, CITSM 2018, 2018, no. Citsm, pp. 1–4, doi: 10.1109/CITSM.2018.8674288.

[29] S. Puspasari, N. Suhandi, and J. N. Iman, "Evaluation of Augmented Reality Application Development for Cultural Artefact Education," Int. J. Comput., vol. 20, no. 2, pp. 237–245, 2021, doi: 10.47839/ijc.20.2.2171.

[30] S. C. Mclellan, Sam; Muddimer, Andrew; Peres, "The Effect of Experience on System Usability Scale Ratings," J. Usability Study, vol. 7, no. 2, pp. 56–67, 2012.

Classification of Osteoporosis in the Lumbar Vertebrae using L2 Regularized Neural Network based on PHOG Features

Kavita Avinash Patil[1]
Department of Electronics & Communication Engineering
Sri Jagadguru Balagangadhranatha Institute of Technology
Faculty, New Horizon College of Engineering,
affiliated to Visvesvaraya Technological University
Bangalore, India

K. V. Mahendra Prashanth[2]
Department of Electronics and Communication Engineering
Sri Jagadguru Balagangadhranatha Institute of
Technology, affiliated to Visvesvaraya Technological
University, Bangalore
India

Dr. A Ramalingaiah[3]
Sapthagiri Institute of Medical Sciences and Research Centre, M.S (Ortho) Orthopedic Doctor, India

***Abstract*—One of the most common bone diseases in humans is osteoporosis, which is a major concern for the public health. Osteoporosis can be prevented if it is detected at an early stage. The research agenda consists of two phases: pre-processing of X-ray images of the spine and analysis of texture features from trabecular bone lumbar vertebrae L1-L4 for detecting osteoporosis. The preprocessing involves image enhancement of texture features and co-register the images in order to segment the L1-L4 regions in the lumbar spine. Range filtering and Pyramid Histogram of Orientation Gradient (PHOG) are used to analyze texture features. Input images are filtered with a range filter to adjust the local sub range intensities in a specified window to detect edges. Then a PHOG algorithm is designed to determine both the local shape of an image texture and its spatial layout. Based on texture features of lumbar vertebrae L1-L4, classify them as normal or osteoporotic using neural network (NN) models with L2 regularization. In an experiment, X-ray images and dual-energy X-ray absorptiometry (DXA) reports of individual patients are used to verify the system. DXA reports describe a statistical analysis of normal and osteoporotic results. However, the proposed work is categorized according to the texture features as normal or osteoporotic. 99.34% classification accuracy is achieved; cross-validation of these classified results is done with the DXA reports. Diagnostic accuracy of the proposed method is higher than that of the existing DXA with X-ray. Further, the area under the Receiver Operating Characteristic (ROC) curve for L1-L4 had a significantly higher sensitivity for osteoporosis.**

Keywords*—Cross validation; image enhancement; lumbar spine; lumbar vertebrae; neural network; osteoporosis; PHOG; regularization; trabecular bone; texture features; x-ray images

I. INTRODUCTION

Osteoporosis is a disease which affects the density and strength of bones. A bone's density is the amount of bones (bone mineral density-(BMD)), while its strength (quality) is the fibers in the bones. Osteoporosis leads to weaker, more porous and more brittle bones, as well as a greater risk of fractures [1]. This paper proposed an efficient model for assessing osteoporosis in the lumbar vertebrae L1-L4 using L2 regularized neural network. The model consists of two parts, first part of the system is a pre-processing of an X-ray spine image in order to visualize the fine textures of the trabecular micro architecture of L1 - L4 vertebrae using the following steps: initially, the X-ray image needs to be enhanced by improving the visual quality, and then, the input images need to be aligned into the same plane by using co-registration, and finally, segment the image. As part of the segmentation process, the primary objective is to determine the region of interest (ROI) in the image, namely between the L1 and L4 vertebrae, in order to improve the actual segmentation. As illustrated in Fig. 1, the second part of the system uses PHOG texture features to detect osteoporosis in the L1-L4 lumbar vertebrae using NN with L2 regularization in order to classify normal and abnormal images.

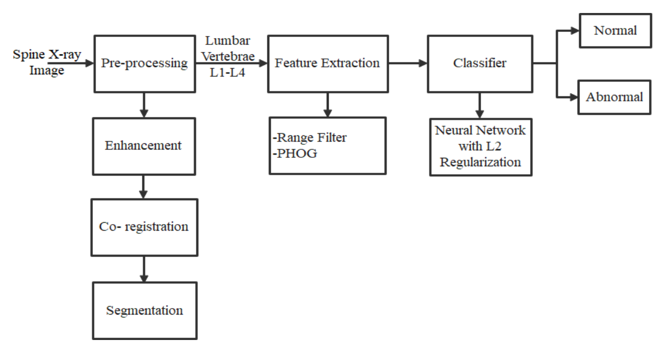


Fig. 1. An Overview of the System Model.

This paper makes the following contributions: The proposed technique for the detection of osteoporosis in the lumbar spine (L1-L4) is described in Section I. Section II outlines the related work of other methods that have been developed so far. The Section III Database detail description is used to test the system. A preprocessing stage was described in Section IV to improve image quality. An in-depth discussion of image texture feature extraction was presented in

Section V. NN networks are regularized by L2 in Section VI, used as a classifier. The Section VII presents a detailed simulation study and discussion. The final Sections VIII and IX of the report summarize the proposed system and the future work required to improve its performance, respectively.

II. Related Work

Osteoporosis is a disorder in which there is a loss of bone mass and abnormally degenerated bone architecture, especially, hips, spine, and wrist, known as osteoporosis [2]. Among osteoporotic fractures, spine fractures are the most common and are a major health concern among the elderly. Consequently, osteoporotic fractures require early diagnosis of patients at high risk. Typically bone density scans are used to detect osteoporosis, among which the dual energy X-ray absorptiometry (DXA) is considered as a common technique to measure bone size and bone mineral density (BMD). Various studies that deal with BMD measurements have been conceded by analyzing image texture features using an X-ray's. An easy and inexpensive method for diagnosing osteoporosis has been proposed based on analysis of image texture features using X-rays. In conjunction with a machine-learning algorithm, a fractal model was utilized to develop the software, using pixel variations for grey levels[3] . An osteoporosis patient's bone structure value (BSV) is estimated from BMD using spine radiograph images. It is necessary to conduct further studies to determine BSV's potential to be a reliable assessment of treatment effects and future fracture risk in individuals with osteoporosis and those without it. The first group had 83 patients treated for osteoporosis alone, while the second group had 76 patients treated for both osteoporosis and lumbar spinal stenosis (LSS) [4]. The T-scores over the first year as well as after a year, two years, and three years were confirmed. Two groups were compared on mean BMD and changes in BMD over three years. In addition, three-year BMD improvements were evaluated along with their relationship to initial BMD change and related factors. Study participants were given ibandronate dose for newly diagnosed osteoporosis, so LSS was examined for its effect on BMD. The study was focused only on whether LSS could improve bone mineral density in the treatment of osteoporosis, clinical outcomes related to osteoporosis treatment, such as osteoporotic fractures, were not assessed in the follow up. An alternative approach to determining the extent of bone loss caused by osteoporosis is fractal analysis when studying spinal CT (computed tomography) images [5]. Based on the results of the study and the K-NN (K-Nearest Neighbor) classifier, a computerized system based on CT images could assist physicians in making initial diagnoses in difficult cases. Overall the system provides 81% classification accuracy; an alternative method is required to improve the system performance. By using the Picture Archiving and Communication System (PACS), the lumbar vertebra was measured in QCT (Quantitative computed tomography) and the HU (Hounsfield unit) of its vertebral body in conventional CT [6,7]. The correlation between the T-score of conventional CT and the T-score of QCT was estimated using a multiple regression algorithm. Further, a logistic regression algorithm was applied to predict osteoporotic and non-osteoporotic vertebrae. With QCT data, the predictor modeling algorithm estimated similar T-scores. In HU, similar results are observed as with QCT, with the exception of one osteoporotic vertebra that did not demonstrate discordance with an accuracy rate of 92.5% was recorded. The predictive accuracy will improve with more collected data. The purpose of study is to find out whether recurrent neural networks are capable of predicting osteoporotic fractures by analyzing spine images [8]. It explores the best design directions for such prediction models by experimenting with various network architectures. Transfer learning gives the advertised benefits, such as faster training speeds and greater suitability for larger datasets. By segmenting and finding vertebral edges, can diagnose the compression and locate the anomaly using Morphometric features and measurements using CT images [9,10]with 88.3% accuracy. There are challenges involved in finding the midpoint in the vertebral body and passing it to the next closest midpoint on the vertebral body boundary to analyse 3D textures, extracted the gray-level co-occurrence matrix Haralick, Wavelets (WL), local binary patterns (LBP), histogram of gradients (HoG), and harmonic alternator patterns (HAR) [11]. The texture-features and vBMD data extraction, fractured vertebrae were excluded. An assessment of prevalence of osteoporosis fourfold cross-validation was conducted to evaluate vertebral fractures. There is a correlation between vertebral level parameters and classification results. Mark-Segmentation-Network and Bone-Conditions-Classification-Network are used to analyze diagnostic CT images to automatically detect bone conditions [12]. The system achieves receiver operating characteristic curve area of 0.9167 and accuracy of 76.62%. Feature extraction from lumbar vertebra CT images as well as other clinical characteristics might be relevant to the diagnosis of the bone condition to improve the system performance. As compared with the traditional Osteoporosis Self-Assessment Tool for Asians (OSTA) model, the ANN, SVM, RF, and LoR models performed significantly better in both men and women [13].

III. Database Details

According to Table I, 80 numbers of samples are used in the process of developing the proposed system. Initially Digital X-ray images of the lumbar spine (lateral view) are taken for processing. Dr. A Ramalingaiah, Orthopedic, provided the images from Abhilasha orthopedic hospital in Banashankari at No.271,3rd Stage, 5th Block, 100ft Road, Bangalore; the lumbar spine of 80 subjects in 2D, JPEG format. An experiment has been conducted to test the system by collecting 20 control subjects (normal) as well as 60 pathological (osteoporotic or abnormal) X-ray images, as well as DXA reports on the same people. In the DXA report, the L1-L4 lumbar spine statistical analysis status details are provided for each person.

TABLE I. Raw Data Base

Lumbar spine X-ray image	No. of Subjects (80)	DXA reports (80)
Controlled Subjects (Normal)	20	20
Osteoporotic Patients	60	60

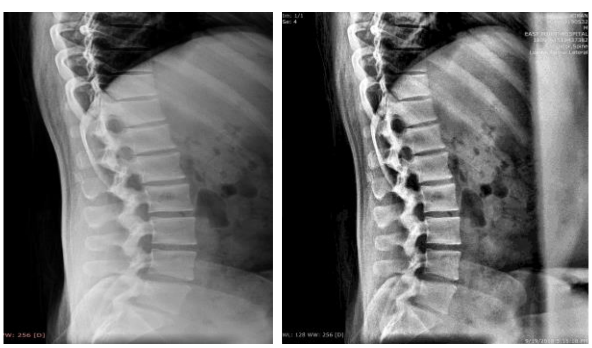
Among 80 subjects, the segmentation algorithm correctly considers 18 (normal) out of 20 and 58 out of 60 (abnormal), which means 18+58=76. A total of 304 ROI images of lumbar spine images (L1-L4) are extracted from each subject, and 18 x 4 = 72 and 58 x 4 = 232 results in 304 ROI images.

IV. Preprocessing Stage

In the preprocessing stage, an adaptive histogram equalization (AHE) and a co-registration are used to obtain L1-L4 region of interested sub images with enhancement.

A. AHE

An image enhancement technique is adapted in this proposed work to provide a more interpretable image, for better quality input to the next phases of the work. A contrast-limited (CL) AHE process is used to enhance the contrast of grey scale image. Rather than processing the entire image, CLAHE focuses on small regions (8×8) of it, called tiles. Divides an image into a number of rectangular contextual regions. It calculates for each region independently by AHE. By enhancing the contrast of each region, the output histogram approximates by the distribution of its parameters. The Rayleigh distribution is used to create a contrast transform function, which depends on the input image type. Using this distribution, imagery appears more natural. Bilinear interpolation [14] is then used to combine the neighbouring blocks, eliminating artificial boundaries. Contrast can be reduced in homogeneous areas to prevent noise from being amplified. With clip-limit (0.01), further prevent the oversaturation of homogeneous areas of an image. In such areas, the histogram of an image is dominated by a high peak because many pixels are within the same gray scale range. The spine lateral view enhanced image (b) and the input image (a) as shown in Fig. 2.



(a) (b)

Fig. 2. (a) Input Image (b) Enhanced Image.

The algorithm-1 comprises into three major components in CLAHE: tile generation, histogram equalization, and bilinear interpolation.

Algorithm 1

Input: M×N image, output: Enhanced image

1. Sections (sub image) are divided in the input image. These sections are referred to as tiles of N×N.
2. Obtain the histogram distribution for each tile.
3. Minimum and maximum values are found through binary search based clipping [15] in the distribution and clipped portions are redistributed.
4. In the next phase, bi-linear interpolation is carried out to combine the neighbouring blocks. The resultant one is an enhanced image with greater contrast.
5. The entropy [16] H is calculated using (1)

$$H = -\sum (h_c \times \log(h_c)) \quad (1)$$

In this case, h_c is the histogram counts.

6. Ensure $N \leftarrow N + 2$.
7. In order to measure the entropy H_{i+1} of the contrast-enhanced image, Where $i = 1, \ldots\ldots, N$ repeat steps 2 through 5 with a new window size.
8. If $H_{i+1} < H_i$ stop, otherwise repeat $N < 8$.
9. It is certain that the enhanced image with the largest H value will also have better quality, and enhanced image is obtained at the output.
10. According to the experimental results, binary search has been shown to be an effective algorithm in terms of clipping and redistributing the pixels.

B. Co-registration

Several biomedical imaging applications require the co-registration of images. The images may be obtained from the same sensor or from different sensors, and the spatial resolution might be the same or different. During co-registration, all images in the series are aligned spatially so that any feature in one image overlaps as well as its footprint in every other image [17,18]. An image is typically selected as the reference to which all other images are aligned during co-registration. The best reference image can be selected from the database, which contains all the features of lumbar spine lateral view. The co-registration process involves identifying common features in the reference and warping the other images that is the ones to be co-registered [19]. In the process, tie points are used to determine the locations of common features. The warped image is aligned to the reference using a polynomial function after enough tie points have been generated.

In Fig. 3, an example of the output of a co-registration process as illustrated. In part (a), the reference image is the normal image that all the target images are aligned to the same feature plane; the target image (b) does not resemble the reference image, as it represents osteoporosis; c) is overlaid image of R and co-registered, the co-registered image is completely aligned with the reference image using Geoscience extended flow optical lucas- kanade iterative (Gefolki) is a co-registration software [20,21]; overlays R and C in different colour bands to form a composite RGB image. A grey region indicates the same intensity of the two images in the composite image. Magenta and green colour regions indicate differences in intensities. (d) An image co-registered with the R image that shows the exact alignment of the T features.

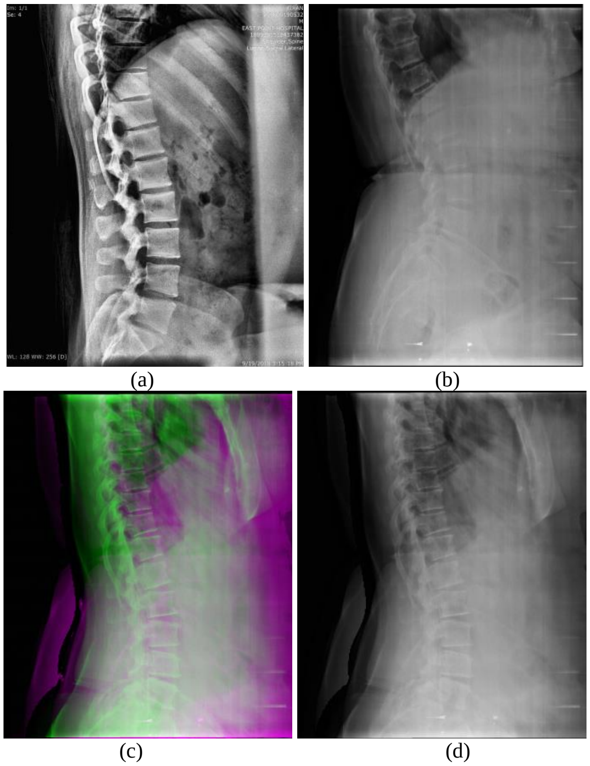


Fig. 3. (a) Reference (R) Image (b) Target (T) Image (c) Overlaid R and Co-Registered (C) Image (d) Co-registered Image.

Algorithm 2

A pixel alignment of two images can be divided into three steps: initialization of input image R and T of the same size, calculation of GeFolki flow, and resampling.

1. Homothety, rotation and scaling is used to make size of T to size R, denoted as T'.
2. GeFolki flow, the transfer from coordinates of T' to R by a matrix W.
   - W is composed of the y-displacement component (column) and the x-displacement component (row) for applying on the T' image.
   - In the software, GeFolki function takes as arguments T', R, and parameters, with order as radius = 64: -8:16, level = 4, iteration = 5, contrast_adapt=false, and rank=4 as parameters.
   - A variety of radius sizes can be tested iteratively; it is a decreasing vector starts from biggest (64) radius to the smallest (16) one in steps of 8. The algorithm is more robust when the radius is large. Radii must be selected as small as possible when the flow on the image changes rapidly.
   - For the purpose of finding large displacements, create a pyramidal structure of the down sampled images in different levels. In a pyramid, the number of levels affects the size of the movements.
   - The total number of iterations required to run the gradient method for the minimum search.
   - Set contrast_adapt is false for homogeneous images to look contrast inversion.
   - Changing the intensity value of T' with lower rank of pixel in the neighborhood within the specified window size (9×9).
3. Finally, transform the T' image to superimpose it with the R image. By resampling the image on a new coordinate grid, this operation creates the flow matrix W. Resampling is done using bilinear interpolation.

C. Region L1-L4 Sub Image Selection

L1-L5 are the five vertebrae that make up the Lumbar Spine. As the largest vertebrae in the body [22][23], the lumbar spine bears the most weight. In contrast to the thoracic spine, Lumbar spines region have more range of motion. There is limited rotational movement in the facet joints of the lower back. In this work L1-L4 is the region of interested subimage. An illustration of the interested region selection process is presented in Fig. 4 of (a)–(g) and explained in Algorithm 3.

Algorithm 3

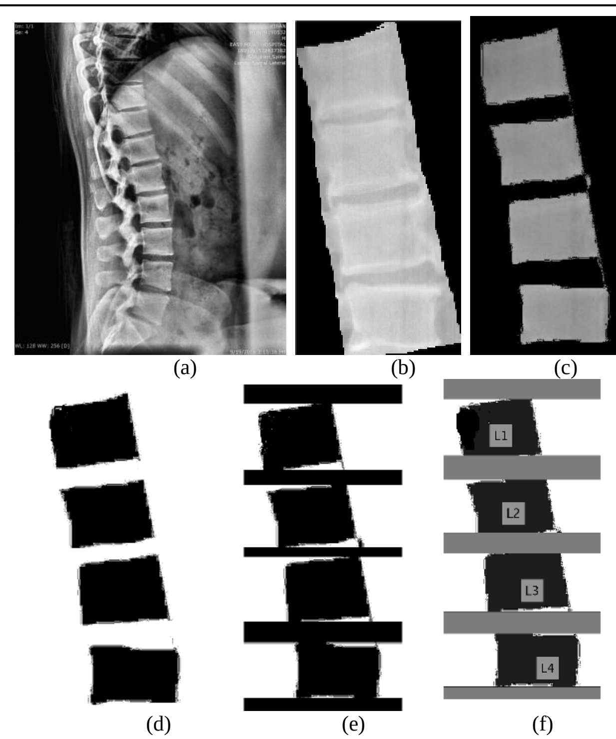
1.The lateral spine view X-ray images are read from the input.
2. Obtain the border coordinates of the freehand drawn region from L1-L4, and divide the result by an array of 2-dimensional matrices that is subimage
3. Using the subimage object, produce a binary image mask.
- It returns a mask that is associated with the subimage object B over the target image.
- The target image must be contained within the same axes as the subimage.
- Mask is a logical image the same size as the target image.
- Mask is false outside the region of interest and true inside.
- Multiplied mask with the input image to produce segmented L1-L4 region.

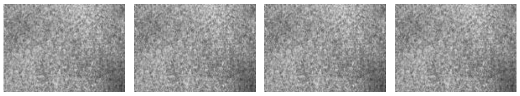
4.Create the inverted binary image from segmented image

```
    segmented image pixels >255;
            Seg_inv=0;
else
Seg_inv= segmented image;
make it into a proper binary image for foreground
Inv_binary_im = Seg_inv < max (Seg_inv (:))/2;
```

5. Create the isolated L1-L4 image with equal distance and equal image size
- find the bar rows
  bar_rows = sum (Inv_binary_im, 2) > 0.9 * size (Inv_binary_im, 2);
- make a bar image
  im_bars = false(size(Inv_binary_im));
  im_bars (bar_rows, :) = true;
- remove the bars from the image
  im_nobars = Inv_binary_im & ~im_bars;

6.Label each lumbar vertebrae by finding its centroid.
7.Select only the connecting pixels on the ROI images for L1, L2, L3, L4.





(g)

Fig. 4. Interested Region Selection Process (a) Original LS view Image (b) Cropped Image (c) Segmented Image (d) Inverted Binary Image (e) Isolated L1-L4 Image (f) Labelled Image (g) ROI Image of L1, L2, L3, L4.

V. Feature Extraction

The feature extraction process involves two stages; Range filter and PHOG. The range filter is used to adjust the local sub range intensities within the specified window size of 3×3 neighborhood pixels in the input image. In a local sub-range the pixels are analyzed according to their statistical range to detect edges.

The PHOG objective is to determine the local shape of an image texture and its spatial layout. By considering the orientations of edge distribution within the sub-image as well as the spatial arrangement, the image is tiled into regions at multiple resolutions to obtain texture features.

A. Range Filter

It operates on morphological utilities such as dilation and erosion of the image to regulate the maximum and minimum pixel values in the specified window [24]. Subsequently, it utilizes the padding operation on these morphological utilities. In Fig. 5, (a) the input and its corresponding histogram are clearly visible, showing the right-hand side of the histogram occupied by the input image and (b) the filter output and its histogram shows that the pixel ranges are narrowing by connecting local neighbors as edges, which reveals more defined texture in the image. In Fig. 6(a) and (b), the range filter response is shown for an abnormal image. A normal and abnormal image can't be distinguished by looking at them. Range filtering connects local neighbors as edges to narrow the histogram of the input.

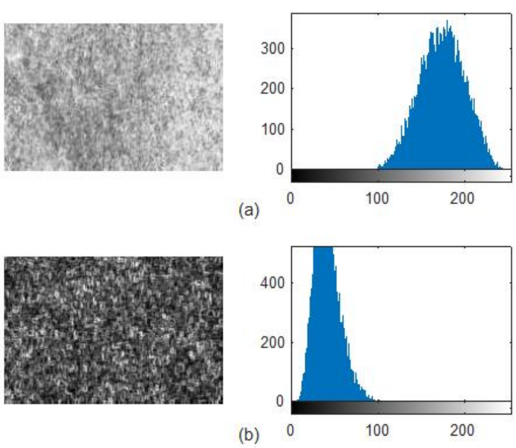


Fig. 5. Range Filter response for Normal Image (a) Input and Histogram (b) Output and Histogram.

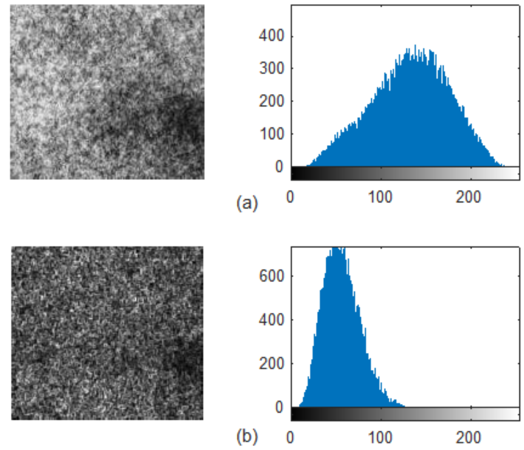


Fig. 6. Range Filter response for Abnormal Image (a) Input and Histogram (b) Output and Histogram.

B. PHOG

By extracting pyramid histograms of oriented gradients (PHOG) from the image are useful to discriminate between normal and abnormal texture features. It was first presented in [25]. Using HOG, a range filter response image can be divided into small cells, HOGs for each cell are computed, normalized with block patterns, and each cell is given as a descriptor [26]. HOG features describe the outline and local structure appearance in the image by describing their distribution of intensity gradients. Using this method, intensity invariance is maintained by counting the appearance of gradient orientations. A descriptor is generated in four main steps:

- Computation of gradients and orientation.
- Bin Orientation.
- Block descriptors.
- Normalization of blocks.

1) Computation of gradients and orientation: In the first step, when an image sample is provided, $f(i,j)$, then the magnitude of the gradient, $M(i,j)$, and angle, $\theta(i,j)$, can be obtained using (2) and (3) by considering the pixel difference.

$$M(i,j)=\sqrt{\left(f(i+1,j)-f(i-1,j)\right)^2+\left(f(i,j+1)-f(i,j-1)\right)^2} \quad (2)$$

$$\theta(i,j)=tan^{-1}\left(\frac{f(i,j+1)-f(i,j-1)}{f(i+1,j)-f(i-1,j)}\right)$$

(3)

Histograms of orientations generated by gradients of samples within a region around the centered point are called orientation histograms. Histograms of orientations are divided into eight bins, each covering 360 degrees. In the histogram, samples are weighted according to their gradient magnitude.

2) Bin orientation: In the second step of the calculation, the cell histogram is created. The algorithm assigns a weighted vote to each pixel within each cell depending on the values found in the gradient calculation and its orientation. Depending on whether the magnitude gradient and its orientations from 0° to 360°, the histogram evenly distributed over 0° to 180° or 0° to 360°. The optimum number of bins in the histogram are considered as 8 bins as mentioned in [27]. Various bin number will be tested in this paper experimentally.

3) Block descriptors: For gradient strength to be normalized locally under different illumination conditions, cells must be organized into larger with the spatially connected blocks. HOG descriptors are then constructed as concatenated vectors of corresponding normalized cell histogram components compiled from all block regions. Each cell contributes to the final descriptor more than once if these blocks overlap. Block geometries consist of two main types: the rectangular R-HOG block and the circular C-HOG block. There are three parameters that characterize R-HOG blocks: the number of cells per block, the number of pixels per cell, and channel per cell histogram. Each R-HOG block is computed at one scale, but with an orientation not aligned. As well as encoding spatial form of information, R-HOG blocks are used. There are two types of circular HOG blocks (C-HOG): those with a single central cell and those with angularly divided central cells. The C-HOG blocks can be defined by four other factors: center bin radius, the additional bins radius expansion factor, and the number of radial bins. Shape context descriptors are similar to C-HOG blocks in appearance, but they differ greatly because they contain multiple orientation, whereas single edge presence count are used to formulate the shape contexts.

4) Normalization of blocks: Gradient magnitude varies greatly over a wide range due to variations in pixel intensity on a local level [28]. Therefore, effective local contrast normalization is essential for good performance. They typically involve separately normalizing each block of cells after grouping them into blocks and normalizing of each block is obtained by using L1-norm and L2 –norm illustrated in (4) and (6). L1- norm is defined as the sum of the vector's absolute values and L2 -norm as a square root of the sum of all squared vector values. Consider a non-normalized vector v_b which contains all histograms in a given block, $\| v_b \|_k$ be its k-norm of the vector for $k=1$ (L1-norm) and $k=2$ (L2-norm) and e be the normalization constants should be small to prevent division by zero.

$$v_{b_{L1}} = \frac{v_b}{\| v_b \|_1 + e}$$

(4)

Square root of $v_{b_{L1}}$ is described in (5), by using the Bhattacharya distance between pixels, descriptor vectors can be represented as a probability distributions.

Square root of $v_{b_{L1}} = \sqrt{\frac{v_b}{\| v_b \|_1 + e}}$

(5)

$$v_{b_{L2}} = \frac{v_b}{\sqrt{\| v_b \|_2^2 + e^2}} \quad (6)$$

L2- hys: By attenuating the maximum pixel values of v_b = 0.2. Renormalizing, hysteresis thresholding can be achieved for L2 norms [29]. Renormalize the univariate feature vector so the values in each are no larger than 0.2. This will limit the effect of large gradient magnitudes. As a result, it is less important to match the magnitudes for large gradients, while the distribution of orientations is more important.

Determine the HOG for Level-0 that indicates for entire image as shown in Fig. 7(a). Throughout the entire HOG operation, $N = 8$ bins are fixed. Each bin in the histogram represents how many pixels fall within a certain range of angles. In the next step, divide the image into four blocks, and compute HOG for each block as shown in Fig. 7(b). Repeat the process recursively until the depth $L = 3$ as shown in Fig. 7(c) and (d). Several levels of pyramids are used to divide the input image, histogram composed of N bins, and Level-0 is characterized by a N - vector. An image with Level-1 is described as a $4 \times N$-vector. The PHOG descriptor represented using (7).

$$D_{PHOG} = N \sum_{l=0}^{3} 4^l$$

(7)

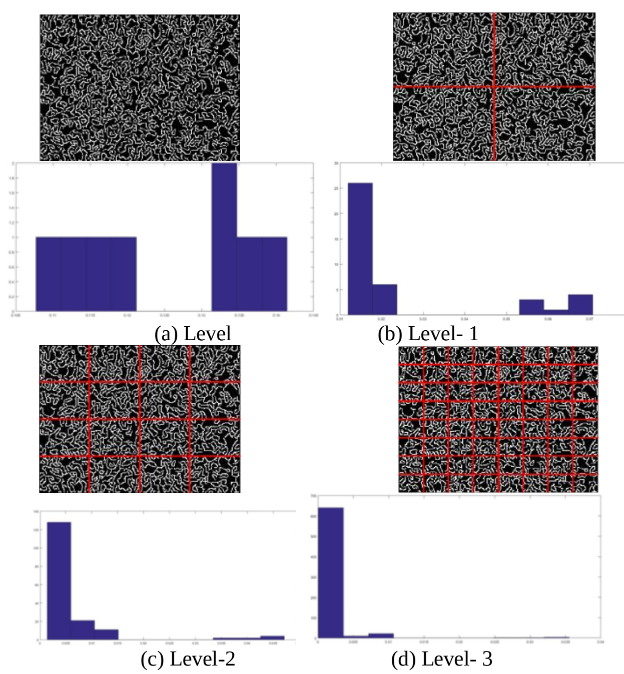


Fig. 7. PHOG Schematics at different Resolution Levels.

Create one vector by concatenating the histograms with the same level. The final PHOG descriptor of an image is obtained by concatenating all the vectors in each pyramid level that pertain to the spatial properties of an image. By setting N = 8 and $L = 3$, the PHOG descriptor (D_{PHOG}) has a length of $8*(1+4+16+64) = 680$.

VI. Classifier

Data can also be classified into identifiable groups or features using neural networks [30]. A neural network based classification becomes a very powerful prediction network. It achieves two goals: Lowering the prediction error by training and testing similar patterns of behavior; and minimizing the number of training samples within a training data set, resulting in more efficient training.

A. Neural Networks with L2 Regularization

A technique known as L2 regularization of neural networks (NNs) reduces the likelihood of the model overfitting, thus improving its performance. NNs are mathematical functions used to train the network and make predictions. The training refers to finding values for weights and biases needed to define the NN. Excessive training of a neural network may lead to model overfitting. Thus, when a neural network model is trained to predict the values from used training data, it does so with little error and high accuracy, but when the model is applied to new, previously unknown data, it does not work well. Parameter optimization is used to determine weights and biases in a way that minimizes error between computed and true results.

1) L2-Regularization: During L2 regularization, a new regularization term was added to the loss function of the neural network [31]. As part of regularization, the proposed work utilizes Euclidean-Norm or L2-Norm for the weight matrices, which is the sum of all the squared weight values of any given matrix. By dividing the scalar regularization parameter λ by two, add the regularization term to the regular loss function. The following steps are used to implement the neural network with L2 regularization.

2) Activation function: The logistic regression [32] hypothesis is defined as in (8)

$$h_\theta(x) = g(z);\ z = \theta^T \times x$$

(8)

θ – weights and x -input

The sigmoid function g is defined as in (9).

$$g(z) = \frac{1}{1+e^{-Z}}$$

(9)

3) Cost function: Probabilities of an observation can be predicted with the cost for the observation, with a minimum amount of error. A given observation x has a cost (error) associated with it if the class label is z can be defined in (10).

$$Cost(h_\theta(x), z) = \begin{cases} -\log(h_\theta(x)) & ;if\ \ z=1 \\ -\log(1-h_\theta(x)) & ;if\ \ z=0 \end{cases} \tag{10}$$

As a result, the total cost for all the m observations in a dataset is given in (11).

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} Cost\left(h_\theta(x^{(i)}), z^{(i)}\right) \tag{11}$$

$\left(x^{(1)}, z^{(1)}\right), \left(x^{(2)}, z^{(2)}\right), \ldots\ldots, \left(x^{(m)}, z^{(m)}\right)$ are the m number of training datasets.

4) Cost function and gradient: The cost function in logistic regression is defined in (12).

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m} z^{(i)} \log\left(h_\theta(x^{(i)})\right) + \left(1-z^{(i)}\right)\log\left(1-h_\theta(x^{(i)})\right)\right] \tag{12}$$

Regularization L2 helps to avoid over fitting in the model by adding a penalizing component for high weights. Equation (13) gives the regularized cost function for logistic regression.

$$J(\theta)_{new} = \frac{1}{m}\sum_{i=1}^{m} Cost\left(h_\theta(x^{(i)}), z^{(i)}\right) + \left(\sum_{j=1}^{n} || \theta_j ||^2\right)\frac{\lambda}{2m} \tag{13}$$

- $\left(\sum_{j=1}^{n} || \theta_j ||^2\right)\frac{\lambda}{2m}$ is the regularization term.
- λ is the regularization factor.
- $\frac{1}{m}\sum_{i=1}^{m} Cost\left(h_\theta(x^{(i)}), z^{(i)}\right)$ loss function.

The gradient of the cost function is a vector where the i[th] element is defined as follows:

$$\frac{\partial J(\theta)}{\partial \theta_{(i)}} = \frac{1}{m}\left[\sum_{j=1}^{m}\left(h_\theta(x^{(j)}) - z^{(j)}\right) x_i^{(j)}\right] \ ;\ for\ i = 0$$

$$\frac{\partial J(\theta)}{\partial \theta_{(i)}} = \frac{1}{m}\left[\sum_{j=1}^{m}\left(h_\theta(x^{(j)}) - z^{(j)}\right) x_i^{(j)} + \lambda\theta_{(i)}\right] \ ;\ for\ i \geq 1$$

5) Learning parameter optimization: In order to efficiently train NN, gradient descent can be implemented to determine the optimal parameters for a regression model [33]. During training and testing, the accuracy of the algorithm was maximized, while the computation time was reduced through parametric λ, θ optimization and number of iterations. Calculate the optimal cost function J (θ) according to parameters θ. Determine the optimal parameters θ for the logistic regression cost function based on a fixed dataset of x and z values. For a given dataset (x, z), computes the logistic regression cost and gradient based on a training data and a specific value of θ. Lastly, it gives the cost and θ values. After the trained data is plotted, the decision boundary will be based on the final θ value as shown in Fig. 8.

6) Prediction: A NN's hidden layers make it better suited for predictive analytics. Predictions are based on the output and input nodes of linear regression models. Consequently, they are cost-prohibitive because they require enormous computing power [34]. Furthermore, neural networks perform best when trained with extremely large data sets. By using supervised learning techniques such as data clustering, each neuron's weight can be determined. In the supervision training process, by feeding sample inputs and outputs to the algorithm, weights are derived until the inputs and outputs are closely matched, that is, weights are optimized. Using a sigmoid activation function on the output layer, it is easy to make predictions for the input test, so predicted label will be either 0 or 1.

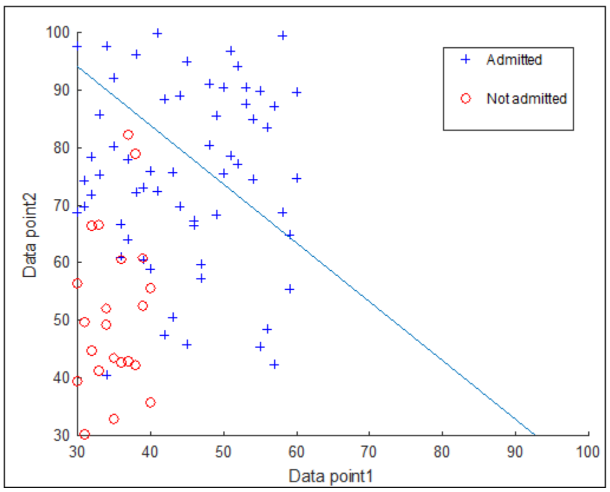


Fig. 8. Trained Data with Decision Boundary.

VII. Result and Discussion

An experimental result is carried out in three stages: *A)* preprocessing *B)* feature extraction *C)* classification.

A. Preprocessing Stage

There are 80 X-ray lumbar spine images collected; 80 x 4 =320, vertebrae of L1-L4 are the actual ROI images, but in reality only 76 ROI images have been correctly segmented (76 x 4 =304), where all four segmentations (L1-L4) are correct in 76 images. Table II shows the true and false segments of L1-L4. Out of 80 images, 77 reflect true segmentation of L1, while 3 reflect false segmentation. The results for L2 were 78 true and 2 false, L3 was 77 true and 3 false, and L4 was 76 true and 4 false.

As shown in Table III, the data base can be used to detect osteoporosis in the lumbar vertebrae L1 - L4. A total of 76 correctly segmented images are considered for the processing (76 x 4 = 304). For testing the system, a 2-fold cross validation is applied. This is accomplished by using 152 samples for supervised samples and 152 samples for testing.

B. Feature Extraction Stage

The texture features were extracted from 304(L1-L4) samples of ROI images using PHOG. For N=8 bins, Table IV shows the levels and the corresponding descriptors. Concatenated level descriptors are used to classify samples based on their attributes.

C. Classification Stage

Table V shows the osteoporosis detection classification accuracy based on the concatenation of all the levels descriptor are 696 with parameter optimization lambda varying from 0 to 1. If lambda is greater than 1, the system takes more iteration and becomes more complicated. Consequently, it was determined that for classification accuracy, when lambda is higher, the likelihood of over fitting during training is reduced, and gives maximum accuracy. During training, the model was iterated 27 times, with $\lambda = 0.9$. In large-scale implementations, a small change in the parameter can produce more effects. By taking the 184 and 40 descriptor with parameter optimization lambda varying from 0 to 1, the classification accuracy is reduced by reducing the size of the descriptor. Fig. 9 displays the graphical presentation of the system's osteoporosis detection classification accuracy with lambda. In order to improve the accuracy, lambda plays a vital role in terms of number of iterations and computation time.

TABLE II. Segmentation of L1-L4

ROI Segmentation	L1	L2	L3	L4
True	77	78	77	76
False	3	2	3	4

TABLE III. Details of the Experimental Data Base

Subjects	No. of Subjects (76)	No. of ROI samples (L1-L4=304)	DXA reports	No. of supervised samples (152)	No. of testing samples (2-fold =152)
Controlled	18	72 (18×4)	18	36	36
Osteoporotic	58	232 (58×4)	58	116	116

TABLE IV. Details of the Levels and Descriptors

No. of levels	No. of descriptor	Concatenated levels	No. of descriptor
L0	8	-	-
L1	32	L0+L1	40
L2	144	L0+L1+L2	184
L3	512	L0+L1+L2+L3	696

TABLE V. System Accuracy with Parameter Optimization

Lambda values	Instances of iteration	Accuracy in Training set (%)	Accuracy of Test sets (%)	Computed time (in seconds)
0.1	25	96.89	99.01	19
0.3	25	97.45	99.05	21
0.6	26	98.56	99.12	22
0.7	27	100	99.34	25
0.9	27	100	99.34	25

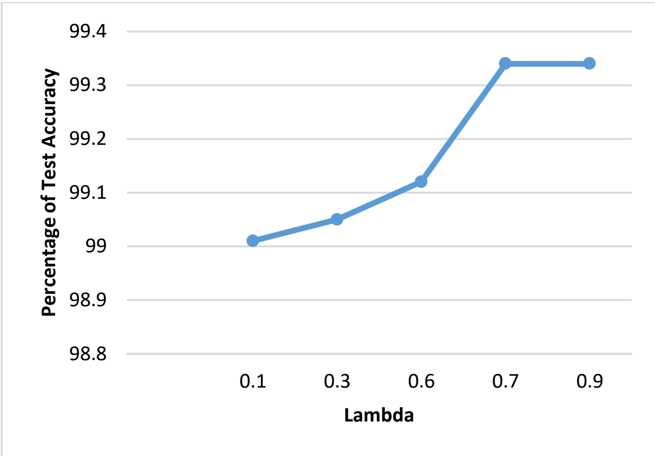


Fig. 9. Graphical Representation of Accuracy with Lambda.

A confusion matrix with multiple classifiers can be used to evaluate system performance as shown in Table VI. Testing the proposed system is based on a detailed analysis of the database, as shown in Table III. With L2 regularization, NN has fewer False Positives (FP) and False Negatives (FN), which leads to less overfitting and more accuracy, flexibility, and works well in classifying events. When compared with other classification techniques, the overall performance of the system leads to better True Positives (TP) and True Negatives (TN). Based on NN with L2 regularization, a 2-fold cross-validation gives good results for L1 through L4. Fig. 10 illustrates 2-fold cross validation using different classifiers for system assessment. Thus, NN with L2 Regularization can produce the smallest FP and FN, resulting in a more accurate system. Using 2- fold cross-validation for the lumbar spines L1-L4, Table VII describes the systems average performance using the various classification methods. The system is tested for 2-fold cross validation for descriptor 696 with lambda 0.9. Sensitivity (True positive Rate-TPR) is the percentage of actual positive results that were correctly predicted. The specificity indicates how many negatives were correctly predicted out of actual negatives. Probability of successful positive predictions is measured by precision [35,36]. The recall measures the proportion of true positives for which the prediction was correct. Precision and recall are expressed as the harmonic mean, this method accounts for both false positives as well as false negatives which is termed as F1-score [37]. NPV (Negative Predictive Value) is the probability that someone will not have the specific disease, even after testing negative for it. According to all the above mentioned measures the NN with L2 regularization techniques attain the best classifications and overall the system measures the trueness of the test samples accurately.

Fig. 11 shows how system performance can be represented from a graphical point of view using the Receiver Operating Characteristic (ROC) curve. The red line represents the area under (AU) ROC. Essentially all of the images of test samples are classified into abnormal and normal images based on the curves occupying 99 percent of the region. This diagonal line shows an average AUROC value of 0.5 is found in ROC curves of random predictors. In order to determine if the model is useful, the random predictor is often used. ROC curves plotted in terms of sensitivity (TPR) versus 1-specificity (false positive rate). This model gives better classification against false positive and true positive rates since the AUC approaches 1, that is it is more accurate. Based on the comparison of different methods in Table VIII, the proposed approach is found to be better than those adopted by the other methods in terms of accuracy, sensitivity, specificity, and F1-score. Because, the overall design will enable more accurate classification methods of osteoporosis detection with the help of preprocessing and significant image texture feature extraction using PHOG.

TABLE VI. EVALUATION OF DIFFERENT CLASSIFIERS

Classifiers	TP	FP	TN	FN
Decision Tree (DT)	113	3	34	2
Naïve Bayes (NB)	108	5	32	7
Support Vector Machine (SVM)	102	9	30	11
K-Nearest Neighbor (KNN)	110	6	32	4
Neural Network with L2 Regularization	115	1	36	0

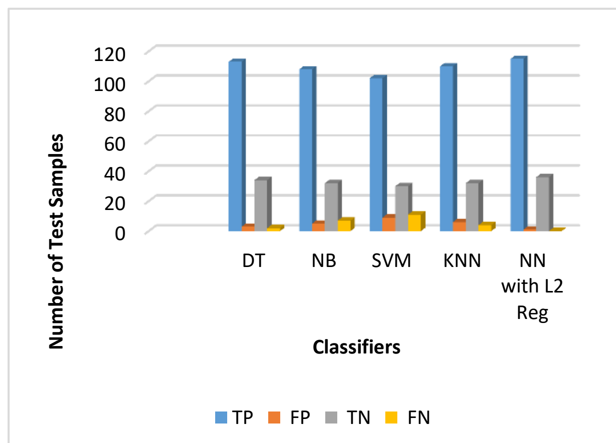


Fig. 10. An Evaluation System represented Graphically.

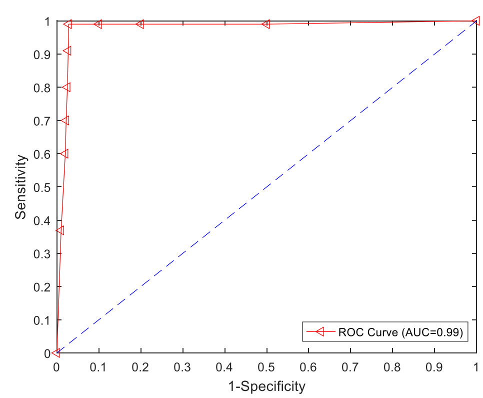


Fig. 11. Analyzing Classifications with AUROC.

TABLE VII. PERFORMANCE OF DIFFERENT CLASSIFIERS

Classifiers	Sensitivity (%)	Specificity (%)	F1-Score (%)	PPV (%)	NPV (%)	Accuracy (%)
Decision Tree (DT)	98.26	91.89	97.83	97.81	94.44	96.71
Naïve Bayes (NB)	93.81	86.48	94.73	95.57	82.05	92.10
Support Vector Machine (SVM)	90.26	76.92	91.06	91.89	73.17	86.84
K-Nearest Neighbor (KNN)	96.49	84.21	95.61	94.82	88.88	93.42
Neural Network with L2 Regularization	**100**	**97.29**	**99.56**	**99.13**	**100**	**99.34**

TABLE VIII. PERFORMANCE COMPARISON WITH DIFFERENT METHODS

Techniques and features	Image modalities	Classifiers	Cross -validation	Specificity (%)	Sensitivity (%)	Accuracy (%)
Fractal Analysis [4]	CT	k-NN	3-fold	90	78	81
Morphometric [7]	CT	k-NN	10-fold	83.3	92.5	88.3
3D feature with vBMD [8]	CT	Random forest	10-fold	78	77	
CNN[9]	CT	BCC- net MS-net	7- fold 2- fold	-	-	76.65
OSTA[10]	Person attribute values	ANN SVM RF KNN LoR	OSTA[10]	64.6 54.7 70 72.9 53.3	91.7 91.7 87.5 83.3 95.8	-
Proposed	**X-ray and DXA Report**	**NN with L2 regularization**	**2- fold**	**97.29**	**100**	**99.34**

VIII. CONCLUSION

Millions of people worldwide suffer from osteoporosis, especially those who are aging. The purpose of this study is to reduce risks of bone fracture by exploiting the effectiveness of osteoporosis with the help of X-ray images. The aim of this study is to reduce the bone fracture risks by exploring the effectiveness of osteoporosis. In this research work, preprocessing step allows experimentally extracting the texture features of trabecular bone structure, allowing visualization of fine texture features of internal trabecular bone. An analysis of texture features is helpful for identifying the significant features that allow the images to be classified. By using these texture features, NN with L2 regularization achieves better system performance in regards to classifying the normal and osteoporotic L1-L4 images. The contribution of this paper is to understand the diagnose of osteoporosis in lumbar vertebrae L1-L4 of spine X-ray image more accurately so that risk factors can be avoided with less cost.

IX. FUTURE WORK

To improve the diagnostic accuracy with the help of spine X-ray image of lumbar vertebrae, it is necessary to make a standard dataset that can be used universally by researchers so that one can set the standard analysis of trabecular bone micro architecture to reduce error classification. A reliable and cost-effective system is needed to overcome the drawbacks of DXA, such as lack of measurement of soft bone (trabecular) and set the standard benchmark to analyze BMD. For the bone density to be calculated, a machine learning algorithms should combine with larger number of clinical reports and X-ray imaging data in order to avoid bone risk factor.

ACKNOWLEDGMENT

The authors would like to thank Dr. A Ramalingaiah M.S (ortho) Orthopedic Doctor, Bangalore. No.271 Abhilasha Orthopedic Hospital 5th Block 100ft Rd Banashankari 3rd stage Bangalore - 560085, India for providing X-ray image of spine (raw-data) and DXA report. We sincerely acknowledge and are grateful for their guidance and suggestions.

REFERENCES

[1] Riggs, B. Lawrence. "Overview of osteoporosis." Western journal of medicine 154, no. 1,63,1991.

[2] On Osteoporosis, NIH Consensus Development Panel, and Diagnosis Prevention, "Osteoporosis prevention, diagnosis, and therapy," JAMA 285, no. 6, pp. 785-795, Feb. 2001.

[3] H. P. Dimai, R. Ljuhar, D. Ljuhar, B. Norman, S. Nehrer, A. Kurth, and A. Fahrleitner-Pammer, "Assessing the effects of long-term osteoporosistreatment by using conventional spine radiographs: results from a pilotstudy in a sub-cohort of a large randomized controlled trial," Skeletalradiology 48, Vol. 7, pp.1023-1032, Dec.2019.

[4] HY. Park, JY. Ha, KW. Kim, I. H. Baek, S. B. Park, and J. S. Lee, "Effect of lumbar spinal stenosis on bone mineral density in osteoporosis patients treated with ibandronate," BMC Musculoskelet Disord 22, Vol.1, pp.1-7, May 2021.

[5] Z. Omiotek, R. Dzierżak, and S. Uhlig, "Fractal analysis of the computed tomography images of vertebrae on the thoraco-lumbar region in diagnosing osteoporotic bone damage," In: Proc.of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, Vol. 233, No. 12, pp. 1269-1281, Oct. 2019.

[6] K. H. Nam, I. Seo, D. H. Kim, J. I. Lee, B. K. Choi, and I. H. Han, "Machine Learning Model to Predict Osteoporotic Spine with Hounsfield Units on Lumbar Computed Tomography," Journal of Korean Neurosurgical Society, Vol. 62, No. 4, pp. 442-449, Jul. 2019.

[7] Patil, Kavita Avinash, and KV Mahendra Prashanth, "Texture Feature Extraction of CTB Radiograph Image Using Derivative Gaussian filter With NN Classification to Diagnose Osteoporosis," International Journal of Innovative Technology and Exploring Engineering, Vol.8, June 2019.

[8] A. S. Shaker, “Detection and Segmentation of Osteoporosis in Human Body using Recurrent Neural Network,” International Journal of Advanced Science and Technology, Vol. 29, No. 2, pp. 1055-1066, Jan. 2020.

[9] H. Yousefi, E. Salehi, O. S. Sheyjani, and H. Ghanaatti, “Lumbar Spine Vertebral Compression Fracture Case Diagnosis Using Machine Learning Methods on CT images”, In: Proc. of 4th International Conference on Pattern Recognition and Image Analysis (IPRIA) IEEE, Tehran, Iran, pp. 179-184, Mar. 2019.

[10] Patil, Kavita Avinash, KV Mahendra Prashanth, and A. Ramalingaiah. "Osteoporosis Detection in Lumbar spine L1-L4 based on Trabecular Bone Texture Features." International Journal of Intelligent Engineering and Systems 14, no. 6 (2021): 80-94.

[11] A. Valentinitsch, S. Trebeschi, J. Kaesmacher, C. Lorenz, M. T. Löffler, C. Zimmer, T. Baum, and J. S. Kirschke, “Opportunistic osteoporosis screening in multi-detector CT images via local classification of textures,” Osteoporosis international, Vol. 30, No. 6, pp. 1275-1285, Jun. 2019.

[12] C. Tang, W. Zhang, H. Li, L. Li, Z. Li, A. Cai, L. Wang, D. Shi, and B. Yan, “CNN-based automatic detection of bone conditions via diagnostic CT images for osteoporosis screening,” arXiv preprint arXiv:1910.06777, Oct. 2019.

[13] WY Ou Yang, CC Lai, MT Tsou, and LC Hwang, “Development of Machine Learning Models for Prediction of Osteoporosis from Clinical Health Examination Data,” International journal of environmental research and public health18, Vol no.14, pp. 7635, Jan. 2021.

[14] RC Gonzalez, "Richard E. woods", Digital image processing, Vol. 2, pp. 550-570, 2002.

[15] DJ Bora, “Importance of image enhancement techniques in color image segmentation: A comprehensive and comparative study,” arXiv preprint arXiv:1708.05081, Aug. 2017.

[16] Mishra, Akshansh. "Contrast Limited Adaptive Histogram Equalization (CLAHE) Approach for Enhancement of the Microstructures of Friction Stir Welded Joints." arXiv preprint arXiv:2109.00886 ,2021.

[17] L Tang, Y Li, T Ling, L He, T Luo, and D. Guo, “UAV image registration algorithm based on overlapping region detrction,” In 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Vol.10, pp. 6706-6709, Jul. 2016.

[18] A Plyer, E Colin-Koeniguer, and Weissgerber F, “A new coregistration algorithm for recent applications on urban SAR images,” IEEE Geoscience and Remote Sensing Letters 12, Vol.11, pp. 2198-202, Aug. 2015.

[19] G Brigot, E Colin-Koeniguer, A Plyer, and Janez F, “Adaptation and evaluation of an optical flow method applied to coregistration of forest remote sensing images,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9, Vol. 7, pp. 2923-39, Jul. 2016.

[20] F Champagnat, A Plyer, G Le Besnerais, B Leclaire, and Y Le Sant, “How to calculate dense piv vector fields at video rate,” In Proceedings of 8th International Symposium on Particle Image Velocimetry-PIV09, Vol. 11, pp. 15-20, Aug. 2009.

[21] A Plyer, G Le Besnerais, F Champagnat, “Massively parallel Lucas Kanade optical flow for real-time video processing applications,” Journal of Real-Time Image Processing 11, Vol. 4, pp. 713-30. Apr. 2016.

[22] Frost BA, Camarero-Espinosa S, Foster EJ, “Materials for the Spine: Anatomy, Problems, and Solutions,” Materials (Basel).12(2):253,PMID: 30646556,Jan 14 2019.

[23] Patil, Kavita Avinash, and KV Mahendra Prashant. "Segmentation of Lumbar [L1-L4] AP Spine X-ray images using various Level Set methods to detect Osteoporosis." In 2021 IEEE Bombay Section Signature Conference (IBSSC), pp. 1-6. IEEE, 2021.

[24] Bailey, Donald G., and Robert M. Hodgson, "Range filters: Local intensity subrange filters and their properties," Image and Vision Computing 3, no. 3 ,pp: 99-110, 1985.

[25] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, pp. 886-893, CVPR.2005.177, June 2005.

[26] OL Junior, D Delgado, V Gonçalves, and U Nunes, “Trainable classifier-fusion schemes: An application to pedestrian detection,” In 2009 12th international IEEE conference on intelligent transportation systems 4, pp. 1-6, Oct. 2009.

[27] A Bosch, A Zisserman, and X Munoz, “Representing shape with a spatial pyramid kernel,” In Proceedings of the 6th ACM international conference on Image and video retrieval, pp. 401-408, Jul. 2007.

[28] EB Sudderth, A Torralba, WT Freeman, and AS Willsky, “Learning hierarchical models of scenes, objects, and parts,” In Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, Vol. 2, pp. 1331-1338, Oct. 2005.

[29] DG Lowe, “Distinctive image features from scale-invariant key points,” International journal of computer vision 60, Vol. 2, pp. 91-110, Nov. 2004.

[30] DR Baughman, and YA Liu, “Classification: fault diagnosis and feature categorization,” In Neural Networks in Bioprocessing and Chemical Engineering, pp. 110-171, Academic Press 1995.

[31] AY Ng, “Feature selection, L 1 vs. L 2 regularization, and rotational invariance,” In Proceedings of the twenty-first international conference on Machine learning, p. 78, Jul. 2004.

[32] Demir-Kavuk, O., Kamada, M., Akutsu, T., & Knapp, E. W, “ Prediction using step-wise L1, L2 regularization and feature selection for small data sets with large number of features ,” BMC bioinformatics, 12, 412, 25 Oct 2011.

[33] N Jain, “Optimization of Regularization and Early Stopping to Reduce Overfitting in Recognition of Handwritten Characters,” In Congress on Intelligent Systems, pp. 305-323, Sep. 2020.

[34] Iftikhar, Asim, Muhammad Alam, Rizwan Ahmed, Shahrulniza Musa, and Mazliham Mohd Su’ud, "Risk Prediction by Using Artificial Neural Network in Global Software Development," Computational Intelligence and Neuroscience 2021 (2021).

[35] Patil, Kavita Avinash, KV Mahendra Prashanth, and A. Ramalingaiah. "Texture Feature Extraction of Lumbar Spine Trabecular Bone Radiograph Image using Laplacian of Gaussian Filter with KNN Classification to Diagnose Osteoporosis." In Journal of Physics: Conference Series, vol. 2070, no. 1, p. 012137. IOP Publishing, 2021.

[36] Powers, David MW, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," arXiv preprint arXiv:2010.16061 ,2020.

[37] Townsend, James T.,"Theoretical analysis of an alphabetic confusion matrix," Perception & Psychophysics 9, no. 1 , pp. 40-50, 1971.

Deep Learning Approach for Spoken Digit Recognition in Gujarati Language

Jinal H. Tailor[1], Rajnish Rakholia[2], Jatinderkumar R. Saini[3]*, Ketan Kotecha[4]
MCA, S. S. Agrawal Institute of Management & Technology, Navsari, India[1, 2]
Symbiosis Institute of Computer Studies and Research, Symbiosis International (Deemed University), Pune, India[3]
Symbiosis Centre for Applied Artificial Intelligence, Symbiosis International (Deemed University), Pune, India[4]

***Abstract*—Speech Recognition is an emerging field in the area of Natural Language Processing which provides ease for human machine interaction with speech. Speech recognition for digits is useful for numbers oriented communication such as mobile number, scores, account number, registration code, social security code etc. This research paper seeks to achieve recognition of ten Gujarati digits from zero to nine (૦ to ૯) by using a deep learning approach. Dataset is generated with total 8 native speakers 4 male 4 female with the age group of 20 to 40. The dataset includes 2400 labeled audio clips of both genders. To implement a deep learning approach, Convolutional Neural Network (CNN) with MFCC is used to analyze audio clips to generate spectrograms. For the proposed approach three different experiments were performed with different dataset sizes as 1200, 1800 and 2400. With this approach maximum 98.7% accuracy is achieved for spoken digits in Gujarati language with 98% Precision and 98% Recall. It is analyzed from various experiments that increase in dataset size improves the accuracy rate for spoken digit recognition. No of epochs in CNN also improves accuracy to some extent.**

Keywords—CNN; Deep learning; digit; Gujarati; speech recognition

I. Introduction

Speech is the most common medium used by humans for communication. Humans express their thoughts in the form of speech. Speech recognition is the emerging field in the area of Natural Language Processing [1]. Performance and Accuracy of Speech recognition process depends upon the speaker's mental state, emotions, accent, style and ease of speaking and environment. Human speech includes various parameters such as words, sentences, digits and fillers. Speaker feels more comfortable when he is communicating in his native language. In general, people have good command and control over their native language as they have learnt it from their childhood. Speech recognition with native language provides better performance and accuracy with the use of algorithms [2]. Generally the speech recognition process focuses upon identification of words and sentences. Very little work has been done by researchers regarding digits recognition in Gujarati language. Digits recognition with deep learning includes zero to nine numbers identification using optimized algorithms.

Gujarati is Indo-Aryan language and used as native language for people living in Gujarat state in India. It is spoken by 55.5 million people in India. Gujarati language includes 34 consonants, 12 vowels, 10 digits and 106 special symbols. Speech recognition related to the Gujarati language is a challenging task due to morphological structure, language barrier, dialects and limited resources [3]. This research paper mainly focuses upon digit recognition in Gujarati language using deep learning approach.

Deep learning is a subfield of machine learning that includes various methods and functions that imitates human behavior and knowledge. Deep learning algorithms apply a hierarchy of nonlinear transformation to the input values and generate statistical models as output that can predict results in future on their own. Deep learning approaches require large amounts of data for training and require more computation resources [4]. There are several forms of neural networks such as Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), Artificial Neural Network (ANN) and Feedforward Neural Network (FNN). Most popular deep neural network is Convolutional Neural Network (CNN) to understand and analyze data in visual form.

In the proposed approach for digit recognition for Gujarati language CNN model was implemented. For the analysis process, dataset was prepared with 8 speakers with 2400 labeled audio clips. Different dataset sizes are used as input to CNN to analyze the impact on accuracy.

II. Literature Review

Gunawan, A. (2010) [5] developed a digit recognition system from zero through nine using HMM and MATLAB for GUI. Spoken digits were in an isolated structure with 34 male and female speakers. They have created two different modules for isolated and continuous speech recognition. To check effect of variation of environment on accuracy they have selected noisy and clean environment. In a clean environment, in multi-speaker mode for isolated and continuous speech average accuracy found is 96%. For speaker independent mode, average accuracy is 78%. In noisy environments, in multi-speaker mode for isolated and continuous speech average accuracy found is 80%. For speaker independent mode average accuracy is 61.6%. They have concluded that recognition accuracy depends upon some factors such as noise rate and microphone quality.

Chapaneri and Jayaswal (2013) [6] have proposed a speaker-independent system for isolated digits using WMFCC. They have used SOLA based techniques and IFDTW to reduce time complexity for recognition. The overall accuracy of the proposed system for recognition 0 to 9 was 99.16%. They have

*Corresponding Author.

concluded that the system performs 22 times faster with higher accuracy and lower computational overhead achieved using WMFCC and FIFDTW.

Datta et al., (2021) [7] have done comparative analysis for methods to recognize zero to nine digits. For comparison they have analyzed cepstral and MFCC. They have concluded that the feature extraction method MFCC performed with better results compared to cepstral in command based systems.

Kurian, C., & Balakrishnan (2013) [8] have developed speaker-independent systems for connected digits for Malayalam language using Perceptual Linear Predictive (PLP) cepstral coefficient and Hidden Markov Model (HMM). For the training phase they have recorded 20 sets of continuous digits of 21 speakers from the age group of 20 to 40 years in an office environment. For the proposed system they have achieved 99.5% recognition accuracy by averaging word accuracy.

Sen O. et al. (2021) [9] created a convolutional neural network model to recognize Bangla spoken digits from 0 to 9. They have used Mel Frequency Cepstrum Coefficient (MFCC) for feature extraction and Convolutional Neural Network (CNN) for digit recognition. They have developed a proposed model using python,keras and tensorflow for 4000 spoken occurrences with parameters such as gender, accent and age. Measured accuracy with 10 fold cross validation was 96.7%.

Zerari N. et al. (2019) [10] proposed a framework for digit recognition using neural network. For feature extraction they have used Mel Frequency Cepstral Coefficient (MFCC) and Filter Banks (FB) coefficients. They have applied a proposed model on two different datasets one for spoken digits and another for spoken TV commands for different age groups for a total of 88 Arabic native speakers from which 44 were male and 44 were female. Total first digit database entries were 8800 tokens and the second TV command database includes 10000 instances. It is analyzed that the delta-delta feature used to characterize speech recognition signals gives better accuracy over 96% compared to MFCC and FB.

Dhanashri D. & Dhonde S. (2017) [11] have proposed isolated word speech recognition using HMM and DNN for acoustic modeling. Dataset for digit sequences samples were collected in a quiet environment at 20 kHz with 55 men and 57 women. System accuracy is measured for three different hidden layers such as 300,400 and 500 units. Accuracy achieved for 300 hidden layer units is 83.67%, for 400 its 85.44% and for 500 units its 86.06%. They have analyzed that proposed model accuracy can be improved by increasing the number of hidden layer units.

Wazir A. & Chuah J. (2019) [4] have proposed Arabic digits speech recognition model using Recurrent Neural Network (RNN). Feature extraction is done with MFCC and extracted features are fed into the Long Short Term Memory (LSTM) network. They have collected total 1040 data points from different countries and dialects. For the training phase 840 data points were used from 1040 data points and achieved 94% with minimum loss. In the testing phase total 200 data points were included and 65% accuracy was achieved. They have found that the most recognized digit is ‘0’ with 80% accuracy and the worst recognized digit is ‘6’ with 60% accuracy.

Hamidi M. et al. (2020)[12] have presented analysis of Amazigh digit recognition with IVR system for speaker-independent systems in noisy environments. They have used open source software Asterisk 1.6, Ekiga softphone and CMU Sphinx Tools. They have selected 22 Moroccan native speakers of both genders male and female. They have concluded that minor degradation in accuracy found for SNR as low 3 and 9 lb. Major degradation was found with decoded speech signal and best recognition rate at 3 lb.

Swedia E. et al. (2018) [13] have proposed a deep learning approach for Indonesian digit recognition using LSTM. For feature extraction they have used LPC and MFCC techniques and compared their performance. They have recorded total 7990 speech digits with 12 coefficients of MFCC and 12 for LPC. They have concluded that Indonesian speech recognition gives better performance using MFCC with 96.58% compared to LPC with 93.79%.

Swastika R. et al. (2017) [14] have performed comparative analysis for four deep learning architectures such as DBN-DNN, LSTM, TDNN and CNN. For feature extraction MFCC and LPC techniques are used. They have used TIDigits dataset which includes 226 speakers from which 111 men and 114 women each uttered 77 digit sequences. Kaldi speech recognition toolkit is used for training and decoding processes. They have concluded that CNN gives the best recognition performance in a noisy and reverberant environment compared to other three architectures. There are also research instances where the authors worked for printed Gujarati characters, for instance [15]. A comparative study of Literature review is presented in Table I.

TABLE I. COMPARATIVE STUDY OF LITERATURE REVIEW

Authors	Year	Spoken Language	Methods / Techniques	Category of Tokens	Accuracy
Gunawan [5]	2010	English	HMM	Digits	96%
Chapaneri, S. V., & Jayaswal, D. J. [6]	2013	English	WMFCC FIFDTW	Isolated Digits	99.16%
Datta et al. [7]	2021	English	Cepstral MFCC Vector Quantization	Digits	MFCC VQ - 93.3 % Cepstral - 86.6 %
Kurian, C., & Balakrishnan, K. [8]	2013	Malayalam	PLP HMM	Connected Digits	99.5 %
Sen O. et al. [9]	2021	Bangla	MFCC CNN	Digits	96.7 %
Zerari N. et al. [10]	2019	Arabic	LSTM MLP	Digits	96 %
Dhanashri D. & Dhonde S. [11]	2017	English	HMM DNN	Isolated Digits	300 layer - 83.67 % 400 layer - 85.44 % 500 layer - 86.06 %
Wazir A. & Chuah J. [4]	2019	Arabic	MFCC LSTM RNN	Digits	Training - 94 % Testing - 65%
Hamidi M. et al. [12]	2020	Amazigh	CMU Sphinx	Digits	72.43 %
Swedia E. et al. [13]	2018	Indonesian	LSTM MFCC LPC	Digits	MFCC - 96.58 % LPC - 93.79 %
Swastika R. et al. [14]	2017	English	DBN - DNN LSTM CNN Kaldi	Digits	CNN with better performance in clean environment

III. METHODOLOGY

The methodology to recognize digits in Gujarati language using CNN includes 6 major steps depicted in Fig. 1.

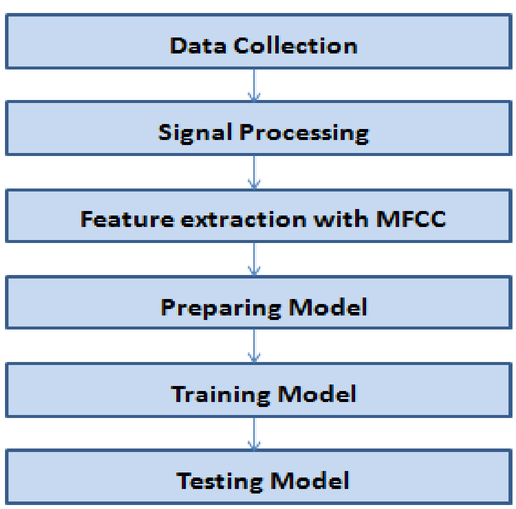


Fig. 1. Flowchart of Steps for Proposed Model.

A. Data Collection

For data collection, 8 native speakers, comprising 4 males and 4 females, of the Gujarati language were selected. All 8 speakers were between the age group of 20 to 40. For recording purposes Audacity software was used and recording is performed in a lab environment. Total 30 occurrences of 0 to 9 digits were spoken by each speaker. Dataset consists of 0 to 9 digits in Gujarati with English language representation listed in Table II. Total 2400 occurrences (8 speakers * 10 digits * 30 occurrences) were used for the proposed speech recognition model. Collected data is split in two parts: Training and Testing. For data division 80:20 ratios are followed from total collected data. From 2400 total occurrences 1920 (80% of 2400) were used for training the model and 480 (20% of 2400) occurrences used for testing purposes. All the audio files are collected and stored in wav format.

TABLE II. DATASET OF DIGITS (0 TO 9) IN GUJARATI AND ENGLISH REPRESENTATION

Digit in Gujarati	Pronunciation in Gujarati	Digit word in English	Digit in English numeric
૦	૦- શૂન્ય	0	zero
૧	૧- એક	1	one
૨	૨- બે	2	two
૩	૩- ત્રણ	3	three
૪	૪ -ચાર	4	four
૫	૫- પાંચ	5	five
૬	૬- છ	6	six
૭	૭- સાત	7	seven
૮	૮- આઠ	8	eight
૯	૯- નવ	9	nine

B. Signal Processing

Speech signals are non-stationary signal combination of voiced and unvoiced signals. From actual speech signals, it is required to extract voiced segments without silence and noise. Data segments with silence are not useful for feature analysis. To remove noise from signals, audacity has a provision with noise reduction option through which we can control by specifying how much dB value we want to consider as noise from 0 to 24 (maximum) dB values.

For processing, first we have removed noise from recorded audio files. Through Audacity noise reduction parameters can be set with residue data that is removed. It provides ease to select a voice sample window which contains useful information about speech signals. The dataset containing a total of 2400 occurrences of the 10 digits, from 0 to 9, were recorded by the speakers in Gujarati. Noise parameters are set to preferences and selected tracks are stored with reduced noise.

C. Feature Extraction with MFCC

Raw data is not used directly in the model training process because it contains noise and other unnecessary parameters that affect accuracy of the model. Feature extraction is an important step in the speech recognition process. Extracted features are used as input to the model to generate text sequences. In the model, extracted features as input, gives better performance compared to raw audio signals. To extract spectral frames from speech signals, we have used the MFCC technique which is a widely used technique for feature extraction. First audio signal is converted from analog to digital with a sampling frequency 16 kHz. To improve model performance, pre-emphasis increases the energy in higher frequency for better phone detection.

To differentiate phones from audio signals, segments are divided into window width 25ms. From each segment, we have extracted 39 features. Hamming window is used to divide signal into segments without noise in a high frequency region. By applying Discrete Fourier Transform (DFT) to convert the time domain of a signal into frequency domain for easy analysis.

By modeling human hearing property into a model at the feature extraction step can improve the performance of the model. For mapping actual frequency with human hearing frequency, mel scale parameter is used. At lower energy humans can easily identify increased levels compared to higher energy. To adapt the same human behavior in the system, log is applied to Mel-filter.

After DFT transformation, the time domain and frequency domain are inverted. The lowest frequency is the highest frequency in the time domain. So inverse transform IDFT (inverse Discrete Fourier Transform) of log of magnitude of signal is generated which is called a cepstrum. After applying IDFT, MFCC includes 12 coefficients to extract signal sample energy as features.

In the dynamic feature generation step, with 13 features, MFCC technique generates first order derivatives and second order derivatives for each feature so total 26 features generated. Derivatives are calculated with differences between coefficients of audio signals to understand transition. Calculated 39 features generated by MFCC for each audio signal that was included as input to the speech recognition system.

D. Deep Learning Approach

Deep learning includes various sub domains such as computer vision, digital signal processing, music classification and automatic speech recognition. With the popularity of virtual assistants like Alexa, Google Home and Siri, the field machine learning systems are promoted at a higher level for speech to text or action. All these systems work upon audio signal processing using deep learning concepts.

E. CNN

Neural networks include a major part as Convolutional Neural Network. CNN is made up of neurons that have learnable weights and biases. It includes various application implementations such as image processing, object identification and classification. CNN can also be used for speech recognition as it can be applied on acoustic models and define convolution layers to each overlap acoustic frames. It extracts phones, pitch, frequency and gender features from acoustic frames according to weights of the network. Convolution Layer is the first layer of CNN that is the core building block and performs major computational tasks. It is a mathematical operation that merges two sets of information. CNN includes three main aspects: locality, weight sharing and pooling. Before start processing input data it is necessary to reduce non-white noise through locality. Weight sharing is used to improve robustness of model and overfitting and reduce weights. Pooling or subsampling process reduces the size. The input image is resized to optimal level to input into the convolutional layer. CNN Architecture is shown in Fig. 2.

In CNN, a filter or kernel is available that is placed in a sliding window over some of the pixels of the image depending upon kernel size to produce feature maps in different convolutional layers. In this step, through element wise multiplication each original pixel value of the image is multiplied with values in the filter. The sum of multiplication values is generated.

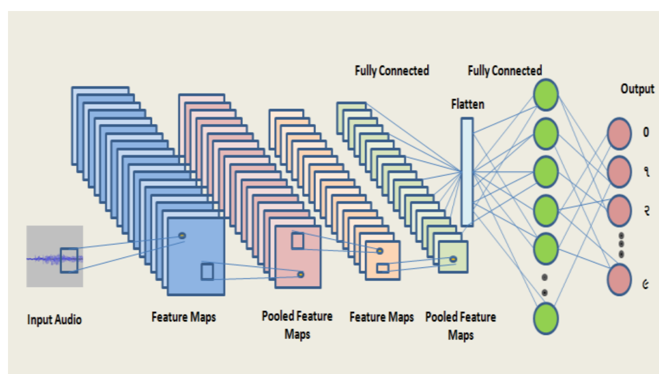


Fig. 2. Convolutional Neural Network.

If the value found is too low then there is nothing to compare in the activation map by curve detector filter. By adding more filters input data can be verified in depth of the activation map. The second layer called the Activation layer, ReLu (Rectified Linear Unit) function is applied to increase non-linearity in the CNN as images consist of objects that are not linear. In the third pooling layer, that involves downsampling of features. Common pooling layer uses 2 X 2 max filters with stride of 2 that is non-overlapping filter. Max filter returns max value in the feature from selected region. Last layer is known as a fully connected layer that pooled a full feature map matrix into a single column that was used for processing in a neural network. All the features are combined in a network through fully connected layers.

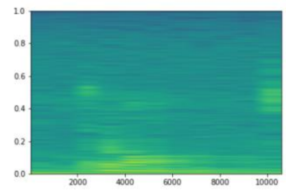
F. Preparing Model

The dataset consists of a total 2400 occurrences both from male and female speakers. It consists of 10 spoken digits recording from 0 to 9. Recording of the dataset was performed using audacity software with 22050 Hz monophonic 16-bit audio files with .wav format. Google colab is used to implement and execute neural networks to get GPU and TPU as a runtime environment. It is free and easy to use. Librosa python library is used to extract spectrograms from audio files. Total dataset is splitted into training and testing data with 80:20 ratios. From a total 2400 audio files, 1920 files were used in the training session. The algorithm is written in python with TensorFlow library using Keras open source library as python interface. Each spectrogram file is stored with PNG format. Spectrogram images of different digits in Gujarati are depicted in Fig. 3. All spectrogram images belong to classes ranging from 0 to 9. Then data is split into training and testing sets. We have used sequential types to build a layer wise model in keras.

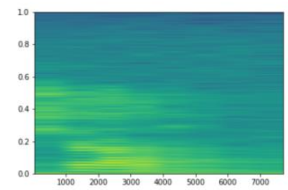
In the training phase, input images of the spectrogram are passed to a 2D convolutional layer (Conv2D) with filter size of 32. There are 2 layers as Conv2D layers. First layer consists of 64 nodes and the second layer consists of 32 nodes. With kernel size of 3, 3 X 3 filter matrix is used. For two layers activation function ReLu (Rectified Linear Activation) is used. A 'Flatten' layer is added between conv2D and dense layer to transform the resultant vector into 1 dimensional vector. Dense is the standard output layer used in neural networks. Mainly 10 different nodes were added in layers ranging from 0 to 9. Resultant vector is passed to a fully connected dense layer with a Softmax function that concludes the output in sum up to 1 to represent it in the form of probabilities. Based on the highest priority, prediction series can be generated. After setting all the parameters for the model, compilation is done with three parameters. Compiling process includes optimizer, loss and metrics. SGD (Stochastic Gradient Descent) optimizer is added from keras as it is an iterative method for optimizing objective function. The 'categorical_crossentropy' is used for loss function which indicates the model performance. Third parameter Accuracy metric is used to show accuracy of model on validation data.

G. Training Model

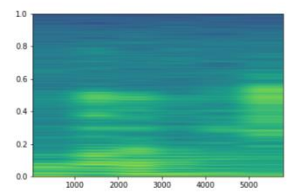
To train a model, fit () function is used with training data and validation data as test data with number of epochs. Epochs specify the total number of model cycles for data. We can get better model performance by increasing the number of epochs up to some extent. For proposed model, 500 epochs were set to analyze neural network performance.



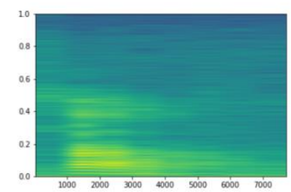
Spectrogram of digit ૩ (ત્રણ)- 3
(a)



Spectrogram of digit ૪ (ચાર) - 4
(b)



Spectrogram of digit ૫ (પાંચ) - 5
(c)



Spectrogram of digit ૭ (સાત) - 7
(d)

Fig. 3. Spectrogram of different Digits Spoken in Gujarati.

H. Testing Model

To predict the output based upon test data, predict_generator function is used. It generates an array with 10 digit numbers. Each digit range represents the digit classification according to probability.

IV. Result and Discussion

The experiment was executed with training and testing of data to create a neural network. The proposed approach includes a total 2400 occurrences of spoken digits 0 to 9 in Gujarati language. The model is developed with Python programming language with Keras and Tensorflow libraries. The model is executed three times with varied dataset sizes as 1200, 1800 and 2400 digits. The result analysis with various data sets is depicted in Table III. In the first iteration with 1200 dataset values, received accuracy was 58.6% with 58% precision and 58% recall. In the second iteration, accuracy achieved is 78% with 78% precision and 78% recall. In the third iteration with 2400 occurrences, 98.7% accuracy is achieved with 98% precision and 98% recalls.

It is analyzed that accuracy of recognition improves with dataset size. The result can be viewed in graphical form depicted in Fig. 4.

TABLE III. Result Analysis with Various Dataset Sizes

Experiment No.	**Data set size**	**Learning Rate**	**No. of Epochs**	**Batch size**	**Accuracy %**	**Precision%**	**Recall %**
1	1200	0.01	500	32	58.6	58	58
2	1800	0.01	500	32	78	78	78
3	2400	0.01	500	32	98.7	98	98

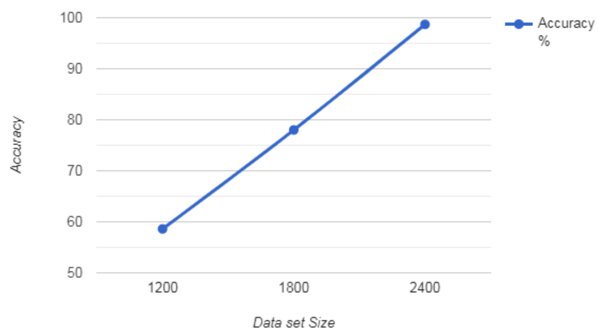


Fig. 4. Accuracy Measure for Various Dataset Sizes.

It is observed that CNN gives better performance with speech recognition compared to other related approaches described in literature review.

V. Conclusion

Speech recognition is a very important and useful application to interact with machines using speech. This emerging field includes many challenges like dataset size, noisy environment, age, gender, dialect and mental state of the speaker and many more. Speech wav files for Gujarati digits are stored and features extracted with MFCC technique. In this research paper, digit recognition for Gujarati language is performed using CNN.

Speech is passed and processed as input into multiple convolutional layers. For model training Tensorflow and Keras are used with python. Maximum Accuracy rate of recognition out of three experiments with different datasets size for all digits 0 to 9 is found 98.7%. It is observed that accuracy for speech recognition models can be improved by increasing the size of the dataset. Deep learning approach implemented with CNN trained with a large dataset gives better performance. In the future the proposed model can be enhanced to observe results for major parameters such as age, gender and dialect of the speaker that affect the speech recognition result.

References

[1] Daniel Jurafsky et al. Speech and Language Processing - An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (3rd ed. draft). https: //web.stanford.edu/˜jurafsky/slp3. 2020.

[2] Tailor, J. H., & Shah, D. B. (2015). Review on Speech Recognition System for Indian Languages. International Journal of Computer Applications, 119(2).

[3] Tailor, J. H., & Shah, D. B. (2016). Speech recognition system architecture for Gujarati language. International Journal of Computer Applications, 138(12).

[4] Wazir, A. S. M. B., & Chuah, J. H. (2019, June). Spoken arabic digits recognition using deep learning. In 2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS) (pp. 339-344). IEEE.

[5] Gunawan, A. (2010). English digits speech recognition system based on hidden markov models “. In Proceedings of International Conference Computer.

[6] Chapaneri, S. V., & Jayaswal, D. J. (2013). Efficient speech recognition system for isolated digits. IJCSET, 4(3), 228-236.

[7] Datta RKS., Rudresh, M. D., & Shashibhushsan, G. (2021). Comparative performance analysis for speech digit recognition based on MFCC and vector quantization. Global Transitions Proceedings, 2(2), 513-519.

[8] Kurian, C., & Balakrishnan, K. (2013). Connected digit speech recognition system for Malayalam language. Sadhana, 38(6), 1339-1346.

[9] Sen, O., & Roy, P. (2021, September). A Convolutional Neural Network Based Approach to Recognize Bangla Spoken Digits from Speech Signal. In 2021 International Conference on Electronics, Communications and Information Technology (ICECIT) (pp. 1-4). IEEE.

[10] Zerari, N., Abdelhamid, S., Bouzgou, H. & Raymond, C. (2019). Bidirectional deep architecture for Arabic speech recognition. Open Computer Science, 9(1), 92-102. https://doi.org/10.1515/comp-2019-0004.

[11] Dhanashri, D., & Dhonde, S. B. (2017). Isolated word speech recognition system using deep neural networks. In Proceedings of the international conference on data engineering and communication technology (pp. 9-17). Springer, Singapore.

[12] Hamidi, M., Satori, H., Zealouk, O., & Satori, K. (2020). Amazigh digits through interactive speech recognition system in noisy environment. International Journal of Speech Technology, 23(1), 101-109.

[13] Swedia, E. R., Mutiara, A. B., & Subali, M. (2018, October). Deep learning long-short term memory (LSTM) for Indonesian speech digit recognition using LPC and MFCC Feature. In 2018 Third International Conference on Informatics and Computing (ICIC) (pp. 1-5). IEEE.

[14] Sustika, R., Yuliani, A. R., Zaenudin, E., & Pardede, H. F. (2017, November). On comparison of deep learning architectures for distant speech recognition. In 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE) (pp. 17-21). IEEE.

[15] Audichya, M.A., & Saini, J.R. (2017, September). A Study to Recognize Printed Gujarati Characters Using Tesseract OCR. International Journal for Research in Applied Science and Engineering Technology, 5(9), 1505-1510. doi: 10.22214/ijraset.2017.9219.

Tweet Credibility Detection for COVID-19 Tweets using Text and User Content Features

Vaishali Vaibhav Hirlekar[1]
Department of Computer Science and Engineering
Sir Padmapat Singhania University
Udaipur, Rajasthan, India

Arun Kumar[2]
Department of Computer Science and Engineering
Sir Padmapat Singhania University
Udaipur, Rajasthan, India

***Abstract*—The deadly COVID-19 pandemic is currently sweeping the globe, and millions of people have been exposed to false information about the disease, its remedies, prevention, and origins. During such perilous times, the propagation of fake news and misinformation can have serious implications, causing widespread panic and exacerbating the pandemic's threat. This increasing threat factor has given rise to considerable research challenges. This article is mainly concerned about fake news identification and experimentation is specifically performed considering COVID-19 fake news as a case study. Fake news is spread intentionally to mislead the people and therefore we need to identify user's involvement and it's correlation with additional features. The aim of this research is to develop a model that can predict the essence of a tweet given as an input with the help of multiple features. Our strategy is to make use of the tweet's text as well as the user's metadata and develops a model using natural processing technique and deep learning method. In this process, we have analyzed the behavior of the accounts, observed the impact of the various factors that can lead to fake news. The experimental analysis shows that hybrid model with text and content features have generated a benchmark result than the existing state of art techniques. We have obtained a best F1-score of 0.976 during the experimentation.**

Keywords*—Fake news; machine learning; natural language processing; deep learning

I. Introduction

Fake news and its consequences have the ability to impact a wide range of institutions, from a citizen's lifestyle to a country's international relations. Widespread use of social media sites been generating and exchanging more content than ever before, some of which is deceptive and has little bearing on fact. So far several related works has been carried out for collecting and identifying fake news, however no commercially viable system exists yet. Fake news isn't a new phenomenon but it has new repercussions and effects. Fake news can lead to the collapse and failure of the world's largest economies by mass exploitation, and it can be one of the most devastating "internet wildfires." Aside from political ramifications, fake news can and has resulted in personal defamation, distorted perspectives, and mass incitement on a variety of topics. It is much easier for the sources to produce this news than it is for people to embrace and share it. Fake news is the biggest danger to our ostensibly functioning democracy; in addition to distorting and corrupting ideologies, it has also resulted in real effects such as cyber defamation, cyber stalking and other cyber-attacks.

Detecting and identifying fake news on a social media is a difficult challenge. The rapid dissemination of false news has an impact on millions of people and their actual surroundings. The propagation of fake news is not a new issue on the social media sites [1]. Several firms and well-known individuals utilize various social media networks to promote their products and build their reputation. All of these operations persuade numerous people to share and enjoy the news. As a result of this process, fake news spreads over the web. In terms of a certain issue, the content, style, and media platform of fake news change with time and fake news tries to falsify linguistic information. Fake news may contain genuine evidence within a fabricated framework to promote a false assertion [2]. The term fake news has coined in the 2016 US Election primarily, which encouraged academicians and researchers to do the research in this direction [3]. The researchers tried to gather the data from various resources and then checked the actual authenticity of the news being spread. Since then people have been utilizing a variety of manual techniques to do the fact-checking, such as using fact-checking websites. These websites are crucial in spotting false news on the internet. There are a number of fact-checking and fake news identification research projects, methods, and applications available, most of which look at the issue from a veracity classification standpoint. Misinformation, disinformation, hoax, and rumor are all terms used in similar literature to describe fake news. The term "misinformation" refers to the spread of incorrect information without consideration of the real intention. The goal of misinformation is to fool the intended recipient of the information. Rumours or hoaxes are purposely crafted to appear accurate. The fake information is for the gullible. The person may not realize the actual authenticity and believes in on what is being spread through social media especially the social sites eager to increase their viewership.

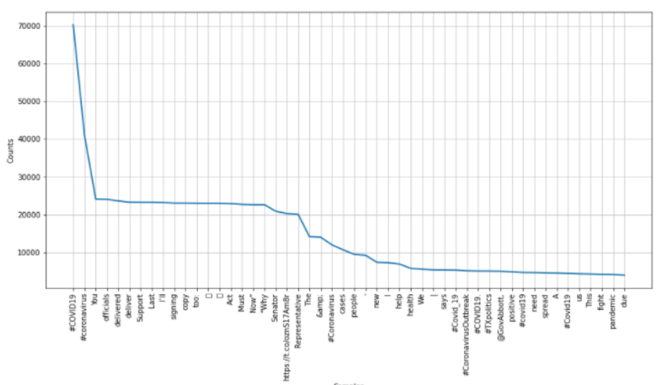
A. *Covid 19 – An Infodemic*

Covid 19 Infodemic has become more like a disease which is spreading rapid faster in the society through the dissemination of false information. Verifying the veracity, authenticity, and accuracy of given information is extremely difficult, especially when it concerns a horrifying disease that poses a threat to humankind. [4] COVID19, a virus that first appeared in Wuhan, China, in December 2019, has spread to 213 nations, regions, or territories throughout the world, resulting in roughly 3,478,418 fatalities as of May 24, 2021.

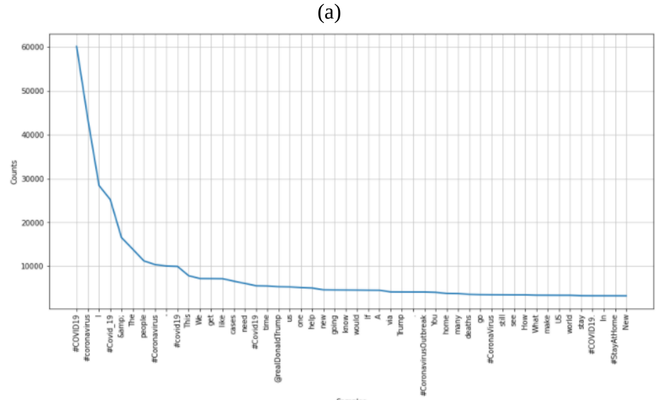
This infodemic posed a serious problem for public health alongwith social media channels including as Facebook, Instagram, WhatsApp and Twitter have become key sources of information on the crisis. In the COVID19 battle, fighting the infodemic is a new front. People trust the information that appeals to their emotions and personal opinions than information that is considered factual or objective in the 'posttruth' era. The following figure shows the news trend encountered during the Covid time. Fig. 1(a) shows plot of 50 most commonly used words in real tweet, whereas Fig. 1(b) shows plot of 50 most commonly used words in fake tweet.

The epidemic and infodemic elements of a pandemic are the two sides of the coin in today's highly digitalized society. This infodemics are usually fuelled by a combination of human and non-human system (bots), all of whom are pursuing essentially unknown aims. In this perspective, we present a methodology to detect fake tweets using the COVID-19 epidemic as a case study by using ensemble learning models with machine learning and Deep Learning Techniques.

In this research work, we have studied the differences between fake and real tweet based on behavioral, content based and comment based features of the tweet. The methodology implemented tackles the problems of fake tweet with the help of natural language processing toolkit and performs tweet text analysis as well. Different from the existing work, we take into consideration not only text characteristics, however also used user account characteristics for better results.



(a)



(b)

Fig. 1. (a) Shows Plot of 50 most commonly used Words in Real Tweet, (b) Shows Plot of 50 most commonly used Words in Fake Tweet.

The related work of fake news detection has been presented in Section II. Section III gives systematic overview of approaches used for Fake Tweet detection process. A Methodology along with model implementation has been discussed Section IV. The experimental setup and results are reviewed in Section V whereas Section VI makes the conclusion and recommendations for further research.

II. Related Work

Since all individuals communicate through a virtual world using social media, the methods of identifying fake information have been researched. Fake information spread by many social network users utilizing various platforms for financial or personal benefits. The evolution of social networking platforms encourages users to transmit and exchange information about their activities. However it has created a challenge for the researchers who want to secure user data from a variety of dangers. For researchers, detecting harmful information in the form of false news has become a significant challenge.

Looking back in time, misleading news isn't a new issue. Since a long time, there has been widespread anxiety about such news. The scientific community began to pay attention to the issue of fake information in the early 2000s, which expressed itself in the form of paid posters, review spam and rumor detections. The term "fake news," on the other hand, was popularized during the 2016 US Presidential Election [5] [6]. The impact of such news was unclear, although it spread some gossip, uncertainty, and deceit among users. Because of the growing media environment erroneous political information has also been circulating widely. The content of the news and the social circumstances are the most important factors in the fake news identification. News is classified into two categories: textual and visual, however emotions are also integral part of the news content. In addition, deep neural networks [7] [8] are used to frame latent textual representation.

Many of the researcher have also used different supervised and unsupervised, adversarial, user response methods. [9] Here author had created a false news detection algorithm utilizing machine learning approaches along with n-gram analysis. Here, author employed numerous characteristics collected using two distinct approaches and evaluated in six different machine learning settings. TF–IDF used as feature extraction method, and Support Vector Machine algorithm has been used in experimentation. In [10] researcher had developed a model to detect hoaxes or non-hoaxes distributed on social media platforms such as Facebook and created an automated fake news credibility inference algorithm to detect fake information. Here, [11] author had analyzed numerous variables such as user profile data and the relationship between users and the originator of the fake news. In [12] author, suggested a graph neural network-based technique and analyzed non-Euclidean data using a graph neural network. They often avoided certain written content by using unseen data for implementation. In addition, [13] author proposed model using twelve classifiers and evaluated on three datasets. The false prediction ratio of these ML classifiers is used to merge them. Based on their performance measures, Linear Support Vector Classifier, Logistic Regression and Passive Aggressive along with TF–IDF, CV, and HV feature extraction methods. In [14] developed a volunteer-based crowd annotation tool by combining the perspectives of multiple stakeholders, the system were based on the Micro Mappers platform for English and Arabic tweets. To promote additional studies, [15] published consolidated content named CORD-19, which comprises 59,000 publications regarding COVID-19 and information associated with coronaviruses. The researchers have tried a variety of techniques to combat the COVID-19 Infodemic in recent months.

In [16] performed spatio-temporal analysis of the flow of information and the transmission of COVID-19. Author proposed a model [17] trained on several Indic Languages wherein fake news dataset tweets had performed better due to syntactic features. In Hindi, the model showed 79 percent and in Bengali 81 percent F-Score. [18] [19] conducted substantial research on the usage of machine learning strategies to resolve numerous COVID 19 difficulties. In [20] analyzed Facebook ads from 64 countries and discovered that about 5% of them included potential disinformation. However, none of these methods helped to address the disinformation issue by providing an explanation for the supplied fake assertion.

In [21] employed a natural language inference (NLI) model that was upgraded by adding internal semantic relatedness scores and ontological WordNet elements and performed claim verification on the FEVER Dataset. In [22] proposed a model developed using natural language processing methods and used different algorithms of machine learning, and deep learning that comprises of a categorization strategy that employs new twitter attributes. The approach is built in tandem with Apache Spark and achieved 79% accuracy using random forest algorithm. The author also noted that the emotion of tweets is essential in tweet categorization. In the process of detecting misleading information, [23] collected data from various fact-checking websites, performed filtering, preprocessing and feature selection operation on the text part of tweets and observed that Neural Network, Logistic Regression and Decision Tree classifiers has given the best performance from the different perspectives.

Here author used [24] two distinct methods to build a model that can implement constraint based task. To improve F1-score across several test sets by executing impact data purification with a high cleansing percentage (25%) and experimented a model with a 99 percent cleaning percentage and obtained the 54.33 percent F1 score and 61.10 percent accuracy. In this paper, author [25] had developed a BERT-based model with other important Twitter features. In addition, the method was extended to many Indian languages, and a mBERT-based model was used with Hindi and Bengali datasets and provided a methodology to solve the issue of data scarcity in low-resource languages. Using the annotated data, the model observed 81%, 79% percent F-Score in Bengali and Hindi Tweets, and 81 % F-score with zero shot model.

Although all of the aforementioned research indicated that studying and detecting fake news in the social media using various techniques is successful however several limitations were found. In some of the approaches various learning algorithm has been used to detect the fake news, which involves more processing time and brings limitation to the various accuracy parameters due to size of the dataset. Most of the technique concentrates only the text pattern of news, however there could be few factors that could differentiate between fake and real content.

In summary, fake content detection on any social media is a challenging task. A deep analysis is needed to identify user's involvement through various features. An exploratory study is performed in this article to explore the link between the content-based, comment-based, sentiment-based, and behavior-based characteristics, as well as the interrelationship between them. Furthermore, a hybrid model is proposed based on the integration of the text and metadata and tested on various standard deep learning and machine learning algorithm.

III. A Systematic Overview of Approach used for Fake Tweet Detection

The methodology implemented in this paper tackles the problems of fake tweet with the help of natural language processing toolkit and deep learning algorithms. Ultimate aim of the research is to classify each tweet into 2 distinct categories i.e. "real" or "fake". Experimentation has been carried out with the perspective wherein combination of user and content features will be used along with the tweet text. We have used multiple input models for the experimentation to handle continuous and numeric features efficiently.

A. Dataset used for the Experimentation

The dataset we selected contains the tweets of users from 29-03-2020 to 15-04-2020 using the following hashtags: #epitwitter, #coronavirusoutbreak, #covid19, #coronavirus, #ihavecorona, #corona, #coronavirusPandemic. From about 11 April 2020, the dataset also included the following additional hashtags: #StayHomeStaySafe, #TestTraceIsolate. The dataset contains variables as given: location, hashtag, title of the tweet, tweet text alongwith tweet account details. Dataset does not include retweets, although a count of retweets is provided as a variable. Alongwith the 'retweet_count' some of the other features are also included in the dataset i.e. favourites_count', 'followers_count', 'friends_count' which has been successfully used for improving the accuracy of the model. Around 303692 tweets have been used for the experimentation during the process, among which 156612 were fake tweets whereas 147080 were real tweets.

B. An Exploratory Analysis for Data Insight

An Exploratory Data Analysis has been performed for preliminary investigations on data in order to identify patterns, spot anomalies, testing hypotheses, and validating assumptions using summary statistics and graphical representations. The process of computationally identifying and classifying viewpoints in a text, with the goal of determining whether the writer has a +ve, -ve, or neutral attitude toward a certain topic, product, etc.

Fig. 2 shows Polarity and subjectivity score with sentiment and subjectivity flag wherein the subjectivity score range between [0.0, 1.0], polarity score shows in the range [-1.0, 1.0], wherein 1.0 being very subjective and 0.0 being highly objective.

	polarity	subjectivity	sentiment_flag	subjectivity_flag
0	0.000000	0.000000	neutral	objective
1	0.000000	0.000000	neutral	objective
2	0.541667	0.708333	positive	subjective
3	-0.100000	0.200000	neutral	objective
4	0.250000	0.333333	neutral	neutral

Fig. 2. Polarity and Subjectivity Score with Flag.

Subjective sentences generally refer to personal opinion, emotion or judgment whereas objective sentence refers to factual information. Subjectivity is a float value which lies in the range of [0,1]. Here the value 0.7 represents that subjectivity is more, which ultimately refers that mostly it is a public opinion and not a factual information. Polarity, also known as orientation is the emotion expressed in the sentence. It can be positive, neagtive or neutral. Polarity is float value which lies in the range of [-1,1], here 0.5 refer to positive sentence.

Next distribution of user and content variables and relationships between other variables has been studied using pair plots in which the favorites, re-tweet, followers and friends count considered as a parameters.

Fig. 3 shows that favorites count and friends count are positively correlated. It also appears that re- tweet count and followers count has partial effect on favorites count. However during the analysis, all the features have been tested individually as well as in the combination to determine their impact on the tweet credibility detection. Furthermore, efficacy is also been tested with tweet content and user features in the deep learning and machine learning environment.

C. Approach used for Fake Tweet Detection

Extracting features from the content of the tweet is a reliable pattern recognition system and hence using natural language processing (NLP) is the most obvious solution to automatically identify fake news. To begin with this approach, data preprocessing has been carried out on tweet text part. Data preparation is a key stage in the modeling process, and the outcomes are dependent on how well the data has been preprocessed. In this approach, text normalization is performed which includes: converting numbers into words, converting letters, removing white spaces and punctuations, removing numbers and stop words. Pre-processed tweet text has been further given to the different vectorizer. For machine learning algorithms, TF-IDF is employed as it assigns a frequency score to words by emphasizing those that occur more often inside a document but not across documents. For deep learning algorithms, Glove has been used to obtain vector representations for words.

Tweet text feature is nothing but the pre-processed and normalized text through which the credibility of the content can be verified. Tweet text analysis done with NLP techniques has been able to evaluate the credibility of the tweet to some extent. However, text feature alone may not be enough to give better accuracy when it comes to detect the credibility. Features that come along with the tweet can be called as metadata of the tweet for exa. quote, favorites, retweet, followers, friends can play major role in the process of evaluation of the tweet because here tweets becomes more than strings with certain metadata added to them. This metadata becomes the features as an additional input dimension for an algorithm. With this perspective, we have tried to assess the tweet credibility by checking media content, account information and text characteristics. So here we have used 4 user features i.e. favourites_count', 'retweet_count', 'followers_count', 'friends_count' along with processing of tweet text that can make significance difference to the prediction of the news. Here we have tried to evolve a methodology to rate the credibility of the tweet based on the correlation with the additional features. In this phase, a tweet and user content feature has been used to perform the classification of the tweet.

- Behavioral Feature - considers user characteristics and linked user account attributes.
- Content based features - considers content of the tweet.
- Comment based features – considers the characteristics of the tweet itself.

The aim of our approach is how well multi-features-based method will be able to differentiate between the fake and real tweet. In this regards, we have tested different baseline algorithms like Random forest, logistic regression, Decision Tree, Naïve Bayes, XGBoost , Convolutional Neural Network , Bidirectional LSTM, and hybrid algorithm like CNN-BiLSTM for the tweet classification using the text and metadata features of the tweet. The methodology and algorithm have been discussed extensively in the next section that ultimately shows the flow of the process. Following tables are used to show features used for studies. Table I shows Tweet content features and Table II shows User content features used during the studies.

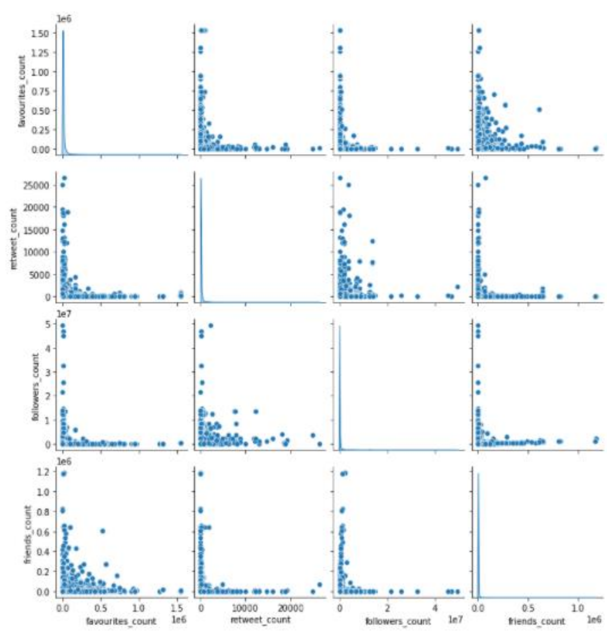


Fig. 3. Correlation between the Tweet Content Features.

TABLE I. TWEET CONTENT FEATURE USED FOR STUDY

Feature	Name	Type	Description
TC1	Title	Textual	Short text of the tweet, it is a summary of the topic's highlight
TC2	Text	Textual	Extended part of tweet that gives topic's details.
TC3	Source	Textual	Indicates source of the tweet.
TC4	is_quote	Boolean	A quote tweet is a re-tweet with some additional text attached. This parameter determines whether or not the selected tweet contains a quote.
TC5	is_retweet	Boolean	A retweet re-sends selected tweet. This field checks whether the selected tweet is a retweet.

TABLE II. USER CONTENT FEATURE USED FOR STUDY

Feature	Name	Type	Description
UC_1	**f**avourites_count	Numerical	It's the number of tweets that given user has marked as favourite or in the account's lifetime, the number of Tweets this user has liked.
UC_2	Retweet_Count	Numerical	Retweet count always apply to the original tweet only, there is no counts for a "retweet" tweet, only the original, retweeted tweet. For example, if tweet B is a retweet of tweet A, and C is a retweet of B, in the end So in this example: B will have a count of 0, and so will C. A will have a retweet count of 2. As more individuals repost the tweet, this number may vary.
UC_3	*Followers_Count*	Numerical	No. of followers of the twitter account.
UC_4	*Friends_Count*	Numerical	It shows the number of friends of twitter account. However the ratio of followers to friends may well impart some useful information about the way in which the twitter account is being used.

IV. MODEL IMPLEMENTATION

A. *Fake Tweet Detection with text and user Features using Statistical Approach*

A statistical model is designed to derive the inference about the relationship between the variables and further used to predict the fake tweets. Based on the exploratory analysis, we have found the relationship between the user and content parameters, which are from different domain (i.e text and numeric). These features are combined further to build a model. In order to create a model that can handle continuous data and text data, following algorithm is used. Our dataset d has total of n data points: (d1, y1), (d2, y2), ……, (dn, yn) respectively , where di is the ith tweet and yi is its label. Each input sample, di comprises 2 input sub-sets here — tweet content feature (d^i_{TC}) and user content feature (d^i_{UC}).

Algorithm for Tweet and User features using Statistical Approach.

- Read tweet content feature (d^i_{TC}).
- nlp_input (V_{TC}) <- Process the d^i_{TC} data using.
- Preprocessing.
- Read all user content feature ($d^i_{UC1,UC2,UC3,UC4}$).
- tw=Convert to a matrix of TF-IDF features (V_{TC}).
- Concatenate the features (tw, UC1, UC2, UC3, UC4).
- Spliting into Train and Test set.
- Model <- classifier ( ).
- Model.fit (features_train data).
- Prediction=Model (features_test data).

Experimental model 1 used Machine learning approach with combination of text and user feature which consists of Text Pre-processing, Tokenization, TF-IDF vectorizer for text part and then converted all the features generated from metadata and n-gram frequencies of the text into a matrix. Each row represents a tweet and each column the value of one of user features. Further we have used different classifier algorithms like Random Forest, Naïve Bayes, Logistic Regression, Naive Bayes, Decision Tree and XGBoost for the experimentation.

B. *Fake Tweet Detection with text and user Features using CRED_Tweet Model*

In this approach, algorithm operates in two phases. In the first phase extraction, d^i_{TC} data i.e. text input is processed using regular expression with the help of nltk library to perform basic data cleaning operation and then pre-trained glove model is used to convert text into embedding. In the next phase, User features d^i_{UC} are then interpolated to higher dimensional dense feature vectors termed meta_input d^i_{UC} through separate convolution kernel. In the proposed (CRED_Tweet) model shown in Fig. 4, a combination of multiple inputs CNN-BiLSTM architecture is successfully used providing text and metadata information through different convolution layers.

Here, vocabulary size V is used to represent the tweet text of length N. The embedded input is passed through 3 different convolution layer Conv1D (128, 5, activation='relu') followed by max pooling layer. Here, in the convolution layer used 128 filters. The convoluted vectors is been given to Bidirectional LSTM layer with 128 internal unit with dropout=0.3 and kernel_regularizer attributes and further to flatten. In the next phase, User features d^i_{UC} are then interpolated to higher dimensional dense feature vectors termed meta_input d^i_{UC} through separate convolution, pooling and flatten layers and further concatenated both the dense tensors, $Z_{i\,(ht)} \oplus X_i'$. The vector Z_i is then passed across a fully connected network, with the probability distribution across the two classes being regularized using a dropout layer.

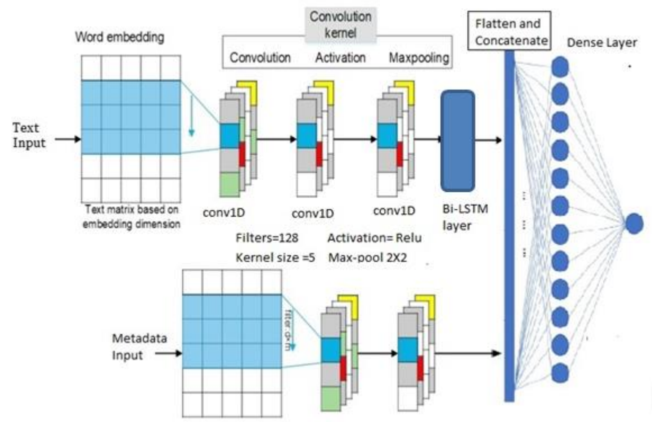


Fig. 4. Block Diagram of CRED_Tweet Model using user and Content Features.

Phase I: Extracting text features

In this model, tweet text features (TC) and User content feature (UC) are based on a variant of CNN and processed individually through two different CNN model as the nature of both the feature are different i.e. text and numerical respectively. Although CNNs are mostly employed in image classification [26] or object identification, they have also demonstrated noteworthy performance in several Natural Language Processing (NLP) applications including text classification [27] [28]. Using the convolutional approach, the neural network produce local features around each word of the neighbouring word in the first phase and later they are combined using a max operation. As a result, we use CNN to model textual characteristics for the identification of fake news. Fig. 4 shows the layered processing of the CRED_Tweet model using user and content features.

Let Z be the text input, here maximum length of the tweet considered as n and padded with sequence length m. $\in d^{i}{}_{TC.}$

$$Z_i^{TC} = Z_{i,1} \oplus Z_{i,2} \oplus Z_{i,3} \oplus . Z_{i,n} \tag{1}$$

Here, each tweet is been re-presented as a matrix and then convolution filter are applied to derive new features. Here, we have used convolutional filters $w \in R$ Hl×Wl×Dl×N to construct the new features. Equation (2) shows default convolution function. When the word matrix is been processed through a convolution layer, it can produce the feature Z i.

$$Z = W^T.X + b \tag{2}$$

$$Z_i = f\,(\,W^T.\,Z_i^{TC} + b\,) \tag{3}$$

Here the b represents the bias, and · is the convolutional operation. A layer or convolution kernel includes convolution layer, activation layer and pooling layer as shown in Fig. 4. The function f is the non-linear transformation and also includes ReLU activation layer here.

$$Z_i = \max\{0, Z_i\} \tag{4}$$

During the experimentation, Zi has been passed through 3 different convolution layers of 128 neurons and having filter size = 5 and activation='relu. The filter generates a feature map by running over every potential window of words in the tweets. A max-pooling layer [24] is utilized to get the maximum feature map. The maximum value is denoted as Z_i'.

$$Z_i' = \max\{\,Z_i\,\} \tag{5}$$

By saving the most significant convolutional findings for false news detection, the max-pooling layer can considerably increase the model's robustness. The CNN has the advantage over the LSTM as it decreases the number of dimensions in the input features that must be provided to a sentiment classifier or a natural inference prediction model after the feature extraction stage.

These token vectors (Z_i') are further encoded using a Bi-LSTM, using the forward and backward layers which processes the N vectors in opposite directions. a hidden state hft is emitted by the forward LSTM at each time-step, which is concatenated with the corresponding hidden state hbt of the backward LSTM to produce a vector ht $\in R^{\,Hl\times Wl\times Dl\times N}$.

$$Z_{i\,(ht)} = Z_{i\,(hft \oplus hbt)} \tag{6}$$

Phase II: Extracting metadata features

Let X be the metadata numeric input $\in d^{i}{}_{UC}$. X_i^{UC} has been passed through 2 different convolution layers of 128 neurons and having filter size = 5 and activation='relu.

$$X_i^{UC} = X_{i,1} \oplus X_{i,2} \oplus X_{i,3} \oplus \ldots X_{i,n} \tag{7}$$

$$X_i = f\,(\,W^T.\,X_i^{UC} + b\,) \tag{8}$$

$$X_i = \max\{0, X_i\} \tag{9}$$

$$X_i' = \max\{\,X_i\,\} \tag{10}$$

At the end, concatenate both the features (text and metadata) feature.

$$Z_i' = Z_{i\,(ht)} \oplus X_i' \tag{11}$$

A convolutional network's fully connected layers are essentially a multilayer perceptron that used to map the $m^{(l-1)}{}_1 \times m^{(l-1)}{}_2 \times m^{(l-1)}{}_3$ activation volume from the preceding various layers into a class probability distribution. As a result, the multilayer perceptron's output layer will contain $m^{(l-1)}{}_1$ output neurons, where i specifies the number of layers in the multilayer perceptron.

$$y_i = f\,(Z_i') \tag{12}$$

Pred (y for given di TC, di UC ; θ) = activation function(f) on (Zi').

Here, Z_i' signifies the changed vector after passing through the relevant feed forward sub-network and sigmoid activation function, while θ denotes the model parameters employed throughout the experiment.

Loss function and optimizer:

The aim of any optimization problem is to minimize the cost function, which means of measuring how accurate the data is. We utilized Binary cross entropy in this case, which compares each of the predicted probabilities to the actual class output, which can be 0 or 1. The score is then calculated, penalizing the probabilities depending on their deviation from the predicted value. This refers to how close or far the value is to the real value. The negative average of the log of corrected projected probability is shown by Binary Cross.

Entropy.Loss= abs(y_pred – y_actual) (13)

```
Layer (type)                     Output Shape         Param #     Connected to
==================================================================================================
input_2 (InputLayer)             [(None, 1000)]       0
__________________________________________________________________________________________________
embedding_1 (Embedding)          (None, 1000, 100)    72004300    input_2[0][0]
__________________________________________________________________________________________________
conv1d_5 (Conv1D)                (None, 996, 128)     64128       embedding_1[0][0]
__________________________________________________________________________________________________
max_pooling1d_5 (MaxPooling1D)   (None, 199, 128)     0           conv1d_5[0][0]
__________________________________________________________________________________________________
conv1d_6 (Conv1D)                (None, 195, 128)     82048       max_pooling1d_5[0][0]
__________________________________________________________________________________________________
metadata_input (InputLayer)      [(None, 4, 1)]       0
__________________________________________________________________________________________________
max_pooling1d_6 (MaxPooling1D)   (None, 39, 128)      0           conv1d_6[0][0]
__________________________________________________________________________________________________
conv1d_8 (Conv1D)                (None, 4, 128)       384         metadata_input[0][0]
__________________________________________________________________________________________________
conv1d_7 (Conv1D)                (None, 35, 128)      82048       max_pooling1d_6[0][0]
__________________________________________________________________________________________________
max_pooling1d_8 (MaxPooling1D)   (None, 2, 128)       0           conv1d_8[0][0]
__________________________________________________________________________________________________
max_pooling1d_7 (MaxPooling1D)   (None, 1, 128)       0           conv1d_7[0][0]
__________________________________________________________________________________________________
conv1d_9 (Conv1D)                (None, 2, 128)       32896       max_pooling1d_8[0][0]
__________________________________________________________________________________________________
bidirectional_1 (Bidirectional)  (None, 256)          263168      max_pooling1d_7[0][0]
__________________________________________________________________________________________________
max_pooling1d_9 (MaxPooling1D)   (None, 1, 128)       0           conv1d_9[0][0]
__________________________________________________________________________________________________
flatten_2 (Flatten)              (None, 256)          0           bidirectional_1[0][0]
__________________________________________________________________________________________________
flatten_3 (Flatten)              (None, 128)          0           max_pooling1d_9[0][0]
__________________________________________________________________________________________________
concatenate_1 (Concatenate)      (None, 384)          0           flatten_2[0][0]
                                                                  flatten_3[0][0]
__________________________________________________________________________________________________
dense_2 (Dense)                  (None, 128)          49280       concatenate_1[0][0]
__________________________________________________________________________________________________
dense_3 (Dense)                  (None, 2)            258         dense_2[0][0]
==================================================================================================
Total params: 72,578,510
Trainable params: 72,578,510
Non-trainable params: 0
```

Fig. 5. Fitting the Convolutional Neural Network with Bilstm Model.

Optimizers are techniques or strategies for changing the characteristics of a neural network, such as weights and learning rate, to minimize losses. To minimize losses, the optimizer determines how to alter the weights or learning rates of the neural network. Adam optimizer has proven benchmark outcomes above existing state-of-the-art algorithms by training the neural network in less time and more effectively. Fig. 5 shows fitting of the convolutional neural network with Bi-LSTM model. From the given figure, we can analyse how multiple inputs of different types are processed here through different layers. Continuous input i.e. text or content features of the tweets has been processed through convolutional layers and then passed through Bi-LSTM layer whereas metadata features i.e. user features has been processed through separate convolution layer, then concatenated output of both layer and provided further to fully connected layer for final predictions. In this process, total trainable parameters encountered are 72,578,510.

Precision: Conversely, precision score represents the ratio of true positives to all events predicted as true. In our case, precision shows the number of articles that are marked as true out of all the positively predicted (true) articles:

Precision = TP/ (TP+FP) (14)

Recall: Recall represents the total number of positive classifications out of true class. In our case, it represents the number of articles predicted as true out of the total number of true articles.

Recall = TP/(TP+FN) (15)

F1-Score: F1-score represents the trade-off between precision and recall. It calculates the harmonic mean between each of the two. Thus, it takes both the false positive and the false negative observations into account. F1-score can be calculated using the following formula:

F1= 2. (Precision.Recall)/(Precision+Recall) (16)

Accuracy: Accuracy is often the most used metric representing the percentage of correctly predicted observations, either true or false. To calculate the accuracy of a model performance, the following equation can be used:

Accuracy = TP+TN / (TP+TN+FP+FN) (17)

The predicted results are evaluated with confusion matrix and other measures like True negative rate (Specificity), True positive rate (TPR), Precision, Recall (Sensitivity), F1-score, accuracy, PRC (Precision-Recall curve) and ROC (Receiver operating curve) etc. Tables III and IV shows the performance results of various parameters. Performance of several user profile categories, tweet content elements, and a combination of both has been examined. The best accuracy found is with decision tree 92.56 among all the ML algorithms along with 92.13 % precision and 92.54 % recall. However the further experimentation with deep learning approaches had shown a benchmark result over the existing state of art techniques.

Deep learning-based analysis has a higher accuracy and detection rate than machine learning. In the experiment 2, CNN with Bi-LSTM is used with the embedding layer and convolution layers with 128 tensors, 5 filters and with relu activation. Further Bi-LSTM is used with 128 neurons and 0.3 dropout and recurrent dropout, with regularizers.l2 (0.01). User features are also separately processed through different CNN layer with with 128 tensors, 5 filters and with relu activation. 'Relu' and 'softmax' activation function used in dense layer. Model is further compiled with 'binary_ cross entropy' loss and 'adam' optimizer and found best accuracy with 97.60 with text and metadata feature as shown in Table IV. Size of train set, test and validation set is as given here: 242953, 151846, 151846. The result shown in Table IV, deep learning scenarios, shows that combining certain features is based on user content and tweet content improves accuracy. Among all the algorithms, CRED_Tweet Approach had achieved 98.44 % precision, 96.56 % recall and 97.60 % accuracy.

TABLE III. SHOWS ACCURACY PARAMETERS FOR DIFFERENT MACHINE LEARNING CLASSIFIERS WITH TEXT AND METADATA FEATURES

Measures in %	**LR**	**NB**	**RF**	**DT**	**XG Boost**
Precision	85.17	91.80	91.72	92.13	90.91
Recall	91.02	74.40	90.97	92.54	91.41
F1 score	87.98	84.15	91.64	92.55	91.41
Accuracy	87.98	84.39	91.65	92.56	91.42

TABLE IV. SHOWS ACCURACY PARAMETERS FOR DIFFERENT DEEP LEARNING CLASSIFIERS WITH TEXT AND METADATA FEATURES

Measures in %	**Bi-LSTM**	**CNN**	**CRED_Tweet Approach**
Precision	87.22	98.41	98.44
Recall	93.02	95.44	96.56
F1 score	90.01	97.04	97.60
Accuracy	90.01	97.05	97.60

C. Analysis of the Result

The performance of the statistical ML models has been observed using ROC and precision-recall curve. ROC curves summaries the trade-off between the true positive rate and the false positive rate for a predictive model with varying probability thresholds whereas Precision-Recall curves illustrate the trade-off between a predictive model's actual positive rate and positive predictive value when different probability thresholds are used. The integral or an estimate of the area under the precision-recall curve is summarized as AUC (Area under curve). Fig. 6 illustrates ROC-AUCs and PRC-AUCs for the deployed models. Here, Random Forest and XG Boost Algorithm gives 0.97 %, Naïve Bayes gives 0.83 %, Logistic Regression and Decision Tree gives 0.93 % ROC-area under curve. The precision-recall curve is constructed by calculating and plotting the precision against the recall for different classifiers at a variety of thresholds. PRC identifies the Positive Predictive Value (precision) for each corresponding value on the sensitivity (recall) scale. Here, Random Forest and XG Boost Algorithm gives 0.97 %, Naïve Bayes gives 0.81 %, Logistic Regression and Decision Tree gives 0.94 % and 0.89 % respectively area under curve for PRC. In our experimentation, Decision Tree had the best precision and recall i.e. 92.13% and 92.54 % with the 92.56 % accuracy which is best among all the algorithms.

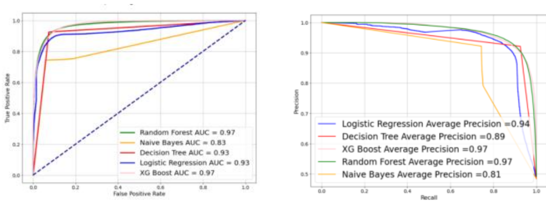


Fig. 6. Shows ROC and Precision-Recall Curve for LR,DT,RF,NB and XG Boost Classifier with Text and Metadata.

During the analysis, all the metadata features have been tested individually as well as in the combination to determine their impact on the tweet credibility detection. However, it has been observed that combination of both news content and user content features improves detection rate. The proposed hybrid model CRED_Tweet trained in order to improve work in this domain. The model outperforms LSTM with similar weights and shorter training time in terms of test accuracy. As a result, quicker training with CNN is feasible, decreasing the training time required for big datasets. Fig. 7 shows the plots of the training and test accuracy and loss values of the model over the 05 epochs. Model loss figure shows good fit learning curve which shows that training and testing loss that decreases to a point of stability remains almost the same in all epochs.

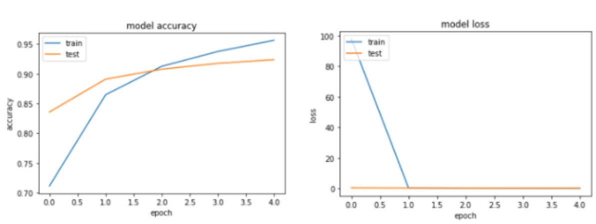


Fig. 7. Shows Accuracy and Loss of CRED_Tweet Model.

D. Comparative Analysis

We compare the CRED Tweet model to a few state-of-the-art approaches, which are listed below. TI-CNN technique used by author [29] with text and picture for fake news detection is by combining implicit and explicit characteristics. This technique has used 8,074 real news and 11,941 fake news which gave 92.2 % precision, 92.7 % Recall, 92.10 % F1-score and accuracy. In [30], author have used machine learning approaches with multiple features extracted from different sources, which used 2282 Buzz Feed news related to US election and found 85% AUC with Random Forest, 80 % with KNN and 86 % with XGB. [13] Here author have used Multilingual Approach for Fake Tweet Detection. For Indic i.e. Bengali and Hindi Languages & English give 92.75% Precision, 62.95% Recall, 75% F1-score and 81% accuracy. In [22] various machine learning algorithm has been experimented on Covid-19 epidemic fake news dataset which contains 5000 real tweets and 5000 fake tweets. The method gives 85 % precision, 82% F1-score and 79 % accuracy. Table V and Fig. 8 shows comparative evaluation with other approaches w.r.t precision, recall, F1-score and accuracy parameter.

TABLE V. SHOWS COMPARATIVE EVALUATION OF CRED TWEET WITH OTHER APPROACHES

Author	Methodology	Dataset	Results	Limitation
Yang et al. 2018 [29]	TI-CNN technique to evaluate picture and text for fake news analysis by combining explicit and latent characteristics..	20,015 news, i.e., 11,941 fake news and 8,074 real news	Precision-92.2 Recall-92.7 F1-score-92.10	The model trained only on CNN which may work better with picture but for text RNN model is needed.
Reis et al. 2019 [30]	Machine learning approaches are used with multiple features extracted from different sources	2282 Buzz Feed news related to US election	RF-85% KNN - 80% SVM - 79%	Accuracy for detecting fake account is very low due to small dataset.
D. Car et al. 2020 [25]	Multi-Indic-Lingual Approach used for COVID Fake-Tweet Detection	COVID-19 multilingual tweet dataset for Indic Languages (Hindi and Bengali) & English	Precision-92.75 Recall-62.95 F1-score-75.00	Accuracy for detecting fake account is very low due to small dataset.
Y. Madani et al. 2021 [22]	The various machine learning algorithm has been experimented on Covid-19 epidemic fake news dataset	FakeNewsNet dataset contains 10,000 fake and real tweets (5000 fake tweets and 5000 Real tweets).	RF-79%, DT-62% LR-60%, SVM-72% NB -53%, MLP-48 %	Accuracy for detecting fake account is very low due to small dataset.

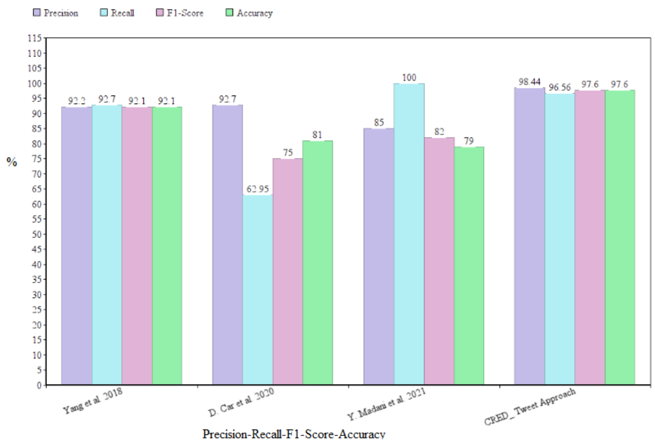


Fig. 8. Comparative Evaluation with other Approaches.

With respect to all the approaches specified above, CRED Tweet Approach uses CNN - Bi-LSTM for Tweet content feature and CNN model for user content features. In this approach 303692 tweets has been used for the experimentation during the process, among which 156612 were fake tweets whereas 147080 were real tweets. CRED_Tweet Approach had achieved 98.44 % precision, 96.56 % recall and 97.60 % accuracy which is considerably better, also the size of the dataset used during the experimentation is quite large as compared to other state-of-the-art approaches.

The methodology implemented in this paper tackles the problems of fake tweet with the help of natural language processing toolkit and deep learning algorithms. Ultimate aim of the research is to classify each tweet into two distinct categories i.e. “real” or “fake”. Experimentation has been carried out with the perspective wherein combination of user and content features will be used along with the tweet text. We have used multiple input models for the experimentation to handle continuous and numeric features efficiently.

V. Conclusion

In this research work, we have studied the differences between fake and real tweet based on behavioral, content based and comment based features of the tweet and further help to classify fake tweets in an extremely dedicated domain of COVID-19.

The methodology implemented tackles the problems of fake tweet with the help of natural language processing toolkit and performs tweet text analysis as well. Different from the existing work, we take into consideration not only text characteristics, however also used user account characteristics for better results and we have found that the efficacy of fake tweet detection is improved using tweet content features and user content features.

Our proposed model outperforms better than other baseline deep learning and machine learning approaches. Overall, the use of ANN in the identification of fake news appears to be promising. Aside from CNN and Bi-LSTM, we'll look at more complex neural network architectures in the future. When traditional models are combined with task-specific function engineering techniques, they can be extremely useful. In future, we aimed at doing in-depth exploratory analysis on the tweet in order to find out the indirect features that can affect the credibility of the news. Despite the enormous amount of existing works on fake news identification and detection, there is still room for improvements, and new profound developments into the nature of fake news can lead to more effective and accurate models.

References

[1] S. Sahoo and B.B. Gupta. Multiple features based approach for automatic fake news detection on social networks using deep learning. Applied Soft Computing, Volume 100, 106983, 2021. doi:10.1016/j.asoc.2020.106983.

[2] S. Feng, R. Banerjee and Y. Choi. Syntactic stylometry for deception detection. In Proc. of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers – Volume 2(ACL '12), Pages 171–175, 2012. doi/10.5555/2390665.2390708.

[3] B. Horne, S. Adali. This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News, 2017. ArXiv, abs/1703.09398.

[4] M. Cinelli, W. Quattrociocchi, A. Galeazzi, C.M. Valensise, E. Brugnoli, A. Schmidt, P. Zola, F. Zollo and A. Scala. The COVID-19 social media infodemic. Sci Rep 10, 16598 (2020). https://doi.org/10.1038/s41598-020-73510-5.

[5] E. Bakshy, S. Messing, L. A. Adamic. Exposure to ideologically diverse news and opinion on Facebook. SCIENCE05 JUN 2015 : 1130-1132. DOI:10.1126/science.aaa1160.

[6] M. Barthel, A. Mitchell and J. Holcomb. Many Americans believe fake news is sowing confusion, Pew Res. Center 15 (12) (2016).

[7] H. Karimi, P. Roy, S. Sadiya and J. Tang. In Proceedings of the 27th International Conference on Computational Linguistics, Pages 1546—1557, 2018. https://aclanthology.org/C18-1131.

[8] S. Hosseinimotlagh and E. E. Papalexakis, Unsupervised content-based identification of fake news articles with tensor decomposition ensembles. In the Proc. of Misinformation and Misbehavior Mining on the Web Workshop held in conjunction with WSDM , 2018.

[9] H. Ahmed, I. Traoré and S. Saad. Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques. ISDDC 2017. DOI:10.1007/978-3-319-69155-8_9.

[10] E. Tacchini, G. Ballarin, M.L. Vedova, S. Moret and L.D. Alfaro (2017). Some Like it Hoax: Automated Fake News Detection in Social Networks. ArXiv, abs/1704.07506.

[11] J. Zhang, B. Dong and P. S. Yu. Fake news detection with deep diffusive network model, 2018. arXiv:1805.08751.

[12] Y. Han, S. Karunasekera and C. Leckie, Graph neural networks with continual learning for fake news detection from social media, 2020. arXiv:2007.03316.

[13] D. Kar, M. Bhardwaj, S. Samanta, A. Azad. No Rumours Please! A Multi-Indic-Lingual Approach for COVID Fake-Tweet Detection, 2020. arXiv:2010.06906.

[14] F. Alam, F. Dalvi, S. Shaar, N. Durrani, H. Mubarak, A. Nikolov, G. Martino, A. Abdelali, H. Sajjad, K. Darwish and P. Nakov. Fighting the covid-19 infodemic in social media: A holistic perspective and a call to arms,2020. arXiv:2007.07996 [cs.IR].

[15] L. Wang, K. Lo, Y. Chandrasekhar, R. Reas, J. Yang, D. Eide, K. Funk, R. Kinney, Z. Liu, W. Merrill, P. Mooney, D. Murdick, D. Rishi, J. Sheehan, Z. Shen, B. Stilson, A.D Wade, K. Wang, C. Wilhelm, B. Xie, D. Raymond, D.S. Weld, O. Etzioni and S. Kohlmeier. Cord-19: The covid-19 open research dataset. ArXiv (2020).

[16] L. Singh, S. Bansal, L. Bode, C. Budak, G. Chi, K. Kawintiranon, C. Padden, R. Vanarsdall, E. Vraga and Y. Wang. A first look at covid-19 information and misinformation sharing on twitter,2020. arXiv preprint arXiv:2003.13907.

[17] F. Alam, F. Dalvi, S. Shaar, N. Durrani, H. Mubarak, A. Nikolov, G. Martino, A. Abdelali, H. Sajjad, K. Darwish and P. Nakov. Fighting the covid-19 infodemic in social media: A holistic perspective and a call to arms,2020. arXiv:2007.07996 [cs.IR].

[18] W. Naude 2020. Artificial intelligence against covid-19: An early review. https://towardsdatascience.com/artificial-intelligence-against-covid-19-an-early-review-92a8360edaba.

[19] J. Bullock, A. Luccioni, K. H. Pham, C. S. N. Lam and M. Luengo-Oroz. Mapping the landscape of artificial intelligence applications against covid-19, 2020.arXiv preprint arXiv:2003.11336.

[20] Y. Mejova , I. Weber and L.Fernandez-Luque . Online health monitoring using facebook advertisement audience estimates in the united states: evaluation study. JMIR public health and surveillance, 2018.doi:10.2196/publichealth.7217.

[21] Y. Nie, H. Chen and M. Bansal. Combining fact extraction and verification with neural semantic matching networks. In Proc. of the AAAI Conference on Artificial Intelligence, volume 33, pages 6859–6866, 2019. doi: 10.1609/aaai.v33i01.33016859.

[22] Y. Madani, M. Erritali, B. Bouikhalene. Using artificial intelligence techniques for detecting Covid-19 epidemic fake news in Moroccan tweets. Results in Physics, Volume 25, 2021. doi:10.1016/j.rinp.2021.104266.

[23] M. K. Elhadad , K. F. Li And F. Gebali. Detecting Misleading Information On Covid-19. In proc. of IEEE Access, vol. 8, pp. 165201-165215, 2020, doi: 10.1109/ACCESS.2020.3022867.

[24] Y. Bang, E. Ishii, S. Cahyawijaya, Z. Ji and P. Fung. Model Generalization on COVID-19 Fake News Detection, 2021.doi:arXiv:2101.03841.

[25] D. Kar, M. Bhardwaj, S. Samanta, A. Azad. No Rumours Please! A Multi-Indic-Lingual Approach for COVID Fake-Tweet Detection, 2020. arXiv:2010.06906.

[26] A. Krizhevsky, I. Sutskever and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Proc. of the 25th International Conference on Neural Information Processing Systems - Volume 1, Pages 1097–1105, 2012. doi: 10.5555/2999134.2999257.

[27] D. Zeng, K. Liu, S. Lai, G. Zhou, J. Zhao. Relation Classification via Convolutional Deep Neural Network. In Proc. of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pages- 2335–2344, 2014. https://aclanthology.org/C14-1220.

[28] Y. Kim. Convolutional neural networks for sentence classification. In Proc. of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages- 1408.5882, 2014. doi:10.3115/v1/D14-1181.

[29] Y. Yang, L. Zheng, J. Zhang, Q. Cui, Z. Li and P. S. Yu. Convolutional Neural Networks for Fake News Detection. 2018. DOI: arXiv:1806.00749.

[30] J. C. S. Reis, A.Correia, F. Murai, A. Veloso and F. Benevenuto. "Supervised Learning for Fake News Detection," in IEEE Intelligent Systems, vol. 34, no. 2, pp. 76-81, March-April 2019, doi: 10.1109/MIS.2019.2899143.

Integration of Ensemble Variant CNN with Architecture Modified LSTM for Distracted Driver Detection

Zakaria Boucetta[1], Abdelaziz El Fazziki[2], Mohamed El Adnani[3]
Computer Science Engineering Laboratory
Faculty of Sciences, Cadi Ayyad University
Marrakech, Morocco

***Abstract*—Driver decisions and behaviors are the major factors in on-road driving safety. Most significantly, traffic injuries and accidents are reduced using the accurate driver behavior monitoring system. However, the challenges occur in understanding human behaviors in the practical environment due to uncontrolled scenarios like cluttered and dynamic backgrounds, occlusion, and illumination variation. Recently, traffic accidents are mainly caused by distracted drivers, which has increased with the popularization of smartphones. Therefore, the distracted driver detection model is necessary to appropriately find the behavior of the distracted driver and give warnings to the driver to prevent accidents, which need to be concentrated as serious issues. The main intention of this paper is to design and implement a novel deep learning framework for driver distraction detection. First, the datasets for driver distraction detection are gathered from public sources. Furthermore, the Optimal Fusion-based Local Gradient Pattern (LGP) and Local Weber Pattern (LWP) perform the pattern extraction of the images. These patterns are inputted into the new deep learning framework with Ensemble Variant Convolutional Neural Network (EV-CNN) for feature learning. The EV-CNN includes three different models, like Resnet50, Inceptionv3, and Xception. The extracted features are subjected to the architecture-optimized Long Short-Term Memory (LSTM). The Hybrid Squirrel Whale Optimization Algorithm (HSWOA) performs both the pattern extraction and the LSTM optimization. The experimental results demonstrate the effective classification performance of the suggested model in terms of accuracy during the detection of distracted driving and are helpful in maintaining safe driving habits.**

Keywords—Distracted driver detection; ensemble variant convolutional neural network; hybrid squirrel whale optimization algorithm; local gradient pattern; local weber pattern; optimal fusion-based pattern descriptors; long short term memory

I. Introduction

The complicated traffic road systems have four different elements like environment, roads, cars, and people. If these four elements are not coordinated, then it leads to road accidents [1]. Recent research reveals that more number of accidents nearly 90% of accidents are caused due to driver error owing to their drunkenness, distraction, and fatigue. Mostly, car accidents highly occur because of the reason of driver distraction. Some of the actions of the drivers like eating or drinking, switching to a favorite radio station and using mobile devices lead to accidents as they only need to focus on the driving towards road without any distraction [2] [3]. Distracted driving mainly causes the wrong lane changes those results in severe traffic accidents. National Highway Traffic Safety Administration (NHTSA) defines distracted driving as “any activity that diverts attention from driving, including talking or texting on the phone, eating and drinking, etc.”. The rapid development of onboard electronics like smartphones and navigation systems are designed as supplementary factors that influence the distraction of the driver [4]. Hence, it is essential for developing in-depth research on determining their occurrence mechanisms, distracted driver attitude, and developing their respective solutions. This can be done for enhancing driving safety and reducing the frequency of distracted driving behaviors [5] [6], [7].

Many research works have been made for identifying the behavior of distracted drivers at different levels. The Gaussian Mixture Model (GMM) model is developed by collecting the parameters of vehicle motion through in-car GPS for judging whether the distraction occurred or not in the driver [7]. The collected vehicle data in the running state is utilized in the existing model based on support vector machine technology and has differentiated them into three classes like high cognitive load distraction, low cognitive load distraction and no distraction [8]. The dynamic parameters regarding the brake baffle position and acceleration baffle position are gathered and involved five diverse machine learning approaches for distinguishing the visual distraction of the driver, lane lateral deviation and longitudinal/lateral acceleration, where the vehicle speed is used in the model. This model achieves an accurate identification of the distraction state of the driver. Recently, the detection approaches of distracted driving relies on the dynamic estimation of vehicle operating parameters like lane deviation, steering wheel angle, acceleration, and vehicle speed for determining the distraction [9] [10]. On the other hand, this methodology requires the installation of additional exclusive data acquisition tools in the vehicle [11]. Concerning the different kinds and specifications, the accuracy and completeness of data acquisition devices are varying which is not beneficial for the application in the practical driving environment.

Various vision and physiological sensors are vastly utilized in monitoring the driver's status. Physiological sensors are commonly limited to estimating particular driving behavior. In order to evaluate the fatigue and somnolence of the driver, EEG and EOG [12] are mostly used in the existing model. However, these sensors are cost-ineffective and need certain prior information like gaze direction. Some approaches are developed for detecting distracted driver behaviors using vision, vehicle driving status, and physiological parameters. According to the development in computing hardware, neural networks, and camera technology, complete features are allowed to be obtained from complex images. Convolutional Neural Networks (CNNs) [13] are commonly used for certain complex tasks in image processing. Recently, different approaches have been developed in the deep learning concept for solving the problems related to image recognition and classification [14] [15] [16]. These approaches are used for extracting the features from the images and utilized for performing the classification. Yet, the distracted behavior is detected by involving a single pre-trained model, which leads to an overfitting problem that further results in detection failures in practical applications. Therefore, a new distracted driver detection model needs to develop using a deep learning approach.

The main contributions of the research work are described as follows:

- To design a new distracted driver detection model with the support of the suggested hybrid optimization algorithm using the developed ensemble learning for feature extraction through the improved deep learning model for distracted driver detection to reduce traffic accidents by warning the distracted drivers.
- To develop an optimal fusion-based pattern extraction approach by combining the extracted patterns of LGP and LWP for maximizing the entropy of the developed HSWOA-based pattern images.
- To integrate an ensemble variant-CNN for extracting more significant features from the ResNet50, Inceptionv3, and Xception networks for ensuring the highly essential features for the proposed model.
- To develop an improved deep learning model named O-LSTM for detecting distracted driver behavior with the help of suggested HSWOA by optimizing certain parameters in LSTM for enhancing the accuracy and precision of the detection in the proposed distracted driver detection model.
- To introduce the hybrid optimization algorithm named HSWOA for optimizing the fusion weights of the optimal fused patterns and for tuning the hidden neurons and number of epochs of the LSTM to achieve accurate detection performance and to improve the overall performance of the proposed model.
- To evaluate the efficiency of the proposed model by comparing it with diverse heuristic algorithms and also with the different classifiers under various quantitative measures.

The rest of the sections in the proposed model are stated as follows. The Section II discusses the exciting works and their challenges. The Section III explains the proposed distracted driver detection model using deep learning approaches. The Section IV conveys the optimal fusion-based pattern extraction of the proposed model. The Section V narrates the ensemble variant-CNN-based feature extraction phase for the proposed model. The Section VI describes the detection phase and suggested HSWOA of the proposed model. The Section VII depicts the achieved results of the proposed distracted driver detection model. The Section VII concludes the developed distracted driver detection model.

II. Literature Survey

A. Related Work

In 2020, Omerustaoglu et al. [17] have developed the distracted driver detection approach based on vision data for improving the generalization ability of the proposed model. The dataset was obtained as the sensor data and driver images that were acquired from the practical drives. The distraction of the driver was identified through the two-stage model with the help of deep learning approaches. This proposed model was validated and shown that proposed model was enhanced the overall performance of both the fused approaches based on the public dataset. In 2021, Huang et al. [18] have implemented a hybrid form of deep learning framework for finding the distracted behaviors of the drivers by processing the image features. The co-operative pre-trained model was developed for enhancing the accuracy while detecting the driving activity in the system with the help of combined deep learning architectures. The simulation analysis has shown that the proposed model has secured better classification accuracy in the detection of distracted driver behavior and also confirmed the potential capability of maintaining the safety driving without any accidents.

In 2021, Jegham et al. [19] have proposed a Kinect sensor for recognizing the safe driver with the multimodal signals from the sensors. Then, the attention-based network was developed for performing the cognitive process towards the deep network by concentrating on the motion and silhouette of the driver that was acquired from the correlated driver scenes. The analysis results have revealed that the proposed model was highly effective in terms of providing high classification accuracy when compared with the conventional methods. In 2020, Ou and Karray [20] have suggested a novel driver distraction detection model using an improved deep learning approach for determining distracted behavior in different driving conditions. Initially, the generative models were designed for generating the images related to diverse driving conditions and then, the discriminative models for classifying the generated images. The collected dataset was correlated with newly generated training samples, and then, the deep learning approach was trained for recognizing the distraction. The suggested framework has proved that the generative models have a high ability to produce driver images under diverse driving conditions.

In 2021, Pal et al. [21] have investigated a deep learning model for classifying and identifying the distracted behavior of drivers to alert them at the right time and also has ensured an effective solution for the problem. The integrated model has divided the driver activity into ten different categories that have also included with the safe driving class. When the distracted driver was observed, the proposed model has detected the event and eliminated the accidents by alerting the drivers. In 2021, Rao et al. [22] have integrated a detection model for identifying the distraction behavior of the driver using the driving images that were captured using an in-vehicle camera. Here, the whitening of the image has occurred for minimizing the correlation and redundancy of the pixel matrix. The analysis results have shown that a higher accuracy was acquired through the proposed model in detecting the distraction behavior when compared with the existing machine learning models.

In 2021, Kumar et al. [23] have proposed a deep learning-based detection model for finding driver distraction behavior. This introduced model has provided superior performance based on the practical environmental inputs, which was determined by comparing it with the other state-of-the-methods. The suggested approaches have secured high scores in all the quantitative measures for classification. In 2021, Dunn et al. [24] have implemented the data from the two different driving studies by incorporating the automatic-equipped vehicles with the help of a driving automation system in order to analyze the driver behaviors. The analysis results have ensured with significant insights of the driving automation systems under diverse operational phases. Here, the overreliance and overtrust on the advanced technologies have reduced certain safety benefits of these proposed systems.

B. Problem Statement

How to correctly identify whether the driver is in a distracted driving state and provide the necessary warnings for the driver to avoid potential safety risks has become one of the most concerning issues. Many techniques have been developed earlier, and some of them are given pros and cons as mentioned in Table I. CNN and LSTM-RNN [17] reduce the overfitting problem and improve the overall accuracy rate. However, it shows the performance loss owing to the viewpoint of the car and camera are not the same as in the public datasets as their collection setup. HCF [18] achieves a good identification rate and also achieves better classification performance. Yet, when the number of training samples is increased, the performance rate can be affected. DSA [19] enhances the classification capability by choosing the highly essential regions from the images. However, it is slightly suffered from high illumination variation particularly focusing at the front view. CNN [20] efficiently reduces the computational time using the feature selection process. Still, it does not be applicable for real-time applications as it needs to be enhancing the effectiveness of the controller. CNN [21] efficiently detects the distracted regions. Yet, it is difficult to train the data and the computation of the network is slow. Deep CNN [22] attains a better accuracy rate at a faster speed. However, more memory is required to train the data. CapsNet [23] attains better accuracy, precision and recall values even with the utilization of less model parameters. On the other hand, it does not consider the detection of drowsiness, which is a significant non-visual feature. VCC NDS and DAF NDS [24] show effective performance on studying the Over-trust and overreliance on driving automation. But, it includes a very less number of samples for investigation. Hence, owing to the consideration of these existing challenges, it is well-determined to develop a new driver distraction model for avoiding traffic accidents caused by drivers.

TABLE I. FEATURES AND CHALLENGES OF EXISTING DISTRACTED DRIVER DETECTION MODELS

Author [citation]	**Methodology**	**Features**	**Challenges**
Omerustaoglu *et al.* [17]	CNN and LSTM-RNN	It reduces the overfitting problem and improves the overall accuracy rate.	It shows the performance loss owing to the viewpoint of the car and camera are not the same as in the public datasets as their collection setup.
Huang *et al.* [18]	HCF	It achieves a good identification rate. It achieves better classification performance.	When the number of training samples is increased, the performance rate can be affected.
Jegham *et al.* [19]	DSA	It enhances the classification capability by choosing the highly essential regions from the images.	It is slightly suffered from high illumination variation particularly focussing at the front view.
Ou and Karray [20]	CNN	It efficiently reduces the computational time using the feature selection process.	It does not be applicable for real-time applications as it needs to be enhancing the effectiveness of the controller.
Pal *et al.* [21]	CNN	It efficiently detects the distracted regions.	It is difficult to train the data. The computation of the network is slow.
Rao *et al.* [22]	Deep CNN	It attains a better accuracy rate at a faster speed.	More memory is required to train the data.
Kumar *et al.* [23]	CapsNet	It attains better accuracy, precision and recall values even with the utilization of less model parameters.	It does not consider the detection of drowsiness, which is a significant non-visual feature.
Dunn *et al.* [24]	VCC NDS and DAF NDS	It shows effective performance on studying the Over-trust and overreliance on driving automation.	It includes a very less number of samples for investigation.

III. Architectural View of Distracted Driver Detection with Data Collection

A. Architectural Description

Before, the classification of a distracted driver is performed in two different ways. In the first way of method, wearable sensors are used for computing physiological and biomedical signals like heart rate, muscular and vascular activities and brain activity. However, the above-mentioned method has certain disadvantages like user involvement and hardware cost. In the second way of method, the cameras are used for classifying distracted drivers, where diverse vision-based approaches are involved for monitoring the distracted behaviors in practice such as fatigue-related features from the driver's face, body postures, gaze and head pose detection. Mostly, the vision-based techniques are incorporated as the two-step structure, where the significant features are obtained from the original data based on the hand-crafted techniques, and next, the classifiers are considered according to the hand-crafted features. If the two-step structure is used in the approach, then the approach was not able to acquire an optimal trade-off among the hand-crafted features and robustness of the trained classifier. The vision-based approaches are involved with the decision trees and Support Vector Machines (SVMs) for detecting distracted drivers, which seeks most of the research. CNNs are generally utilized for complex tasks related to image processing. Deep learning models have been developed rapidly into speech recognition, natural language processing, and computer vision and ensured promising outcomes for solving the distracted driver detection problem. Recently, various deep learning approaches like DenseNet, ResNet, Inception, VGGNet, and AlexNet have been developed for handling the problems related to image recognition and classification. These techniques are used for extracting the significant features from the images and are considered for image classification. But, the recognition of distracted driver behavior based on the single pre-trained model may lead to overfitting issues that further result in detection failure. Therefore, it is required to develop a new distracted driver detection model using ensemble learning approaches that are diagrammatically represented in Fig. 1.

A new distracted driver detection model is developed and processed under various steps for detecting distracted driver behaviors while driving to ensure safe driving and to avoid traffic accidents with the help of deep learning and a hybrid optimization algorithm. The input driver images with both safe driving and distracted behavior images are obtained from publicly available resources. The collected input images were considered for the optimal fusion-based pattern extraction phase. This phase concatenates the pattern extracted from two different pattern extraction approaches like LGP and LWP, where the optimized weights for respective patterns are incorporated for maximizing the entropy of the extracted patterns. The fusion weights are optimized with the help of a suggested hybrid algorithm named HSWOA, which is implemented by combining the features of SSA and WOA. Then, the optimized patterns are given into the developed EV-CNN, in which the deep features are extracted from the pooling layers of the ensemble variants of CNN like ResNet 50, inceptionv3 and xception for obtaining the most accurate features of the proposed model to the detection phase. Further, the extracted features are subjected into the developed O-LSTM, where the suggested HSWOA is used for optimizing the number of hidden neurons and number of epochs of LSTM to improve the accuracy and precision of the proposed model. At last, the distracted driver behaviors are detected using the implemented O-LSTM.

B. Data Collection

The proposed distracted driver detection model collects the required images from the "https://www.kaggle.com/c/state-farm-distracted-driver-detection/data: access date: 2021-12-27". Here, the dataset consists of the driver images of doing some actions such as reaching behind, doing makeup, talking on the phone, eating and texting. The dataset is comprised of ten classes to predict the driver distraction behavior with both the test and train data. The input images of the proposed distracted driver detection model are represented as IM_p^{indriv}, where $p = 1, 2, \ldots, P$ and P show the total number of collected driver images from the dataset. The sample images for ten classes are listed in Fig. 2.

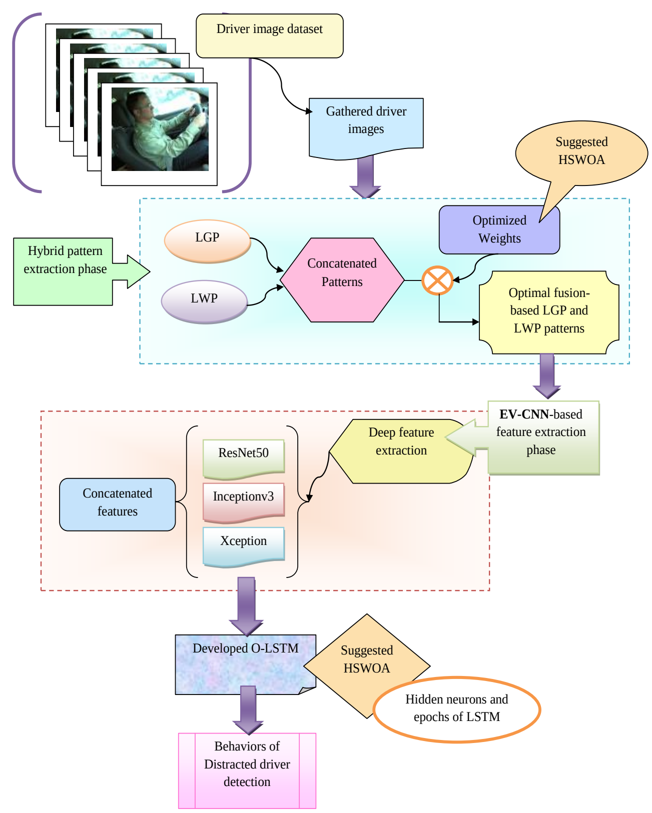


Fig. 1. The Proposed Architecture of a New Distracted Driver Detection Model.

Class description	Image 1	Image 2	Image 3
Class 1: Safe driving			
Class 2: Texting-right			
Class 3: Talking on the phone-right			
Class 4: Texting-left			
Class 5: Talking on the phone-left			
Class 6: Operating the radio			
Class 7: Drinking			
Class 8: Reaching behind			

Class 9: Hair and makeup			
Class 10: Talking to passenger			

Fig. 2. Sample Distracted Driver Images considered for the Proposed Distracted Driver Detection Model

IV. Primary Steps of Distracted Driver Detection: Optimal Fusion Approach

A. Optimal Fusion-based Pattern Extraction

The proposed distracted driver detection model utilizes the LGP and LWD for extracting the patterns from the input images IM_p^{indriv} to attain the accurate performance while detecting distracted drivers while driving. LGP [25] is used for providing the transformation of output image, which is independent of the global intensity variations. This approach is employed for ensuring the facial features but, it is highly sensitive to the local variations along with the edge components in the human face. Similarly, LWP [26] is utilized in the proposed model as it provides superior performance over other descriptors. However, this model considers only four pixels on both the vertical and horizontal directions that are insufficient to represent the local information. Therefore, the pattern extracted from the LGP and LWP is integrated with the optimized weights, respectively using the suggested HSWOA for enhancing the detection performance and to overcome the existing challenges. Here, the optimized fusion weights with the LGP-based pattern is denoted by α, and the optimize weights with the LWP-based pattern is indicated by β. The optimal fusion-based patterns are obtained through in Eq. (1).

$$Pt_z^{of} = \left(Pt_x^{LGP} * \alpha\right) + \left(Pt_x^{lwd} * \beta\right) \tag{1}$$

The main objective FF_1 of optimizing the weights of the extracted patterns is to maximize the entropy of the proposed model that is given in Eq. (2).

$$FF_1 = \underset{\{\alpha,\beta\}}{\arg\max}\left(\frac{1}{EN}\right) \tag{2}$$

Here, the term EN denotes the entropy, which is computed through the Eq. (13). Entropy EN gives "the amount of loss of data in a given input which provides the required information for feature extraction". This is shown in the Eq. (3).

$$EN = \sum_{x=0}^{X-1} Pt_x \log Pt_x \tag{3}$$

The total number of patterns is expressed as X. The entropy needs to be maximized for acquiring the efficient optimized pattern for the distracted driver detection model. Finally, the Optimal fusion patterns are denoted by Pt_z^{of}. The optimal fusion-based pattern extraction with LGP and LWP is represented in Fig. 3.

B. LGP

The input images IM_p^{indriv} are subjected to the LGP technique for extracting the local gradient pattern in the neighboring by involving the intensity gradient profile of the neighborhood measure. Therefore, the LGP value of IM_p^{indriv} is computed for each pixel $\left(a_d, b_d\right)$ using Eq. (4).

$$LGP\left(a_d, b_d\right) = \sum_{i=0}^{I} t\left(h_I - \bar{h}\right)2^i,$$
$$t(a) = \begin{cases} 1, & a \geq 0 \\ 0, & otherwise \end{cases} \tag{4}$$

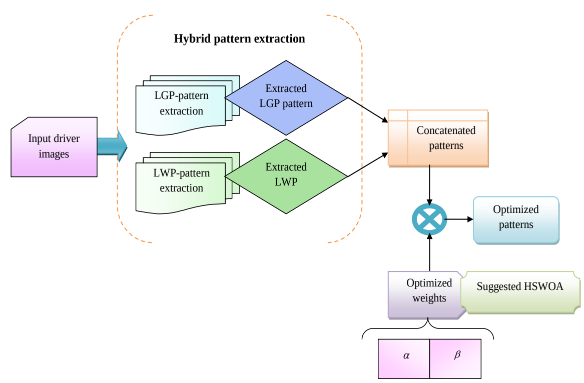


Fig. 3. Optimal Fusion-based Pattern Extraction for the Proposed Distracted Driver Detection Model.

Here, the term h_I represents the gradient of the neighboring pixel I and the average gradient of the neighboring pixels are denoted as $\bar{h}$ that are calculated as given in Eq. (5) and Eq. (6).

$$h_I = \left| T_I - T_C \right| \tag{5}$$

$$\bar{h} = \left(\frac{1}{8}\right)\sum_{i=0}^{I} h_I \tag{6}$$

The extracted patterns are obtained from the LGP technique that is expressed as Pt_x^{LGP}.

C. LWP

The proposed distracted driver detection model employs the input images IM_p^{indriv} for extracting the patterns to determine the distracted driver. LWP is the pattern extraction technique by determining the ratio of variations present in the pixel intensity. This is said to be stimulus information under the visual perception. Differential excitation computes “the ratio of change in pixel intensity between center pixels against its neighbors”. It identifies the local salient visual patterns. The differential excitation is estimated at central pixel $\varepsilon(y_s)$ using Eq. (7).

$$\varepsilon(y_s) = \arctan\left\{\sum_{j=0}^{q-1}\left(\frac{y_j - y_s}{y_s}\right)\right\} \tag{7}$$

Here, the term q denotes the number of neighbors on a circle of radius B and y_s indicates the intensity value of the central pixel. If $\varepsilon(y_s)$ is determined as positive, then it influences the surroundings pixel to be lighter than the present pixel. In contrast, if $\varepsilon(y_s)$ is determined to be negative, it influences the surrounding pixels are darker than the present pixel. The orientation component of LWD is estimated through Eq. (8).

$$\theta(y_s) = \arctan\left\{\frac{y_{\left(\frac{q}{2}-\mathrm{B}\right)} - y_{(\mathrm{B})}}{y_{(q-\mathrm{B})} - y_{\left(\frac{q}{2}-\mathrm{B}\right)}}\right\} \tag{8}$$

The extracted patterns using LWD are denoted as Pt_x^{lwd}.

V. Ensemble Learning-based Feature Extraction for Distracted Driver Detection

A. Ensemble Learning Model

The proposed distracted driver detection model extracts diverse features from the optimized patterns Pt_z^{of} for performing efficient detection performance in the developed model using the suggested HSWOA. CNN-based techniques [27] are selected in the proposed model as it provides better recognition features with the utilization of its more layers and it is independent of the human efforts when implementing its functionalities. Then, the extracted deep features from the pooling layer of the ensemble variant model, where the ResNet50 [28], inceptionv3 [29] and xception [29] approaches are used. ResNet50 in the proposed model contains huge number of layers, which can be trained without enhancing the training error percentage. Similarly, inceptionv3 in the proposed model gains high performance and effectively utilizes the computing resources with less computation load. Xception provides superior performance when utilizing in the larger image classification dataset and ensures higher computational efficiency. The significant features are obtained from the ensemble variant model from the extracted deep features for the distracted driver detection phase. Finally, the extracted features from the three approaches are concatenated and put forwarded to the O-LSTM-based detection. The developed EV-CNN-based feature extraction is depicted in Fig. 4.

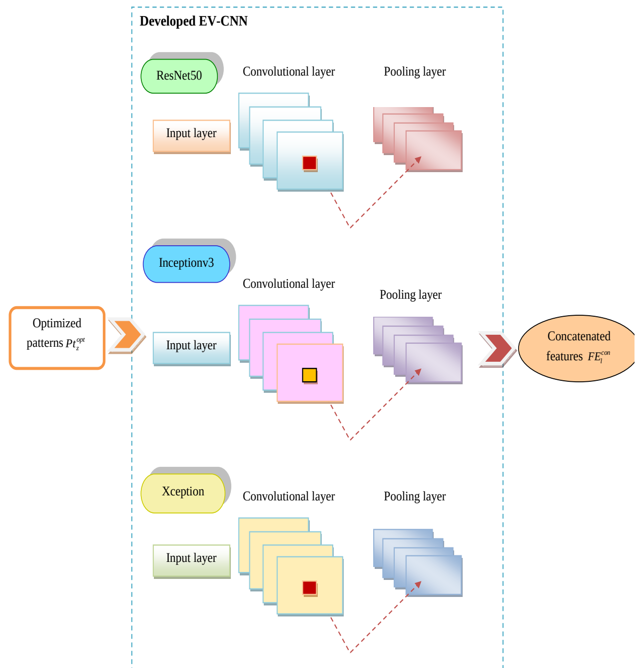


Fig. 4. Feature Extraction of Proposed Distracted Driver Detection Model using EV-CNN.

B. ResNet Model

The proposed distracted driver detection model uses the ResNet50 technique for extracting the significant features from the optimal fusion-based pattern images. The ResNet [30] ensures improved performance even when more number of images is given into it. This model is developed using the skipping connections that is made on the two to three layers with ReLU and batch normalization. The residual learning is applicable to the multiple layers of architecture. The residual block of the ResNet is computed using the Eq. (9).

$$Ol = J\left(in, U + in\right) \tag{9}$$

Here, the term Ol denotes the output layer, in shows the input layer, and the residual map are represented by the function J. The residual block on the ResNet is processed when the dimensions of the input layer and output layer are the same. Additionally, each block of the ResNet contains two or three layers. The first two layers of the ResNet express the "GoogleNet by doing convolution 7×7" and max-pooling at the size of 3×3. The extracted features from the pooling layer of ResNet50 are represented as FE_g^{RES} that is computed with the number of 400 features.

C. Inception Model

The extracted deep features DF_f^{cnn} are used for obtaining the essential features from Inceptionv3 approach for the proposed distracted driver detection model. Inception model is used for clustering the sparse convolution kernel structure into the diverse dense sub-convolution kernel combinations. Convolutional filters with various sizes are employed for acquiring the diverse receptive fields. The training procedure of the inception is described initially with the input deep features into the three convolution layers of the inception model. Three convolution layers are performed with kernels of 3×3 along with the one max pooling layer for extracting the low latitude features from the input deep features. Further, two convolution layers are used with the help of kernels of 1×1 and 3×3 and also with the one max pooling layer are utilized for extracting the features. Then, the three pooling layer is used for obtaining the high-dimensional features from the input deep features. The extracted features from the pooling layer of inception model are counted to be 400, which are indicated by FE_h^{ine}.

D. Xception Model

The proposed distracted driver detection model uses the xception technique for extracting the significant features from the deep features DF_f^{cnn}. Xception model employs 36 convolutional layers for performing the feature extraction based on the CNN. These convolutional layers are sorted accordingly into the 14 modules, which are covered by the linear residual connections. The input deep features DF_f^{cnn} are considered to the Convolutional kernels for extracting the features. The output of the convolution layer is computed using Eq. (10).

$$B_P = f\left(CK_P * B_{P-1} + c_P\right) \tag{10}$$

Here, the term B_P denotes the output of P^{th} convolution layer; the activation function is indicated by $f(\cdot)$, the convolution kernel is represented by CK_P, convolution operation is denoted by "$*$" and the offset parameter is shown by c_P. The feature maps are given into the separable convolution for feature extraction. Here, the separable convolution is utilized for minimizing the parameters count and computation complexity in the model. The extracted features from the pooling layer of the Xception model are counted to be 400 that are denoted by FE_k^{xcep}.

Finally, the extracted features from ResNet50, Inceptionv3 and Xception are concatenated into a single set of features, which is expressed by $FE_l^{con} = FE_g^{RES} + FE_h^{ine} + FE_k^{xcep}$, where $l = 1, 2, \ldots, L$ and L denotes the total number of concatenated features for the proposed distracted driver detection model.

VI. Hybrid Meta-heuristic-based Distracted Driver Detection

A. LSTM-based Detection Model

The concatenated features FE_l^{con} are utilized in the O-LSTM [31] for classifying distracted driver behaviors to avoid traffic accidents. The classification phases are made stronger by the recurrent structures of the deep learning algorithm and so, the external memories are needless for storing the output. The recurrent structures of the LSTM classifiers provide less complexity in computation. The four components such as "cells, input gate, output gate and forget gate" are presented in the LSTM network. The cell carries the data and passed to the input and output gate. The forget gate is initially used for determining the information passed through the network that is shown in Eq. (11).

$$c_t = \sigma\left(B_c \cdot \left[k_{t-1}, \left(FE_l^{con}\right)_t\right] + w_c\right) \tag{11}$$

Here, the terms σ and k_t are correspondingly shown as sigmoid activation function and the output of the hidden state. The weight matrices are described as B_c, B_g, B_h, B_q and the input variable is given as FE_l^{con}. Then, the cell output, output gate and forget gate are represented as g_t, q_t and G_t, respectively. The biased values of these gates are portrayed as w_c, w_f, w_h, w_q. The input gate is formulated in Eq. (12).

$$g_t = \sigma\left(B_g \cdot \left[k_{t-1}, \left(FE_l^{con}\right)_t\right] + w_f\right) \tag{12}$$

Further, it updates a new cell states using sigmoid function that generates the new vector $\widehat{G}_t$ that is shown in Eq. (13).

$$\widehat{G}_t = \tan k\left(B_h \cdot \left[k_{t-1}, \left(FU_{g^*}^{opt}\right)_t\right] + w_h\right) \tag{13}$$

For updating the old cell into new cell, the earlier state is integrated with forget gate and added more parameters that is given in Eq. (14).

$$G_t = g_t * G_{t-1} + c_t * \widehat{G}_t \tag{14}$$

Finally, the output gate provides the cell state using the output of the sigmoid of output gates that are given in Eq. (15) and Eq. (16).

$$q_t = \sigma\left(B_q \cdot \left[k_{t-1}, \left(FE_l^{con}\right)_t\right] + w_q\right) \tag{15}$$

$$k_t = q_s * \tan k\left(d_t\right) \tag{16}$$

The sigmoid activation function of the LSTM classifier is represented as σ with hyperbolic tangent $\tanh$. The distracted driver behaviors are detected using the developed O-LSTM model that is depicted in Fig. 5.

B. Objective Model

The detected distracted behavior of the driver shows the accurate detection outcomes based on the developed O-LSTM with the suggested HSWOA. The detected results of the proposed distracted driver detection model are obtained as the various classes of safe driving and distracted driving behaviors. The objective function of the proposed distracted driver detection model is to maximize the detection accuracy and precision that is given in the Eq. (17).

$$Ofun_{(2)} = \underset{\left\{HN_{lstm}^{hn}, Ep_{lstm}^{ep}\right\}}{\arg\min}\left(\frac{1}{ay + pr}\right) \tag{17}$$

Here, the term HN_{lstm}^{hn} and Ep_{lstm}^{ep} are represented as hidden neurons and the number of epochs of the LSTM, respectively. The developed HSWOA optimizes the number of hidden neurons in the range of [5,255] and number of epochs in the interval of [5, 20]. Accuracy ay is measured as the "closeness of the measurements to a specific value" as given the Eq. (18).

$$accy = \frac{\left(t^p + t^n\right)}{\left(t^p + t^n + f^p + f^n\right)} \tag{18}$$

Here, the true positive and true negative values are shown as t^p and t^n, respectively and false positive and false negative values are given as f^p and f^n, respectively. Precision pr is explained as "the fraction of relevant instances among the retrieved instances" as given in Eq. (19).

$$pr = \frac{t^p}{t^p + f^p} \tag{19}$$

The solution encoding of the suggested HSWOA is shown in Fig. 6.

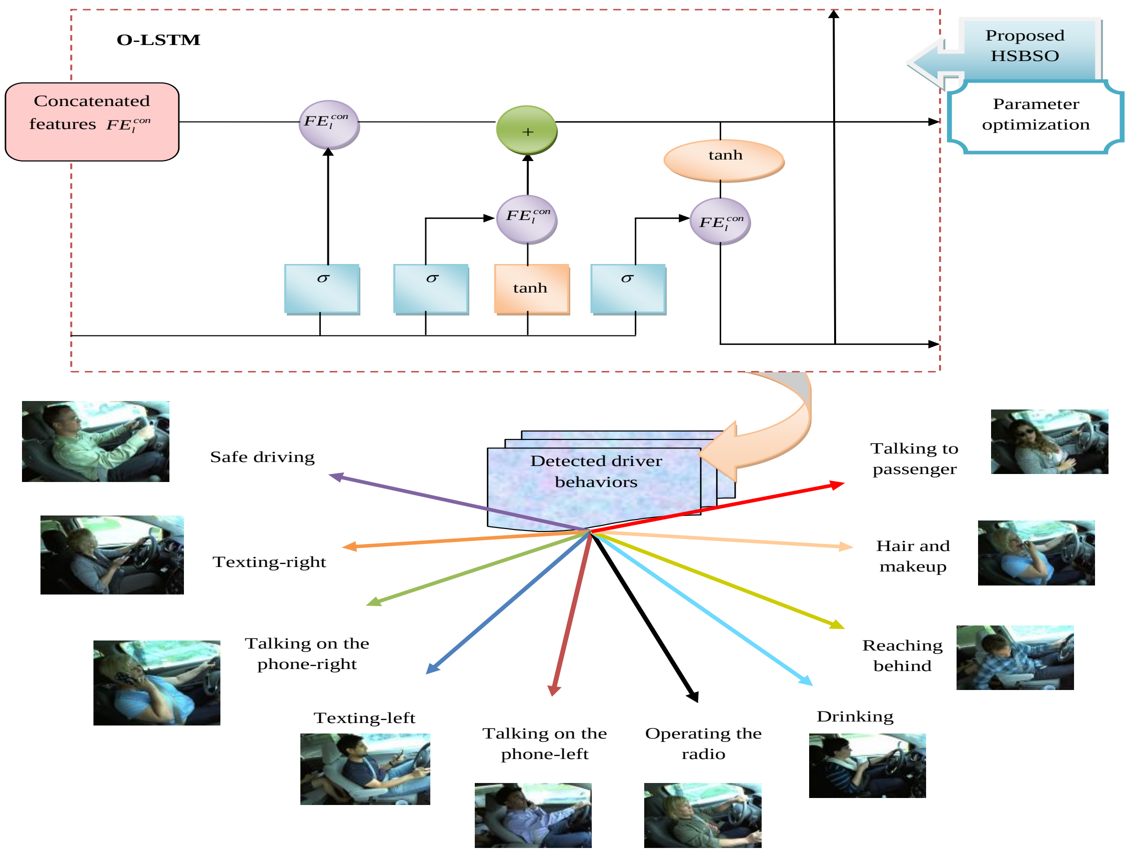


Fig. 5. Proposed O-LSTM-based Distracted Driver Detection Model

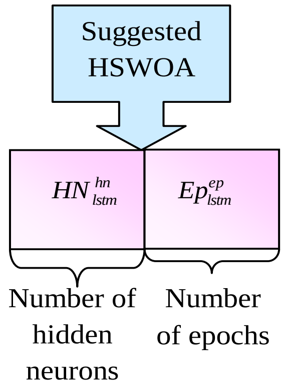


Fig. 6. Solution encoding of the Proposed Distracted Driver Detection Model using the Developed HSWOA.

C. Proposed HSWOA

The proposed distracted driver detection model employs the suggested HSBSO for reducing the training complexity in LSTM by optimizing its hidden neurons and epochs and also tuning the weights of the extracted deep features from the developed EV-CNN. WOA [31] algorithm is chosen for the proposed model due to the ability to solve the real-time optimization problems, which is also simple and easy to implement. But, it does not contain the capability to balance the exploration and exploitation phase. Also, it is easily falls into the local optimum problem. To overcome these challenges of WOA, a meta-heuristic algorithm named SSA [32] is adopted into WOA. Thus, the HSBSO algorithm is developed by adopting the features of SSA into the WOA algorithm. The SSA algorithm is capable of generating the optimal solution for high-dimension optimization problems at the limited time period. In the proposed HSBSO, two random variables c and d are introduced, which are determined by the fitness-based computations that are shown in Eq. (20) and Eq. (21).

$$c = abs\left(f\left(j \right) - bestfit \right) \tag{20}$$

$$d = abs\left(worstfit - f\left(j \right) \right) \tag{21}$$

Here, the term $f\left(j \right)$ denotes the fitness of the current solution and the best fitness value and worst fitness value are denoted by $bestfit$ and $worstfit$, respectively. If the condition $\left(c > d \right)$ is satisfied, then the position update takes place using the SSA otherwise, the position update takes place according to the WOA algorithm.

SSA [26] is implemented based on the motivation of the jumping mechanism and gliding strategies of the flying squirrels. The entire optimization process was carried out in the summer and winter phase. Initially, considered the count of the population is PP and the upper bound and lower bound of the search space are regarded as z_U and z_L. Every individual in the population is generated according to Eq. (22).

$$z_a = z_L + rnd\left(1, dd\right) \times \left(z_U - z_L \right) \tag{22}$$

Here, the term rnd denotes the random number at the range of 0 to 1, z_a indicates the a^{th} individual and the dimension of the search space is expressed by dd. The SSA algorithm needs only one squirrel at the tree and so, equal number of trees and squirrels are present in the search space. Here, the squirrels are categorized based on the fitness of the population into three diverse varieties such as individuals located at hickory trees f_{hk} with the minimum fitness, individuals located at acorn trees f_{ac} with the ranking of second fitness and individuals located at normal trees f_{nt}. Then, the foraging behavior is designed mathematically as follows.

The position of the flying squirrels in the acorn nut trees f_{ac} that jumps to hickory nut tree. Here, the new position of the squirrels is computed in Eq. (23).

$$z_{ac}^{new} = \begin{cases} z_{ac}^{old} + rd_G GC\left(z_{hk}^{old} - z_{ac}^{old} \right) & if\ r_1 \geq PR_{pd} \\ random\ position & others \end{cases} \tag{23}$$

Here, the gliding constant is indicated by GC and the random function r_1 is determined from the interval $\left[0,1\right]$. The random gliding distance is shown by rd_G that is computed in Eq. (24).

$$rd_G = \frac{hh_G}{\tan\left(\phi\right) \times cv} \tag{24}$$

The terms cv and hh_G are considered as the constant values and the gliding angle $\tan\left(\phi\right)$ is calculated in Eq. (25).

$$\tan\left(\phi\right) = \frac{F}{M} \tag{25}$$

The drag force is represented by F and the lift force is expressed by M and these two forces are computed as follows.

$$F = \frac{1}{2\rho v^2 s C_{dd}} \tag{26}$$

$$M = \frac{1}{2\rho v^2 s C_{ll}} \tag{27}$$

Here, the terms ρ, v, s, C_{ll} and C_{dd} are indicated as the constants. Several numbers of squirrels are located on the normal trees that are migrated to the acorn nut trees for getting food resources, where the new position is decided by Eq. (28).

$$z_{nt}^{new} = \begin{cases} z_{nt}^{old} + rd_G GC\left(z_{ac}^{old} - z_{nt}^{old} \right) & if\ r_2 \geq PR_{pd} \\ random\ position & others \end{cases} \tag{28}$$

Here, the term r_2 denotes the random function that is determined from the interval $\left[0,1\right]$. Similarly, some squirrels in the normal tree are changed its location to the hickory nut tree, in which the new position of the squirrel is formulated in Eq. (29).

$$z_{nt}^{new} = \begin{cases} z_{nt}^{old} + rd_G GC\left(z_{hk}^{old} - z_{nt}^{old}\right) & if\ r_3 \geq PR_{pd} \\ random\ position & others \end{cases} \tag{29}$$

Here, the term r_3 indicates the random function that is determined from the interval $\left[0,1\right]$.

Estimate the seasonal monitoring condition: This step is to avoid the local minima problem. Here, the seasonal constant SE and the respective minimum values are correspondingly computed in Eq. (30) and Eq. (31).

$$SE_c^m = \sqrt{\sum_{t=1}^{q}\left(z_{ac,t}^m - z_{hk,t}\right)^2}, m = 1,2,3 \tag{30}$$

$$SE_{cMIN} = \frac{10E-6}{365^{\frac{j}{\left(j_{\max}\right)} / 2.5}} \tag{31}$$

Here, the number of iterations is indicated by j and maximum number of iterations is denoted by $j_{\max}$. The season is determined through the condition $SE_c^m < SE_{cMIN}$. When $SE_c^m < SE_{cMIN}$ is satisfied, the winter season is started or else the season will be unchanged. When the summer season arises, all individuals update their position to f_{hk} that is shown in Eq. (32).

$$z_{nt}^{new} = z_L + LEVY\left(nn\right) \times \left(z_U - z_L\right) \tag{32}$$

The levy distribution is known to be strong tool for enhancing the global exploration capability for most of the optimization algorithms that is given in Eq. (33).

$$LEVY\left(y\right) = 0.01 \times \frac{dr_p \times \sigma}{\left|dr_q\right|^{1/\beta}} \tag{33}$$

Here, the two different functions denoted by dr_p and dr_q that is in the range of $\left[0,1\right]$ and the constant is indicated by β and also the term σ is computed through Eq. (34).

$$\sigma = \left(\frac{\Gamma\left(1+\beta\right) \times \sin\left(\frac{\pi\beta}{2}\right)}{\Gamma\left(\frac{\left(1+\beta\right)}{2}\right) \times \beta \times 2^{\left(\frac{\left(\beta-1\right)}{2}\right)}}\right) \tag{34}$$

Finally, the algorithm stops when the maximum count of iterations is fulfilled.

WOA [31] is encouraged by the hunting behavior of humpback whales. These whales usually catch the pack of small fishes in their close surfaces for their food sources. This foraging happens by generating various bubbles in the circular direction around their prey. This unique behavior is known to be “bubble-net feeding method of humpback whales”.

1) “Encircling prey”: Every whale is indicated as the search agent. The abilities of the humpback whales are intended to identify the prey location and encircle them. WOA is initiated the searching without knowing the position of the optimal solution in the search space and then, the target food source location is considered as the current optimal solution. If the best candidate is identified, then all other candidates will update their position based on the best solution. The encircling behavior of the whales is represented as given in the Eq. (35).

$$\vec{B} = \left|\vec{A} \cdot \vec{z}^*\left(j\right) - \vec{z}\left(j\right)\right| \tag{35}$$

$$\vec{z}\left(j+1\right) = \vec{z}^*\left(j\right) - \vec{C} \cdot \vec{B} \tag{36}$$

Here the current iteration among the search agent is given as j, the position of the current solution and the best solution are defined as z and z^*, respectively and the coefficient vectors are denoted as $\vec{C}$ and $\vec{A}$ that are formulated in the Eq. (37) and Eq. (38).

$$\vec{C} = 2 \cdot \vec{b} \cdot \vec{V} - \vec{b} \tag{37}$$

$$\vec{A} = 2 \cdot \vec{V} \tag{38}$$

Here, the random vector is given as V that lies in the interval of $\left[0,1\right]$ and the convergence factor is shown as b, which simultaneously reduces from 2 to 0 with maximizing the number of iterations.

2) “Bubble net attacking method”: This hunting behavior of whales is classified into two phases namely “shrinking and encircling of the prey” and “spiral upward encirclement and suppression”. The shrinking and encircling method is attained by decreasing the value of $\vec{V}$. If $\left|C\right| \leq 1$ satisfied, then the new position is updated for the search agent using the Eq. (36). The spiral-based position updating takes place for reaching their food destination which is equated in Eq. (39).

$$\vec{z}\left(j+1\right) = \vec{B}' \cdot e^{sr} \cdot \cos\left(2 \cdot \pi \cdot r\right) + \vec{z}^*\left(j\right) \tag{39}$$

Here, the random number is denoted as r in the range of $\left[-1,1\right]$ and the constant s is utilized for indicating the shape of logarithmic spiral. Term $\vec{B}' = \left|z^* - z\right|$ shows the distance between the search agent and best solution. WOA selects the hunting mechanism among these two behaviors by assuming the same probability for updating the positions of the search agents that is shown in the Eq. (40).

$$\vec{z}\left(j+1\right)=\begin{cases}\vec{z}^{*}\left(j\right)-\vec{C}\cdot\vec{B} & if\ P<0.5\\ \vec{B}^{'}\cdot e^{sr}\cdot\cos\left(2\cdot\pi\cdot r\right)+\vec{z}^{*}\left(j\right) & if\ P\geq 0.5\end{cases} \tag{40}$$

3) Searching for prey: WOA contains the ability to balance the "exploitation and exploration phase" due to the size of vector C. If $C\geq\left|1\right|$ satisfied, then the position of search agents is updated through the random vector for establishing the global optimal solution by reducing the local minima. This is depicted in the Eq. (41).

$$\vec{B}=\left|\vec{A}\cdot\vec{z}_{RN}-\vec{R}\right| \tag{41}$$

$$\vec{z}\left(j+1\right)=\vec{z}_{RN}-\vec{C}\cdot\vec{B} \tag{42}$$

Here, the random vector is indicated as z_{RN} and the best optimal solution that is position of prey is identified as z^{*}. The pseudo code of the proposed HSBSO is given in Algorithm 1.

Algorithm 1: Proposed HSBSO

Initialize the population along with its parameters

Evaluate the fitness of every individuals

Compute two variables c and d using Eq.(20) and Eq. (21), respectively

Determine the best and worst fitness values

While (until reaching the stopping criterion) do

If $\left(c>d\right)$

Solution updating based on SSA

Else

Solution updating based on WOA

End

End while

Obtain the best optimal solution

End

The hybrid optimization algorithm is used for enhancing the overall performance of the proposed model especially in the pattern extraction phase and detection phase, which reduces the overall burden of the proposed model. The flowchart of the proposed HSWOA is depicted in Fig. 7.

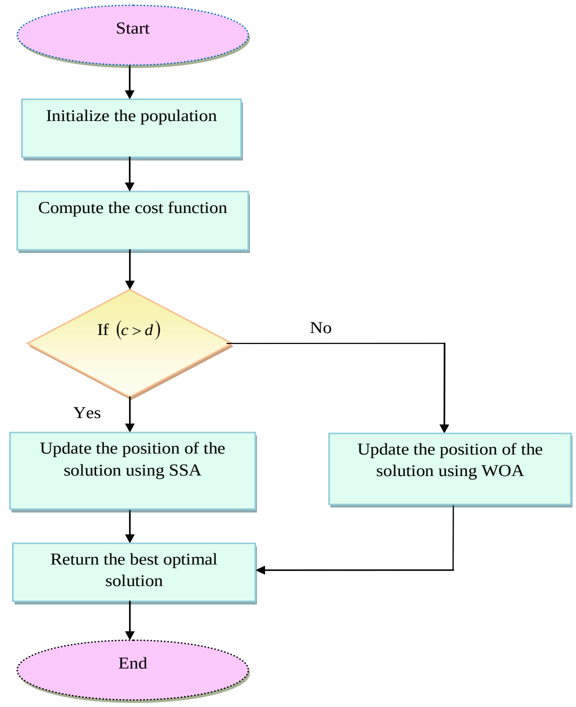


Fig. 7. Proposed HSWOA for Distracted Driver Detection Model.

VII. Result and Discussion

A. Experimental Setup

The platform used for implementing the proposed distracted driver detection model was Python and certain experimental analysis was performed for testing the suggested model using several quantitative measures. The experimental analysis was mainly focused on comparing the performance between the proposed model and conventional meta-heuristic algorithms and also with different classifiers. The involved quantitative measures were classified as positive or Type I measures and negative measures or Type II measures. This experimental analysis was undergone with a population count as 10 and maximum iterations count as 10 for the proposed distracted driver detection model. The proposed HSWOA was compared with other meta-heuristic algorithms like "Particle Swarm Optimization (PSO) [33], Grey Wolf Optimizer (GWO) [34], SSA [32] and WOA [31] and machine learning algorithms like Neural Network (NN) [30], Convolutional Neural Network (CNN) [27], Long Short Term Memory (LSTM) [35], and Ensemble Learning Model (ELM) [17]".

B. Performance Metrics

The performance of proposed distracted driver detection model is evaluated using various quantitative measures that are given as follows.

1) MCC (mc) is "a measure of the quality of binary classifications of testing" as given in the Eq. (43)

$$mc=\frac{t^{P}\times t^{N}-f^{P}\times f^{N}}{\sqrt{\left(t^{P}+f^{P}\right)\left(t^{P}+f^{N}\right)\left(t^{N}+f^{P}\right)\left(t^{N}+f^{N}\right)}} \tag{43}$$

2) Specificity (spt) is “the proportion of negatives that are correctly identified” as represented in the Eq. (44):

$$spt = \frac{t^N}{t^N + f^P} \tag{44}$$

3) NPV (nv) is described as “the sum of all persons without disease in testing” as denoted in Eq. (45):

$$nv = \frac{t^N}{t^N + f^N} \tag{45}$$

4) F1-score (F1) is determined as “the measurement of the accuracy in the conducted test” as given in Eq. (46):

$$F1 = 2 \times \frac{2t^P}{2t^P + f^P + f^N} \tag{46}$$

5) FDR (fdr): is “a method of conceptualizing the rate of errors in testing when conducting multiple comparisons” as denoted in Eq. (47):

$$fdr = \frac{f^P}{f^P + t^P} \tag{47}$$

6) Sensitivity (sen): is “the proportion of positives that are correctly identified” as denoted in Eq. (48):

$$sen = \frac{t^P}{t^P + f^N} \tag{48}$$

7) FPR (frp): is defined as “the ratio between the numbers of negative events wrongly categorized as positive (false positives) and the total number of actual negative events” as given in Eq. (49):

$$fpr = \frac{f^P}{f^P + t^N} \tag{49}$$

8) FNR (fnr): is “the proportion of positives which yield negative test outcomes with the test” as given in Eq. (50):

$$fnr = \frac{f^N}{f^N + t^P} \tag{50}$$

C. *Optimal Fusion-based Pattern Extracted in the Proposed Distracted Driver Detection Model*

The resultant image of the optimized patterns from the LGP and LWP in the proposed model is given in Fig. 8.

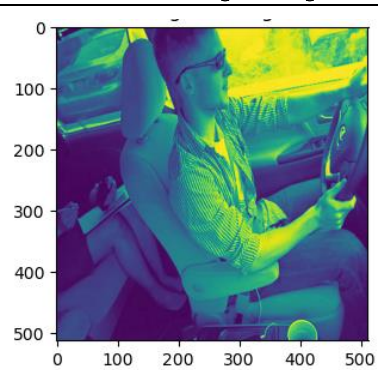
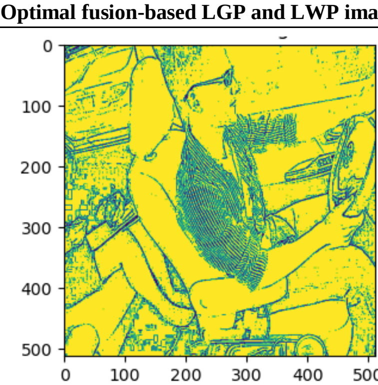
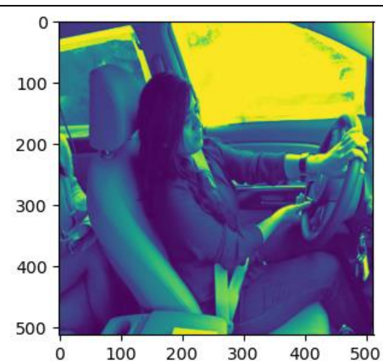
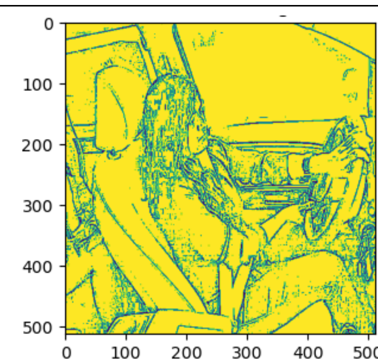
Image description	Original images	Optimal fusion-based LGP and LWP images
Image 1		
Image 2		

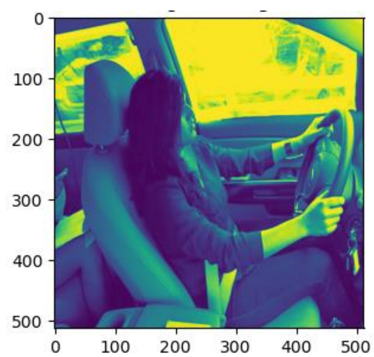
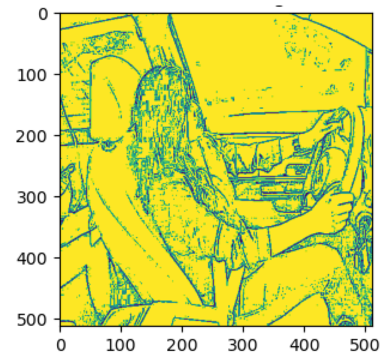
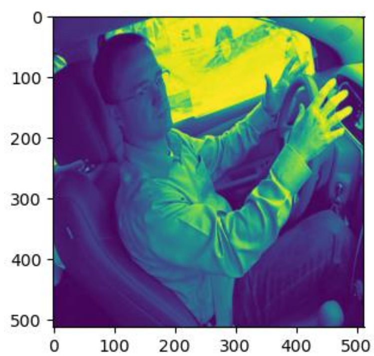
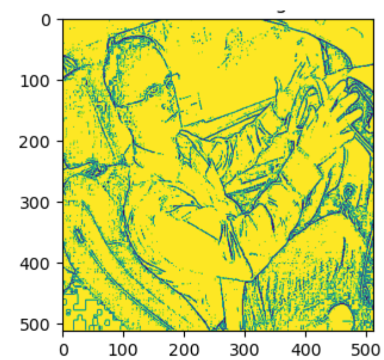
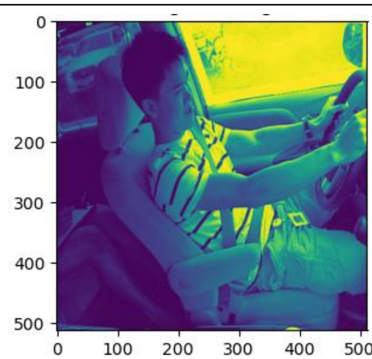
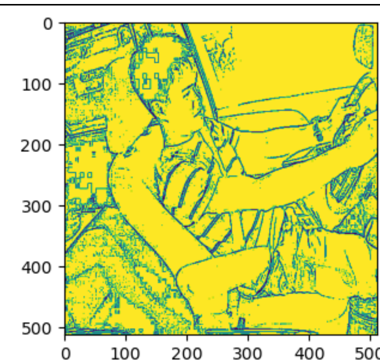
Image 3		
Image 4		
Image 5		

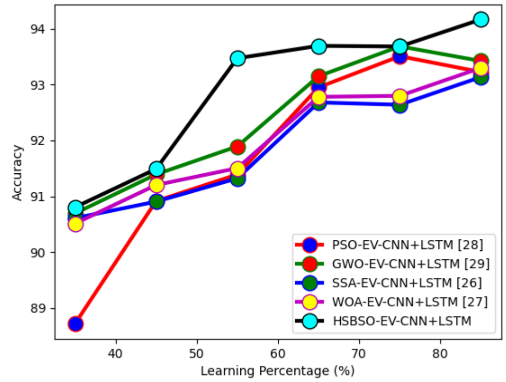
Fig. 8. Original and Optimal Fusion-based LGP and LWP Images of the Safe Drivers and Distracted Drivers.

D. Performance Analysis on Accuracy based on different Meta-heuristic Algorithms

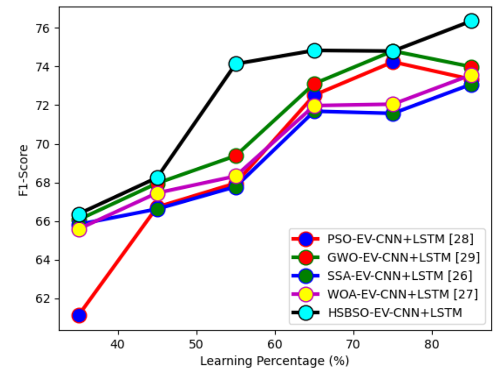
The performance analysis on accuracy for the proposed distracted driver detection model is tested by comparing with the different meta-heuristic algorithms as shown in Fig. 9 at varying learning percentages. The performance of the proposed HSWOA-EV-CNN+LSTM shows 0.21%, 0.64%, 0.53%, and 0.21% improved accuracy than the PSO-EV-CNN+LSTM, GWO-EV-CNN+LSTM, SSA-EV-CNN+LSTM and WOA-EV-CNN+LSTM, respectively at the learning percentage as 70 on dataset 2. While observing the accuracy and F1-Score of the proposed model, it shows a slight low performance at the iteration of 40, which has been increased at the iterations of 50, 60 and 70 that performance is maintained with a slight increase in performance until the iteration of 80. Thus, the overall performance of the proposed algorithm for the developed distracted driver detection model secures enhanced performance than other existing methods.

E. Performance Analysis on Proposed Model with different Classifiers

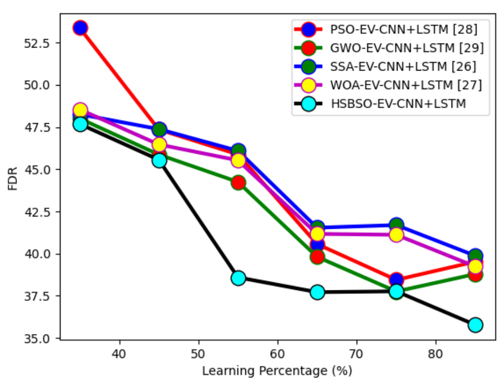
The performance analysis for the proposed distracted driver detection model is evaluated by comparing with different classifiers is depicted in Fig. 10 at varying learning percentages. The performance of the proposed HSWOA-EV-CNN+LSTM contains 6.66%, 6.66%, 7.86% and 4.34% higher accuracy than the NN, CNN, LSTM and ELM, respectively at the learning percentage of 35 on dataset 3. When considering the FNR of the proposed model, it ensures an increased performance rate on increasing the number of iterations. In the MCC analysis on the proposed model, it secures 0.6 values at iteration 35 that is increased to 0.7 at the iteration of 55 and in the further two iterations of 65 and 75, the MCC value has been enhanced than in the prior iterations. Thus, the overall performance of the proposed HSWOA-EV-CNN+LSTM for the proposed distracted driver detection model shows higher performance than other conventional methods of distracted driver detection.


Accuracy
Learning Percentage (%)
PSO-EV-CNN+LSTM [28]
GWO-EV-CNN+LSTM [29]
SSA-EV-CNN+LSTM [26]
WOA-EV-CNN+LSTM [27]
HSBSO-EV-CNN+LSTM

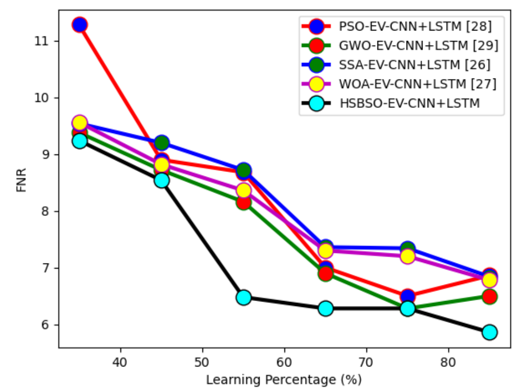
(a)


F1-Score
Learning Percentage (%)
PSO-EV-CNN+LSTM [28]
GWO-EV-CNN+LSTM [29]
SSA-EV-CNN+LSTM [26]
WOA-EV-CNN+LSTM [27]
HSBSO-EV-CNN+LSTM

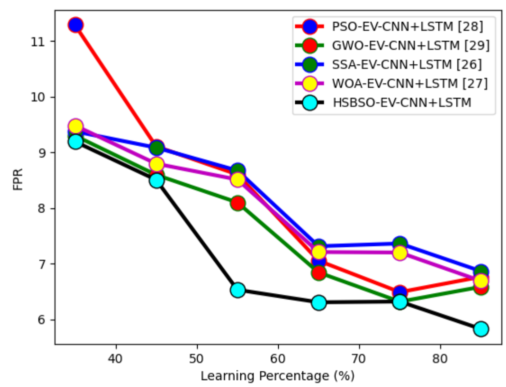
(b)


FDR
Learning Percentage (%)
PSO-EV-CNN+LSTM [28]
GWO-EV-CNN+LSTM [29]
SSA-EV-CNN+LSTM [26]
WOA-EV-CNN+LSTM [27]
HSBSO-EV-CNN+LSTM

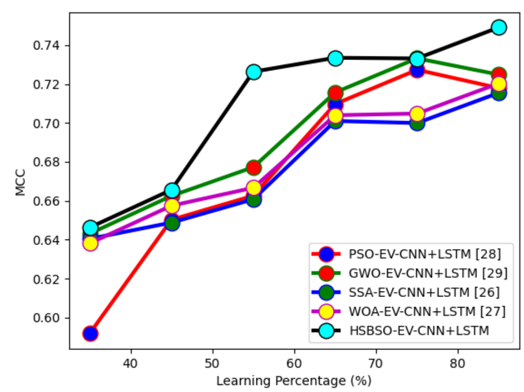
(c)


FNR
Learning Percentage (%)
PSO-EV-CNN+LSTM [28]
GWO-EV-CNN+LSTM [29]
SSA-EV-CNN+LSTM [26]
WOA-EV-CNN+LSTM [27]
HSBSO-EV-CNN+LSTM

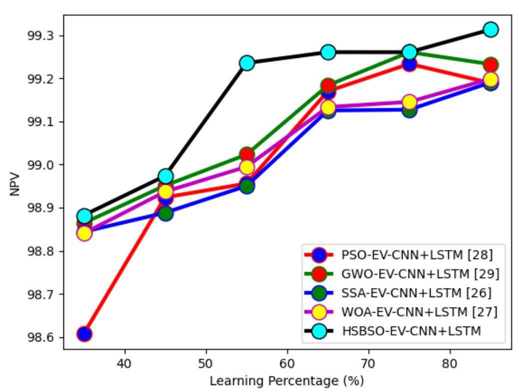
(d)


FPR
Learning Percentage (%)
PSO-EV-CNN+LSTM [28]
GWO-EV-CNN+LSTM [29]
SSA-EV-CNN+LSTM [26]
WOA-EV-CNN+LSTM [27]
HSBSO-EV-CNN+LSTM

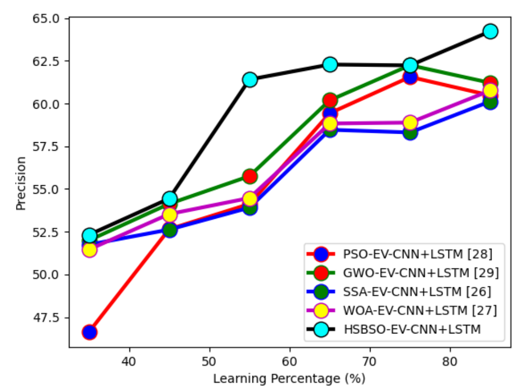
(e)


MCC
Learning Percentage (%)
PSO-EV-CNN+LSTM [28]
GWO-EV-CNN+LSTM [29]
SSA-EV-CNN+LSTM [26]
WOA-EV-CNN+LSTM [27]
HSBSO-EV-CNN+LSTM

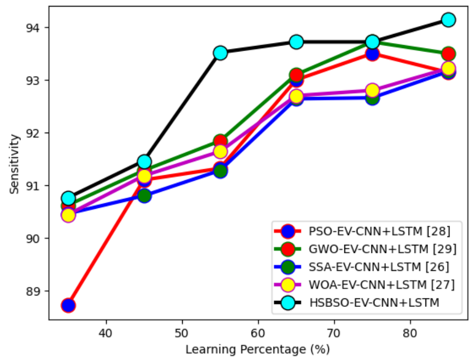
(f)


NPV
Learning Percentage (%)
PSO-EV-CNN+LSTM [28]
GWO-EV-CNN+LSTM [29]
SSA-EV-CNN+LSTM [26]
WOA-EV-CNN+LSTM [27]
HSBSO-EV-CNN+LSTM

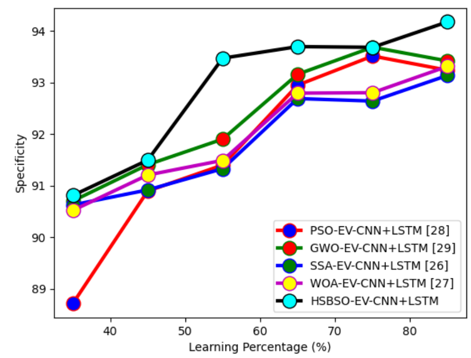
(g)


Precision
Learning Percentage (%)
PSO-EV-CNN+LSTM [28]
GWO-EV-CNN+LSTM [29]
SSA-EV-CNN+LSTM [26]
WOA-EV-CNN+LSTM [27]
HSBSO-EV-CNN+LSTM

(h)

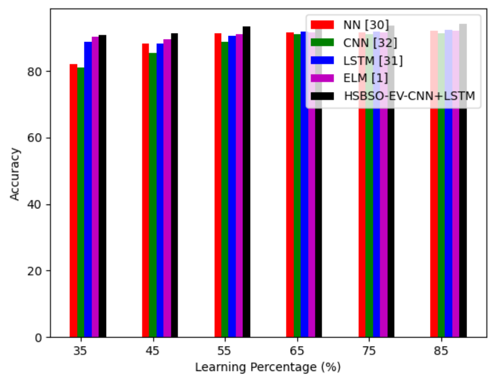


(i)

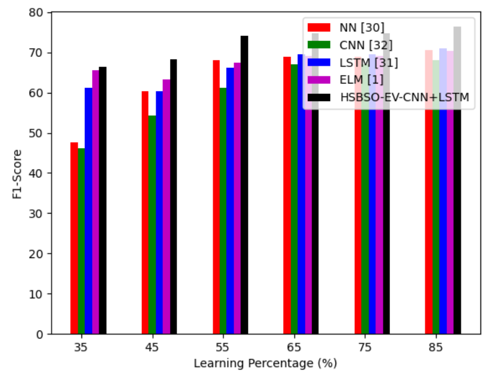


(j)

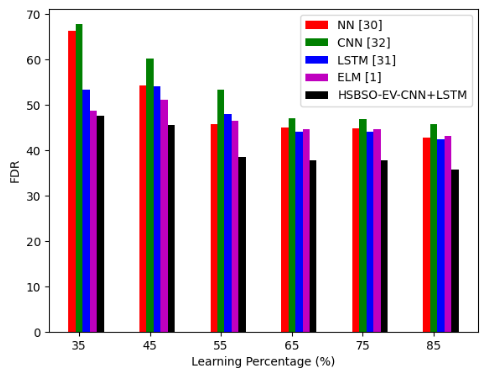
Fig. 9. Performance Analysis on Suggested Distracted Driver Detection Model with Conventional Meta-heuristic Algorithms in Terms of "(a) Accuracy, (b) F1-Score, (c) FDR, (d) FNR, (e) FPR, (f) MCC, (g) NPV, (h) Presicion, (i) Sensitivity, (j) Specificity".



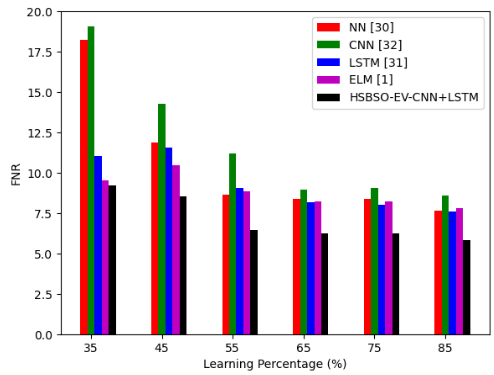
(a)



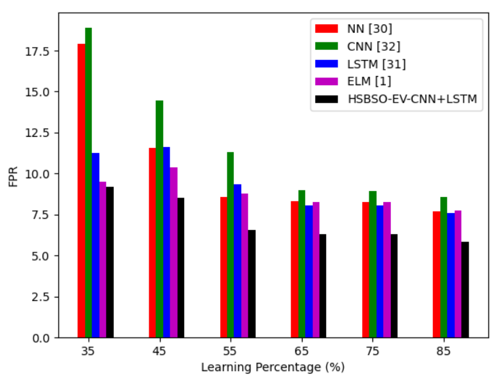
(b)



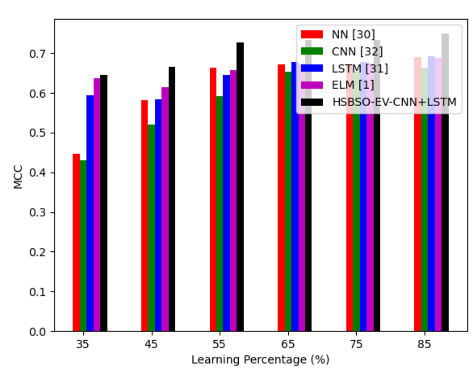
(c)



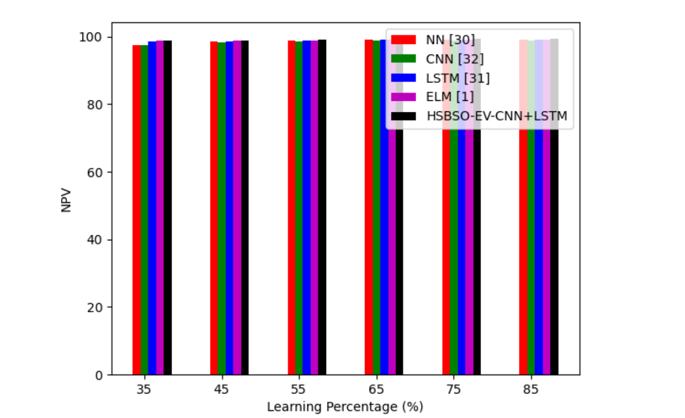
(d)



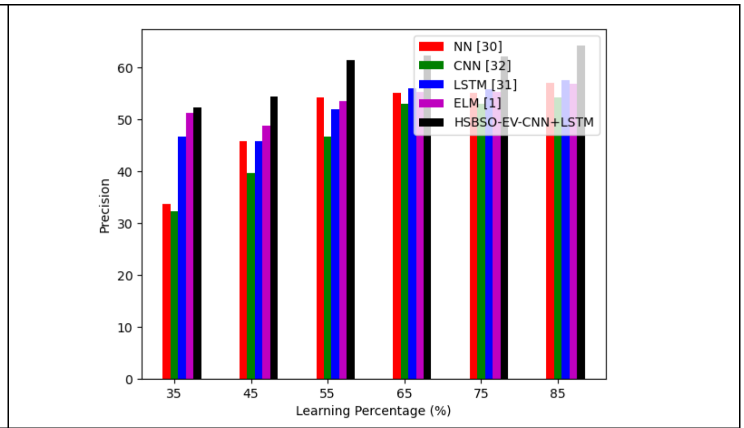
(e)



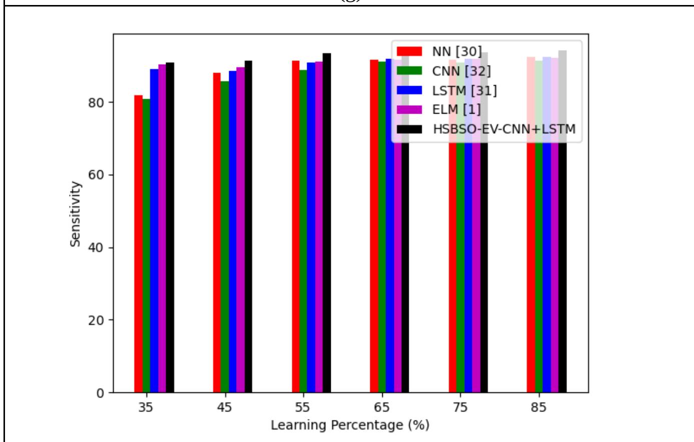
(f)



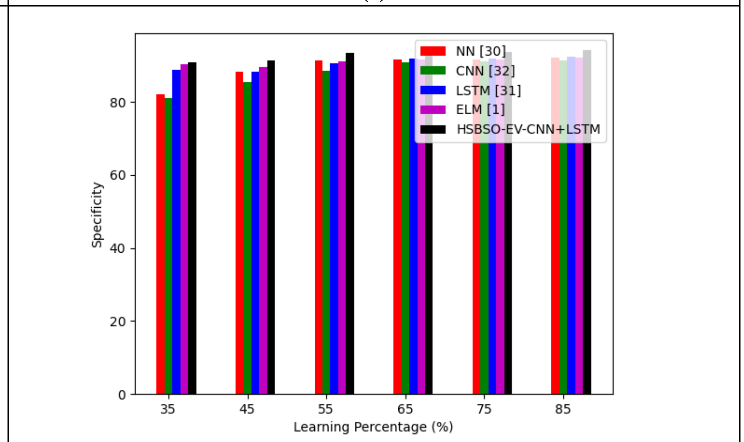
(g)



(h)



(i)



(j)

Fig. 10. Performance Analysis Accuracy of the Suggested Distracted Driver Detection Model with Existing Classifiers in Terms of "(a) Accuracy, (b) F1-Score, (c) FDR, (d) FNR, (e) FPR, (f) MCC, (g) NPV, (h) Presicion, (i) Sensitivity, (j) Specificity".

F. Overall Performance Analysis for the Proposed Model based on different Meta-heuristic Algorithms

The overall performance analysis for the proposed distracted driver detection model is estimated by comparing it with the different meta-heuristic algorithms as depicted in Table II. The performance of the proposed HSWOA-EV-CNN+LSTM is 0.36%, 0.59%, 0.62% and 0.38 superior to the PSO-EV-CNN+LSTM, GWO-EV-CNN+LSTM, SSA-EV-CNN+LSTM, and WOA-EV-CNN+LSTM, respectively when the observing the performance of precision. In all the performance measures, the proposed model shows enhanced performance in detecting distracted driver behaviors through the given images. Therefore, the proposed distracted driver detection model enhances its performance more than the existing methods.

G. Overall Performance Analysis for the Proposed Model based on different Classifiers

The overall performance analysis of the proposed distracted driver detection model is tested by comparing it with the different classifiers as expressed in Table III. The performance of the proposed HSWOA-EV-CNN+LSTM provides 1.88%, 3.55%, 4.4% and 1.06% enhanced MCC than the NN, CNN, LSTM and ELM, respectively. The proposed distracted driver detection model using the suggested HSBSO shows enhanced performance in all quantitative measures particularly provides higher values in the accuracy, precision, specificity and MCC than the conventional techniques. Therefore, the proposed distracted driver detection model provides improved performance than the existing methods.

TABLE II. OVERALL PERFORMANCE ANALYSIS OF SUGGESTED DISTRACTED DRIVER DETECTION MODEL WITH DIFFERENT META-HEURISTIC ALGORITHMS

Measures	*PSO [28]*	*GWO [29]*	*SSA [26]*	*WOA [27]*	*HSWOA-EV-CNN+LSTM*
"Accuracy"	0.9351	0.93686	0.9264	0.928	0.93684
"Sensitivity"	0.935	0.9372	0.9266	0.928	0.9372
"Specificity"	0.935111	0.936822	0.926378	0.928	0.9368
"Precision"	0.615537	0.622393	0.583061	0.588832	0.622311
"FPR"	0.064889	0.063178	0.073622	0.072	0.0632
"FNR"	0.065	0.0628	0.0734	0.072	0.0628
"NPV"	0.992336	0.992607	0.991273	0.991453	0.992607
"FDR"	0.384463	0.377607	0.416939	0.411168	0.377689
"F1-Score"	0.742358	0.748025	0.715742	0.720497	0.747965
"MCC"	0.727266	0.73316	0.699924	0.704787	0.733101

TABLE III. OVERALL PERFORMANCE ANALYSIS OF SUGGESTED DISTRACTED DRIVER DETECTION MODEL WITH DIFFERENT CLASSIFIERS

Measures	**NN [30]**	**CNN [32]**	**LSTM [31]**	**ELM [1]**	**HSWOA-EV-CNN+LSTM**
"Accuracy"	0.9173	0.91058	0.91936	0.91762	0.93684
"Sensitivity"	0.916	0.9094	0.92	0.9178	0.9372
"Specificity"	0.917444	0.910711	0.919289	0.9176	0.9368
"Precision"	0.55214	0.530881	0.558795	0.553091	0.622311
"FPR"	0.082556	0.089289	0.080711	0.0824	0.0632
"FNR"	0.084	0.0906	0.08	0.0822	0.0628
"NPV"	0.989929	0.989067	0.990423	0.990145	0.992607
"FDR"	0.44786	0.469119	0.441205	0.446909	0.377689
"F1-Score"	0.688981	0.670402	0.695284	0.690231	0.747965
"MCC"	0.672149	0.653005	0.678935	0.673661	0.733101

VIII. CONCLUSION

This paper has presented a new distracted driver detection model using suggested HSWOA along with the developed EV-CNN and improved LSTM approach. The gathered input images were initially given into the optimal fusion-based pattern extraction, where the two methods like LGP and LWP were used for generating the optimally fused patterns. These patterns were integrated with the optimized weights based on the proposed HSWOA. Then, the optimized patterns were acquired and considered for the developed EV-CNN approach for extracting the most significant features of the patterns. Further, the extracted features were subjected into the developed O-LSTM, which has been improved by optimizing certain parameters using the developed HSWOA. At last, the distracted driver behaviors were detected using the developed O-LSTM. Through the performance analysis, the proposed model using the developed HSWOA-EV-CNN+LSTM has secured 1.33% enhanced than NN, 1.83% improved than CNN, 2.2% better than LSTM and 0.84% elevated than ELM in terms of accuracy. Therefore, the proposed distracted driver detection model has ensured enhanced performance using developed O-LSTM along the suggested HSWOA than the existing methods for future work; we intend to increase the efficiency of the proposed solution by addressing additional driving style data along with road surface information, which will help improve the generalization of the distracted driver detection model.

REFERENCES

[1] M. Gjoreski, M. Z. Gams, M. Luštrek, P. Genc, J. U. Garbas, and T. Hassan, "Machine Learning and End-to-End Deep Learning for Monitoring Driver Distractions from Physiological and Visual Signals," IEEE Access, vol. 8, pp. 70590–70603, 2020, doi: 10.1109/ACCESS.2020.2986810.

[2] Z. Li, S. Bao, I. v. Kolmanovsky, and X. Yin, "Visual-Manual distraction detection using driving performance indicators with naturalistic driving data," IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 8, pp. 2528–2535, Aug. 2018, doi: 10.1109/TITS.2017.2754467.

[3] Y. Xing, C. Lv, H. Wang, D. Cao, E. Velenis, and F. Y. Wang, "Driver activity recognition for intelligent vehicles: A deep learning approach," IEEE Transactions on Vehicular Technology, vol. 68, no. 6, pp. 5379–5390, Jun. 2019, doi: 10.1109/TVT.2019.2908425.

[4] C. Y. Chiou, W. C. Wang, S. C. Lu, C. R. Huang, P. C. Chung, and Y. Y. Lai, "Driver Monitoring Using Sparse Representation with Part-Based Temporal Face Descriptors," IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 1, pp. 346–361, Jan. 2020, doi: 10.1109/TITS.2019.2892155.

[5] M. Rezaei, M. Terauchi, and R. Klette, "Robust Vehicle Detection and Distance Estimation Under Challenging Lighting Conditions," IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 5, pp. 2723–2743, Oct. 2015, doi: 10.1109/TITS.2015.2421482.

[6] S. Wang, Y. Zhang, C. Wu, F. Darvas, and W. A. Chaovalitwongse, "Online prediction of driver distraction based on brain activity patterns," IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 1, pp. 136–150, Feb. 2015, doi: 10.1109/TITS.2014.2330979.

[7] N. Vinayaga-Sureshkanth, A. Maiti, M. Jadliwala, K. Crager, J. He, and H. Rathore, "A practical framework for preventing distracted pedestrian-

related incidents using wrist wearables," IEEE Access, vol. 6, pp. 78016–78030, 2018, doi: 10.1109/ACCESS.2018.2884669.

[8] Y. Xing et al., "Identification and Analysis of Driver Postures for In-Vehicle Driving Activities and Secondary Tasks Recognition," IEEE Transactions on Computational Social Systems, vol. 5, no. 1, pp. 95–108, Mar. 2018, doi: 10.1109/TCSS.2017.2766884.

[9] T. Liu, Y. Yang, G. bin Huang, Y. K. Yeo, and Z. Lin, "Driver Distraction Detection Using Semi-Supervised Machine Learning," IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 4, pp. 1108–1120, Apr. 2016, doi: 10.1109/TITS.2015.2496157.

[10] C. Craye, A. Rashwan, M. S. Kamel, and F. Karray, "A Multi-Modal Driver Fatigue and Distraction Assessment System," International Journal of Intelligent Transportation Systems Research, vol. 14, no. 3, pp. 173–194, Sep. 2016, doi: 10.1007/s13177-015-0112-9.

[11] A. Gumaei et al., "A deep learning-based driver distraction identification framework over edge cloud," Neural Computing and Applications, 2020, doi: 10.1007/s00521-020-05328-1.

[12] M. Zhu, J. Chen, H. Li, F. Liang, L. Han, and Z. Zhang, "Vehicle driver drowsiness detection method using wearable EEG based on convolution neural network," Neural Computing and Applications, vol. 33, no. 20, pp. 13965–13980, Oct. 2021, doi: 10.1007/s00521-021-06038-y.

[13] M. Shahverdy, M. Fathy, R. Berangi, and M. Sabokrou, "Driver behavior detection and classification using deep convolutional neural networks," Expert Systems with Applications, vol. 149, Jul. 2020, doi: 10.1016/j.eswa.2020.113240.

[14] A. K. Dey, B. Goel, and S. Chellappan, "Context-driven detection of distracted driving using images from in-car cameras," Internet of Things (Netherlands), vol. 14, Jun. 2021, doi: 10.1016/j.iot.2021.100380.

[15] J. Xie, K. Hu, G. Li, and Y. Guo, "CNN-based driving maneuver classification using multi-sliding window fusion," Expert Systems with Applications, vol. 169, May 2021, doi: 10.1016/j.eswa.2020.114442.

[16] B. Baheti, S. Talbar, and S. Gajre, "Towards Computationally Efficient and Realtime Distracted Driver Detection with MobileVGG Network," IEEE Transactions on Intelligent Vehicles, vol. 5, no. 4, pp. 565–574, Dec. 2020, doi: 10.1109/TIV.2020.2995555.

[17] F. Omerustaoglu, C. O. Sakar, and G. Kar, "Distracted driver detection by combining in-vehicle and image data using deep learning," Applied Soft Computing Journal, vol. 96, Nov. 2020, doi: 10.1016/j.asoc.2020.106657.

[18] C. Huang et al., "HCF: A Hybrid CNN Framework for Behavior Detection of Distracted Drivers," IEEE Access, vol. 8, pp. 109335–109349, 2020, doi: 10.1109/ACCESS.2020.3001159.

[19] I. Jegham, A. ben Khalifa, I. Alouani, and M. A. Mahjoub, "Soft Spatial Attention-Based Multimodal Driver Action Recognition Using Deep Learning," IEEE Sensors Journal, vol. 21, no. 2, pp. 1918–1925, Jan. 2021, doi: 10.1109/JSEN.2020.3019258.

[20] C. Ou and F. Karray, "Enhancing Driver Distraction Recognition Using Generative Adversarial Networks," IEEE Transactions on Intelligent Vehicles, vol. 5, no. 3, 2020, doi: 10.1109/TIV.2019.2960930.

[21] A. Pal, S. Kar, and M. Bharti, "Algorithm for Distracted Driver Detection and Alert Using Deep Learning," Optical Memory and Neural Networks (Information Optics), vol. 30, no. 3, 2021, doi: 10.3103/S1060992X21030103.

[22] X. Rao, F. Lin, Z. Chen, and J. Zhao, "Distracted driving recognition method based on deep convolutional neural network," Journal of Ambient Intelligence and Humanized Computing, vol. 12, no. 1, pp. 193–200, Jan. 2021, doi: 10.1007/s12652-019-01597-4.

[23] D. K. Jain, R. Jain, X. Lan, Y. Upadhyay, and A. Thareja, "Driver distraction detection using capsule network," Neural Computing and Applications, vol. 33, no. 11, pp. 6183–6196, Jun. 2021, doi: 10.1007/s00521-020-05390-9.

[24] N. J. Dunn, T. A. Dingus, S. Soccolich, and W. J. Horrey, "Investigating the impact of driving automation systems on distracted driving behaviors," Accid Anal Prev, vol. 156, p. 106152, Jun. 2021, doi: 10.1016/j.aap.2021.106152.

[25] B. Jun and D. Kim, "Robust face detection using local gradient patterns and evidence accumulation," in Pattern Recognition, Sep. 2012, vol. 45, no. 9, pp. 3304–3316. doi: 10.1016/j.patcog.2012.02.031.

[26] L. Zhang, F. Liu, and J. Tang, "Real-time system for driver fatigue detection by RGB-D camera," ACM Transactions on Intelligent Systems and Technology, vol. 6, no. 2, Mar. 2015, doi: 10.1145/2629482.

[27] M. Wollmer et al., "Online driver distraction detection using long short-term memory," in IEEE Transactions on Intelligent Transportation Systems, Jun. 2011, vol. 12, no. 2, pp. 574–582. doi: 10.1109/TITS.2011.2119483.

[28] C. Ou, Q. Zhao, F. Karray, and A. el Khatib, "Design of an end-to-end dual mode driver distraction detection system," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2019, vol. 11663 LNCS, pp. 199–207. doi: 10.1007/978-3-030-27272-2_17.

[29] J. M. Mase, P. Chapman, G. P. Figueredo, and M. T. Torres, "Benchmarking Deep Learning Models for Driver Distraction Detection." [Online]. Available: https://www.researchgate.net/publication/341625555.

[30] R. Sayed and A. Eskandarian, "Unobtrusive drowsiness detection by neural network learning of driver steering," 2001.

[31] S. Mirjalili and A. Lewis, "The Whale Optimization Algorithm," Advances in Engineering Software, vol. 95, pp. 51–67, May 2016, doi: 10.1016/j.advengsoft.2016.01.008.

[32] M. Jain, V. Singh, and A. Rani, "A novel nature-inspired algorithm for optimization: Squirrel search algorithm," Swarm and Evolutionary Computation, vol. 44, pp. 148–175, Feb. 2019, doi: 10.1016/j.swevo.2018.02.013.

[33] Institute of Electrical and Electronics Engineers, Particle swarm optimization of feedforward neural networks for the detection of drowsy driving.

[34] X. Zhang, Z. Liu, Q. Miao, and L. Wang, "An optimized time varying filtering based empirical mode decomposition method with grey wolf optimizer for machinery fault diagnosis," Journal of Sound and Vibration, vol. 418, pp. 55–78, Mar. 2018, doi: 10.1016/j.jsv.2017.12.028.

[35] S. Masood, A. Rai, A. Aggarwal, M. N. Doja, and M. Ahmad, "Detecting distraction of drivers using Convolutional Neural Network," Pattern Recognition Letters, vol. 139, pp. 79–85, Nov. 2020, doi: 10.1016/j.patrec.2017.12.023.

A Review of the Integration of Cyber-Physical System and Internet of Things

A Cyber-Physical Systems Perception of Internet of Things

Ramesh Sneka Nandhini[1]
School of Computer Science and Engineering
Vellore Institute of Technology, Vellore, India

Ramanathan Lakshmanan[2]*
School of Computer Science and Engineering
Vellore Institute of Technology, Vellore, India

***Abstract*—Cyber-physical system has a bigger impact on the present and future of the engineered systems. The integration of cyber-physical system with other technologies often poses challenges that are unique in the fields of design, implementation, and applications. This study explores the definitions of the integrated cyber-physical system from different perceptions, how cyber-physical systems evolved, emerging fields of research in cyber-physical systems. Efficiency, reliability, predictability, and security are some of the challenges that cyber-physical systems pose in the overall implementation. cyber-physical systems when integrated with the Internet of Things evolves into a hybrid technology that advances the technology aspects. The CPS-IoT models complement each other offering useful insights to deal with engineering systems and control modules.**

Keywords*—*Cyber-physical system; internet of things; industrial internet (ii); industry 4.0

I. Introduction and Motivation

Cyber-Physical Systems are modelled on the collaboration of 3C's - computation, communication and control that are discrete and logical (cyber entity) and the natural and artificial systems that are administered by the laws of physics (physical entity). Both the entities are linked through sensors and actuators. Sensors, computational and physical world are closely integrated in CPS. CPS includes both conventional embedded systems and control systems that are presumed to be remodelled by emerging methodologies and integration of the Internet of Things.

IoT is the base or enabling technology for cyber-physical systems. Cyber-physical systems can be contemplated as the advancement of IoT as complete conception and perception, and which possess a strong ability to control the physical world. Cyber-physical systems also include the traditional embedded and control systems, evolving them into innovative methodologies. IoT connects the information acquiring devices like sensors and RFID (Radio Frequency Identification) wireless sensor networks and cloud computing technology for reliable transmission and information processing. On the other hand, CPS is a control technology that is scalable and reliable and is the integration of computing, communication, and control of IoT. However, IoT concentrates on information transmission and processing, whereas CPS not only has the can sense but also possesses a strong ability to control. In CPS, cyber and physical elements are interlinked with one another, which are running on spatial and temporal scales, which reveal numerous and definite behavioural process and collaborate with each in many forms that vary the context.

CPS is transdisciplinary which merges mechatronics, computation theory, Internet of things, wireless sensor networks, theory of cyber nets, design and process science. The process of control is related to embedded systems using feedback loops. CPS and the Internet of Things share the same architecture which makes them similar but a higher between physical and computational elements is defined in CPS. In contrast to traditional embedded systems, CPSs are specifically represented as a structure of interacting elements with physical input and output rather than standalone devices [1]. Applications of CPS include healthcare and industrial automation, control technology, distributed energy system, aircraft control and so on [2]-[4]. CPS will bring change to the existing physical systems in engineering and advances in economic welfare.

II. CPS History and Definition

CPS refers to next generation engineered systems. The word cyber-physical system came around 2006 by Helen Gill at the National Science Foundation (NSF). Cyber-physical system is often confused with "cyber security" where there is no relation with the physical processes. CPS is the tight combination of computations, algorithms, and physical devices. The perception of the technologies is that it associates the real world with the information world. CPS connects by means of the popular technologies such as Internet of Things (IoT), Industry 4.0, the Industrial Internet (II), Machine-to-Machine (M2M), Industrial Internet of Things (IIoT).

A. Definition of CPS

This field of research overlaps with various fields of science and engineering, computer science and networking should come together in collaboration with multiple fields such as automation and control, civil engineering, mechanical engineering, etc. Therefore, CPS definition may vary from their viewpoint.

E.A.Lee describes CPS as "the integration of computations and physical processes. Embedded computers and network monitor and control the physical process, usually with feedback loops where physical processes affect computations and vice versa [5]."

*Corresponding Author.

R.Baheti describes CPS as "the new generation of systems with integrated computational and physical capabilities that can interact with humans through many new modalities. The ability to interact with, and expand the capabilities of, the physical world through computation, communication, and control is a key enabler for future technology developments [6]."

Department of Science and Technology (DST) – Government of India launched a Interdisciplinary Cyber Physical System (ICPS) to promote research and development in this emerging field. The simple architecture of CPS is shown in Fig. 1. The collaboration of the physical and cyber world in CPS will assure real-time, reliable, trustworthiness of the systems. Physical systems collect the data from sensors in CPS systems dynamically guaranteeing the accuracy of the data collected. The data is sent to the cyber world for the processing of information based on the service needs mainly uncertainty management, statistical analysis, and feedback control. These systems are designed in such a way that they self-adapt. CPS is a 3C technology i.e., computation, communication, and control that takes wireless sensor network and Internet of Things.

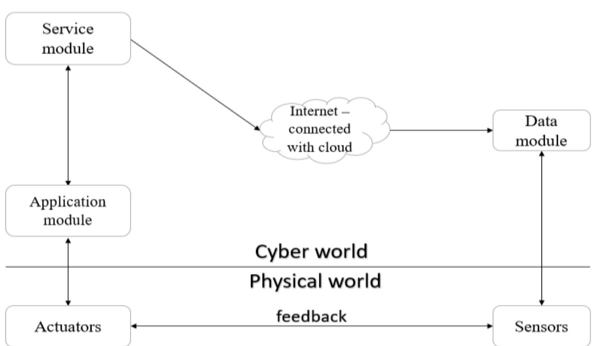


Fig. 1. Working idea of Cyber-Physical System

B. Characteristics of CPS

CPS has both physical and cyber-world. In the physical-world, CPS interacts through sensors adapting embedded computing technology. The cyber-world of the CPS deals with computations and control. The CPS upholds the following characteristics [7]:

- Cyber capability in every physical component: The physical system design is as important as the logical design. It involves hardware design and size, management, connectivity establishment, system testing. Engineers have an understanding in their fields of systems, they acquire the technical characteristics of sensors, actuators and how they process the raw data to the next layer. The higher degree of automation and control in CPS demands the maximum network coverage.

- Networked at multiple and extreme scales: CPS entrust networks to sensors and actuators to collect data and process them. The infrastructure of networking has to be very large frequently. Networks should be able to perform on different domains such as personal transportation systems, healthcare, remote interaction, industrial automation. These domains require a new communication notion that provides concrete and outline together with online adaption for productive use of bandwidth. This leads to scalable, open and adaptive communication [8]. CPS demands networking to be in large and multiple scales to meet the need without compromising the Quality of Service (QoS).

- Multiple spatial-temporal constraints: The CPSs are conceptualized to 3C technology i.e communication, computation and control with the physical world. These require the interaction between the components of CPS in time and space. Constraints such as event detection and action decision should be implemented accurately and appropriately to meet spatial and temporal correctness [9].

- Unconventional computational and physical substrates: The need for unconventional computation in CPSs is because of its dynamic reconfiguration and its communication with the physical world in real-time to adapt them to specific needs. Since the evolution in computation takes place, the physical substrate should comply with the computation as well.

- High degree of automation, control loops closed at many scales: The emerging amount of data on CPS, coupled with the need for systems to be robust and handle real-time environment which leads to self-learning. The automation is important for the system to go forward. The control object is the inbuilt operation of CPS. This object can reform itself and change the physical flow command. The control aspect of the CPS is the basis for dynamic systems and plays a crucial role in feed forwards control loops.

- Dynamic recognizing termed as open system: This characteristic of the CPS includes reorganizing and reconfiguring which can automatically adapt to the environment and develop rules based on the domain requirements and alters the external physical environment to meet the conditions according to present rules.

III. Research Directions of CPS

CPS is the integration of the multidisciplinary heterogeneous systems and has decentralized models. These systems exhibit behaviours that are not predictable in advance, which makes the research possibilities expansive. The researchers recently concentrate on the architecture pattern, computational theory, and control of the feedback system. Other research fields such as system reliability, information processing, system design software, and CPS security system are focused on recently.

A. Architecture of CPS

Many architectures for CPS have been proposed so far, the 5 Component architecture has 5 levels: Connection level, Conversion level, Cyber level, Cognition level, Configuration level [10]. These levels emerge the process for such interdependent systems and overcome the limitations to meet the demand of the target system. These levels process and monitor the information, and coordinate with the physical systems that relate to the cyber computational space. 8 Component architecture is an evolution of 5C architecture that includes three other components in addition to Connection level, Conversion level, Cyber level, Cognition level, Configuration level. The included components are Coalition, Customer, and Content [11].

ACPS architecture where not only the highest level of abstraction is provided but also integrates the physical, cyber, and human components [12]. Service-Oriented Architecture (SOA) supports this model to achieve the adaptive and dynamic vision. 3C architecture is the most commonly used architecture in all applications [13]. This architecture reinforces smart manufacturing in industries and supports the industrial revolution i.e., Industry 4.0. The three components in this architecture are the human component (HC), Physical component (PC), and Cyber component (CC). A smart factory framework is developed which considers the physical and cyber aspects of the system architecture to offer the solution and improve efficiency [14].

Though many Cyber-Physical Systems' architectures meet the expectations of the target system or applications, there is yet a need for further developments as most of these models are clearly defined for a certain set of applications. Researches on architecture are restricted to specific applications and their advancements, the development conditions due to which the interdependence of various demands are not studied. CPS architecture is the foundation of research and progress, so the proposed models must be integrated and improvised based on the present system structure and should accommodate the development of CPS. However, a unified architecture for CPS is required which calls for further research in the architecture.

B. Control of CPS

The foundations of control theory in Cyber-Physical Systems are yet to be established strongly. Converting the input actions into appropriate control actions, in a multimodal application that involves sensors and actuator parameters based on the user actions [15]. An independent way to control the cyber and physical resources, where a new control strategy is provided for data centers to attain stability for the CPS applications [16].

A robust control approach that is modular, promises the stability and performance of the target system automatically due to which the input-output analysis where a standard analytic method is used [17]. The role of control in the cyber-physical system must be addressed because of their impact on the future Internet of Things, wearable electronics, human interaction machines, and many other fields. The analysis of control theory in CPS helps in building the foundations of CPS stronger. A unified approach must be dealt with towards the control strategies that include temperature set-points, adjusting the pressure set-points, monitor air temperature, dimming or turning off the lights, etc. Predictability is one of the challenges to the control of CPS. To ensure predictability, aspects of the model should be accounted such as input and output behaviour, execution, and programming. Uncertainty affects the model or system's guaranteed behaviour [18].

The traditional control theory only meets the targets of CPS control based on the applications or present event models, but the theory should also manage the transmission and response of feedback control where further research prevails.

C. Data Sciences and Data Processing Technologies for CPS

The data is collected by the systems, must be transformed into knowledge to create value. The scope of future smart systems lies in the methodologies and techniques of Data Science. A vast amount of data is collected from the sensor networks and many solutions for data science problems have been discovered. CPS follows a data-driven approach to learn patterns and predict the models, also analyse their environment, and based on their observations. Data processing, in general, includes collection, preparation, input, processing, interpretation, feedback control, and the response from the physical environment after receiving the commands.

Reliability is an essential need for a cyber-physical system as it is applied for safety and critical systems. CPS is intended to process a huge volume of data and run continuously in real-time, where an Automatic Reliability Improvement System (ARIS) is proposed to improve the reliability. ARIS works as a data-centric runtime monitoring system, where anomalous data is indicated and a data quality analysis is performed using computational intelligence and self-tuning techniques for system reliability [19]. The existing Internet of Things is turned into a complex cyber-Physical Social System (CPSS) when people use their social networks and personal computational resources which leads to the generation of huge volume passive data to process. A data processing pattern that is based on stream processing technology, that uses a cluster of edge devices to distribute the workload is introduced. The feasibility of this approach is guaranteed through a demonstration of an intelligent surveillance system that is stationed on the edge device [20]. Present researchers on the methodologies and techniques of Data Science in CPS face challenges in processing and transmission technology of data, safety control, real-time capability, robustness, control and hybrid systems and reliability.

D. Security and Privacy of CPS

The security and privacy protection of the established internet security does not meet the requirements of physical system security in CPS which is randomly distributed sensors over the pervasive wireless network. Most of the present security policy uses a centralized network that trusts a third-party operation. The new security framework that considers the flexibility, privacy protection, transaction security, transparency, efficiency, trust mechanism, data integrity, resilience, the trustworthiness of data that combines with control and information.

Blockchain technology, which is a distributed register that is shared between peer networks is proposed as a unified three-level architecture to identify the possibilities and also adapt, develop and incorporate this with the manufacturing in industry 4.0. The data flow and the communication are incentivized in the structure of the existing cyber-physical system structure to assure safety and reliability [21]. Differential privacy techniques are considered efficient to protect privacy in CPS which can be achieved through certain modifications in the architecture of CPS and can be implemented in cyber-physical systems applications [22].

However, there are some challenges in cyber-physical system from a research perspective. Robustness, safety, and security due to the environment and errors in the physical devices making it a critical challenge. Location, time, and tag-based data can be a threat to security. Merging the systems to create a hybrid system to create an efficient feedback control is required. The architecture of the CPS must be designed in a way that is very rational at every level and it should consider all the physical device information, protocols should be designed at a large-scale. Collecting the raw data at a vast amount from the sensor and mobile networks is one of the major challenges to focus on. Integration and modification should be done at every level of CPS design, communication and computation are to be predicted at all scales and model the time information.

IV. Cyber-Physical System-Internet of Things

The CPS can be considered as an evolution of IoT with more intelligence and interactive operations, adapting the architecture of IoT. CPS enriches the interaction between human-to-human, human-to-machine, machine-to-machine not only in the physical world but also in the virtual world, also includes conventional embedded and control systems that will not be altered by novel approaches. IoT and CPS when introduced in the industry, brings many benefits and extends the scope such as awareness, prediction, comparison, reconfiguration, maintenance. The evolution of IoT and the development of CPS is prominent in certain applications - Real-Time Physical Systems (RTPS), based on real-time parameters with a suitable feedback control policy based on dynamic adaptions [23]. A controller is designed for a nominal model that stabilizes the large-scale CPS [24]. The vulnerability of robotic surgery is reduced by designing a collaborative robotic cyber-physical system [25].

CPS is the integration of physical and cyber systems where embedded computing devices act as physical entities and communication, computation, and control are cyber entities, whereas IoT shows the interconnection of different end-devices that communicates with each other through the internet. Decentralized control among the interconnected end-devices is intended by IoT. Sensing, processing, storing, and networking are some of the common capabilities of both CPS and IoT.

When CPS was emerged and introduced alongside IoT some of the similar characteristics of CPS and IoT must be identified, familiarized, and contrasted to understand their respective roles in the system where they are used. New strategies for communication and control have been introduced making the system more safe, secure, and sustainable which are the fundamental attributes of the CPS.

The advanced communication technologies, alongside wireless communications, lead to the development and implementation of CPSs considered as the higher-level systems of IoT. These technologies include wearable devices, sensors in healthcare, intelligent systems where cloud computing is applied. CPS in IoT deals with the interaction of the intelligent systems, applications that are user-friendly sand interactive, optimization in the IoT-enabled CPS, and control of the distributed systems. The characteristic of the CPS must be enhanced due to the higher requirements and efficient performance. The performance of the CPS elevates the levels of IoT. In short, CPS is the next evolution of M2M systems where smart information processing and IoT are regarded as the future technicalities.

A. Internet of Things enabled Cyber-Physical System

CPS and IoT are intertwined for better and unified solutions, where the users, machines and the environment interact with each other. Defining the interface for the IoT technologies, control of the global system, integration of the IoT devices and physical environment, which is tied together with the communication infrastructure, along with control module are some of the challenges when IoT has enabled alongside CPS on a large scale. The performance of real-time analytics should be considered as another challenge as the large amount of data collected has to undergo a mining process for better results.

The development of the CPSs which adopt the IoT paradigms aids in several applications of the CPS. Design complexity, modeling the hybrid systems, smart control design, sensor reliability, specified IoT sensor and actuation design for CPSs, and security and privacy are some of the key challenges for IoT-enabled CPS.

The infrastructure and control of the large-scale system and the network monitoring are used for better communication and coordination among the modules. This may have a negative effect on the network structure and needs to be upgraded for efficiency and accuracy. Enhanced Expander Membership Protocol [26], having features such as addressing the real-time network issues, limiting the negative effects on the network structure, and recovering the global connectivity efficiently and is also fault-tolerant and scalable. These support the challenges addressed by the CPS network structure. A Fragmented-Iterated Bloom Filter [27] is introduced to manage the distributed network in the event-based cyber-physical system. This is efficient in memory management and computation.

Considering the functional and non-functional requirements during the implementation of the IoT-enabled CPS, the challenges such as scheduling mechanisms, vulnerability to security attacks, applications development, etc. An integrated development environment that is model-driven in cloud computing supports the application development, code generation, and compilation and simulations for the IoT-enabled CPS applications.

COMFIT [28] has two modules, the App Development Module and the App Management and Execution Module, providing an environment where every needed action is performed for the code to develop the app. The scheduling mechanism in the implementation of IoT-enabled CPS needs to be efficient and scalable. The Bayesian Network [29] is used for scheduling algorithms where stability and operational capability have high priority. The vulnerabilities to the security attacks are tackled using game-theoretic principles, providing a scheduling mechanism that is effective based on the game.

B. Integration of CPS and IoT

A Cyber-physical system, in general, is the collaboration of communication, computation, and control with the physical world on a large scale. They are considered as a form of wireless sensors and actuator networks. On the other hand, the interconnection of heterogeneous devices that communicate through the internet is called the Internet of Things. These heterogeneous devices can be referred to as smartphones, laptops, RFID, etc. are intelligent agents that share the information with the systems and people which makes them a part of CPS. Therefore, the integration of CPS-IoT is a solution to many technological challenges and is used in many application domains.

The integration of Environmental IoT with the cyber-physical clouds [30] results in a complex distributed system that supports the comprehension of interdependencies and management in the natural environment. A three-stage approach is deployed - The Abstraction representation, Network-centric representation, Node-centric representation. These technologies provide higher opportunities in the application domain by supplementing each other.

The integration of wireless devices, wearables that are fast and have a working real-time algorithm to cyber-physical systems allows in numerous aids to the medical applications. The author in [31] presents a system by integrating the complimenting technologies that provide comfort to the patients and the data collection from devices is efficient.

An adaptive interface approach [32] that supports both the development and the utilization of the CPS-IoT is proposed. An IoT toolkit with a multi-layer interface that allows users to improve the interaction with sensor and actuator networks is presented. The full potential of the IoT-CPS integration interface can be utilized, with the aid of the presented adaptive interface.

The integration of the CPS for real-time IoT operations [33], resulting in the effective control of actions based on the environmental objects. This helps in the real-time operations in the manufacturing units and Industry 4.0 applications. The integration of CPS with various technologies leads to emergent properties and unpredictable behaviors. The development of IoT and high-speed network leads to the possibilities of the physical world meeting the cyber world. The convergence of Body Area Sensor Networks (BASN) and CPS made the technologies reach their threshold point, making hand pattern recognition smart and safe [34].

C. CPS-IoT Models

The integration of IoT and CPS evolved over time to complement each other and has some very useful applications and insights; however, IoT and CPS have distinct origins. CPS primarily deals with engineering the systems and the control stance, in contrast, the IoT deals with networks and the information technology that digitalizes the physical world, also having the sensing and actuating capabilities. The vision of IoT is the effective interaction between digital and physical entities, enabling the information and communication technologies to their full potential [35]. Despite the distinct origins of CPS and IoT, their models overlap each other. The author in [36] present four overlap models: Partial overlap, Equivalence, CPS as a subset of IoT, IoT as a subset CPS.

IoT and CPS, despite having the same goals as reducing the communication gap between the cyber world and physical world; also has some evident differences [37] - IoT is an open network infrastructure that connects the objects in the physical world through the internet and the data is processed through information technologies. The CPS is all about the control of the physical world through feedback, implementing an Automatic Reliability Improvement System, forming operator-in-the-Loop.

CPS encompasses both open and closed-loop control systems, addressing the human-interaction and for the feedback implements the Human-in-the-Loop (HITL) control system [38], where IoT is all about open-loop systems. IoT and CPS overlap with each other in many aspects making the terms and definitions distinct. IoT usually senses the physical world, processes the data by monitoring these technologies. The IoT application is also used in the control module in the CPS through integrating certain edge devices [39].

Cyber-Physical Systems, link the physical world with that of the virtual world which is defined as the Internet of Things and its services making CPS a subset of IoT. The applications such as IoT-connected smart home systems, a tracking device based on the geolocation [40] are IoT specialized which excludes the CPS making the proof for the above statement. Cyber-physical systems feature collaboration and the coordination of 3C's, physical elements, and its integration with the cyber world of information processing. The greater focus is on the control for the CPS, leaving IoT a simple platform of CPS. IoT senses the physical world through sensors, the internet keeps them connected with each other, collects data, and processes them - CPS also does the same, additionally targets the control of the physical processes addressing the human-machine interaction through feedbacks, which IoT does not feature making IoT as a subset of CPS. The sensor or actuators' activity is collaborated to achieve a goal where CPS uses IoT systems as a platform [41].

The Industrial Revolution is also known as Industry 4.0 or Industrial Internet (II) is one of a kind and next-generation advancements in CPS, where everything is smart and connected in the real-time environment. The unique features of CPS, that IoT does not comply makes it distinct and also clarifies that IoT is not everything that CPS is [42]. Control, platform, internet, and human-machine interactions give the perspective for CPS-IoT overlap models. The definitions of CPS and IoT evolve based on hybrid systems. Physical and human components in the systems and their applications. The perspective of CPS/IoT models allows the idea for the innovation of new applications domains, making its use unified. The integration of CPS-IoT proposes a solution for many challenges making the approach towards a problem simpler.

V. Applications of Cyber-Physical System

The use of CPS in various application domains is to make large-scale systems adaptable, efficient, functional, reliable, and autonomous. CPS has three major advantages - they ensure real-time data collection from the physical world, feedback to the cyber world to perform the control module, a computational process that makes the cyber world stable [43]. These advantages are contributing factors to use CPS in certain domain applications. There are some definite areas where CPS has made potential progress such as technological innovations, system architecture, production systems, design and model of a system, and real-world applications. The CPS can almost be applied to many disciplines and also have trans-disciplinary applications. Numerous categories such as agriculture, energy management, education, weather, and environmental monitoring, medical devices, process control, smart homes, industry 4.0/smart manufacturing, security, wearable devices, transportation system, and traffic prediction are the areas where CPS can be applied.

A. Environmental Monitoring

Wireless sensor networks can be considered as a pillar for CPS. CPS analyzes the surrounding environment through sensor nodes that are deployed in diverse locations with human intervention. These sensor nodes detect both natural and man-made disasters, such as flooding, fire, release of toxic gases, abundant rainfall, etc. which in turn influences the balance of the environment. [44] investigates the impact of the abnormal environmental conditions and the failures caused to the system, which is dealt with by a framework that interacts between cyber-physical system and its environment. A Functional Failure Identification and Propagation (FFIP) framework detects the component failures by analyzing the environmental conditions or loss of the functionality in the system in the environmental surroundings.

The author in [45] uses technologies based on Wireless Sensor Networks (WSN), multi-agents, and cloud computing to monitor the real-time environment status. The CPS paradigm using the technologies operates in a volatile environment providing flexibility and scalability. The three-layer architecture of CPS, where WSN makes sure the data is acquired from physical nodes integrated with sensors and transferred to other layers to supervise the operations of CPS using the decision rules and data analysis.

B. Transportation System

Smart transportation deals with real-time information analysis. Autonomous driving vehicles, Vehicular Cyber-Physical Systems (VCPS), and Intelligent transportation are major advancements of CPS in the transportation domain. Autonomous driving vehicles show almost no fatalities compared to human-driven vehicles ranging from cars to planes. Coordination is the key to smart transportation; it is achieved through the means of 3C's - communication, computation, and control.

Considering the traffic measurement, a privacy-preserving point-to-point function is considered where the number of traveling vehicles is measured based on geographical location by the intelligent cyber-physical road systems. Bit arrays are used to collect data and maximum likelihood estimation (MLE) is used to obtain the measurement result [46]. The co-optimization and co-regulation schemes for both cyber and physical resources are considered by examining the sampling patterns in time-variations, scheduling of sensors, control mechanism, feedback scheduling, and motion planning and resources for Cyber-Physical Vehicle Systems (CPVSs) [47].

The prediction of road traffic has been one of the most useful applications of CPS and is accurate. Here, an operator-in-the-loop is added to the framework so that based on the prediction, the traffic in the certain location can be prevented and a retrospective analysis can be done to prevent the road traffic.

VI. Conclusion

The CPS has an impact on the socio-economic life of its users. This paper conducts a review on the definition of CPS and its history, different fields of research, the integration of IoT-CPS, and how they complement each other and its applications. The research areas of CPS and its advancements have been part of many emerging trends in the information technology fields such as IoT, cloud computing, big data, cloud computing, Industrial Internet, Industry 4.0.

References

[1] A. Ahmadi, C. Cherifi, V. Cheutet, and Y. Ouzrout, "A review of CPS 5 components architecture for manufacturing based on standards," in *2017 11th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)*, 2017, pp. 1-6.

[2] Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, "Health-CPS: Healthcare cyber-physical system assisted by cloud and big data," *IEEE Systems Journal*, vol. 11, 2015, pp. 88-95.

[3] P. J. Mosterman and J. Zander, "Industry 4.0 as a cyber-physical system study," *Software & Systems Modeling,* vol. 15, 2016, pp. 17-29.

[4] R. Rajkumar, "A cyber–physical future," *Proceedings of the IEEE,* vol. 100, 2012, pp. 1309-1312.

[5] E. A. Lee, "Computing foundations and practice for cyber-physical systems: A preliminary report," *University of California, Berkeley, Tech. Rep. UCB/EECS-2007-72,* vol. 21, 2007.

[6] R. Baheti and H. Gill, "Cyber-physical systems," *The impact of control technology,* vol. 12, 2011, pp. 161-166.

[7] R. Rajkumar, I. Lee, L. Sha, and J. Stankovic, "Cyber-physical systems: the next computing revolution," in *Design automation conference*, 2010, pp. 731-736.

[8] Almeida, Luís, "Network Challenges in Cyber Physical Systems," Communications of the China Computer Federation, 2012, pp. 26-31.

[9] Y. Tan, M. C. Vuran, and S. Goddard, "Spatio-temporal event model for cyber-physical systems," in *2009 29th IEEE International Conference on Distributed Computing Systems Workshops*, 2009, pp. 44-50.

[10] B. Bagheri, S. Yang, H.-A. Kao, and J. Lee, "Cyber-physical systems architecture for self-aware machines in industry 4.0 environment," *IFAC-PapersOnLine,* vol. 48, 2015, pp. 1622-1627.

[11] J.-R. Jiang, "An improved cyber-physical systems architecture for Industry 4.0 smart factories," *Advances in Mechanical Engineering,* vol. 10, 2018, p. 1687814018784192.

[12] B.-C. Pirvu, C.-B. Zamfirescu, and D. Gorecky, "Engineering insights from an anthropocentric cyber-physical system: A case study for an assembly station," *Mechatronics,* vol. 34, 2016, pp. 147-159.

[13] A. Ahmadi, A. H. Sodhro, C. Cherifi, V. Cheutet, and Y. Ouzrout, "Evolution of 3C cyber-physical systems architecture for industry 4.0," in *International Workshop on Service Orientation in Holonic and Multi-Agent Manufacturing*, 2018, pp. 448-459.

[14] Chen, G., Wang, P., Feng, B., Li, Y. and Liu, D, "The framework design of smart factory in discrete manufacturing industry based on cyber-

physical system," *International Journal of Computer Integrated Manufacturing*, *33*(1), 2020, pp.79-101.

[15] V. K. Singh and R. Jain, "Situation based control for cyber-physical environments," in *MILCOM 2009-2009 IEEE Military Communications Conference*, 2009, pp. 1-7.

[16] L. Parolini, N. Tolia, B. Sinopoli, and B. H. Krogh, "A cyber-physical systems approach to energy management in data centers," in *Proceedings of the 1st acm/ieee international conference on cyber-physical systems, 2010,* pp. 168-177.

[17] J. L. Ny and G. J. Pappas, "Robustness analysis for the certification of digital controller implementations," in *Proceedings of the 1st ACM/IEEE International Conference on Cyber-Physical Systems*, 2010, pp. 99-108.

[18] Sun, B., Li, X., Wan, B., Wang, C., Zhou, X. and Chen, X, "Definitions of predictability for cyber physical systems," Journal of Systems Architecture, 63, 2016, pp.48-60.

[19] L. Wu and G. Kaiser, "An autonomic reliability improvement system for cyber-physical systems," in *2012 IEEE 14th International Symposium on High-Assurance Systems Engineering*, 2012, pp. 56-61.

[20] R. Dautov, S. Distefano, D. Bruneo, F. Longo, G. Merlino, and A. Puliafito, "Data processing in cyber-physical-social systems through edge computing," *IEEE Access,* vol. 6, 2018, pp. 29822-29835.

[21] J. Lee, M. Azamfar, and J. Singh, "A blockchain enabled Cyber-Physical System architecture for Industry 4.0 manufacturing systems," *Manufacturing letters,* vol. 20, 2019, pp. 34-39.

[22] M. U. Hassan, M. H. Rehmani, and J. Chen, "Differential privacy techniques for cyber physical systems: a survey," *IEEE Communications Surveys & Tutorials,* vol. 22, 2019, pp. 746-789.

[23] T. Facchinetti and M. L. Della Vedova, "Real-time modeling for direct load control in cyber-physical power systems," *IEEE Transactions on Industrial Informatics,* vol. 7, 2011, pp. 689-698.

[24] K. Sakurama, "Control of large-scale cyber-physical systems with agents having various dynamics," *IEEE Transactions on Big Data,* 2017, pp. 691-701.

[25] D. D'Auria and F. Persia, "A collaborative robotic cyber physical system for surgery applications," in *2017 IEEE International Conference on Information Reuse and Integration (IRI),* 2017, pp. 79-83.

[26] P. Poonpakdee and G. Di Fatta, "Robust and efficient membership management in large-scale dynamic networks," *Future Generation Computer Systems,* vol. 75, 2017, pp. 85-93.

[27] C. Munoz and P. Leone, "A distributed event-based system based on compressed fragmented-iterated bloom filters," *Future Generation Computer Systems,* vol. 75, 2017, pp. 108-127.

[28] C. M. de Farias, I. C. Brito, L. Pirmez, F. C. Delicato, P. F. Pires, T. C. Rodrigues et al. "COMFIT: A development environment for the Internet of Things," *Future Generation Computer Systems,* vol. 75, 2017, pp. 128-144.

[29] B. K. Chejerla and S. K. Madria, "QoS guaranteeing robust scheduling in attack resilient cloud integrated cyber physical system," *Future Generation Computer Systems,* vol. 75, 2017, pp. 145-157.

[30] R. V. Rodrigues Filho, B. Porter, and G. Blair, "Environmental iot: Programming cyber-physical clouds with high-level system specifications," in *2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing*, 2014, pp. 947-950.

[31] D. De Venuto, V. F. Annese, and A. L. Sangiovanni-Vincentelli, "The ultimate IoT application: A cyber-physical system for ambient assisted living," in *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2016, pp. 2042-2045.

[32] M. Kalverkamp and C. Gorldt, "IoT service development via adaptive interfaces: Improving utilization of cyber-physical systems by competence based user interfaces," in *2014 International Conference on Engineering, Technology and Innovation (ICE)*, 2014, pp. 1-8.

[33] V. Maru, S. Nannapaneni, and K. Krishnan, "Internet of things based cyber-physical system framework for real-time operations," in *2020 IEEE 23rd International Symposium on Real-Time Distributed Computing (ISORC)*, 2020, pp. 146-147.

[34] G. Fabbri, C. M. Medaglia, A. Pecora, L. Maiolo, and M. Santello, "Cyber physical systems and body area sensor networks in smart cities," in *2016 IEEE 25th International Symposium on Industrial Electronics (ISIE)*, 2016, pp. 980-985.

[35] D. Miorandi, S. Sicari, F. De Pellegrini, and I. Chlamtac, "Internet of things: Vision, applications and research challenges," *Ad hoc networks,* vol. 10, 2012, pp. 1497-1516.

[36] C. Greer, M. Burns, D. Wollman, and E. Griffor, "Cyber-physical systems and internet of things," ed, 2019.

[37] H.-D. Ma, "Internet of things: Objectives and scientific challenges," *Journal of Computer science and Technology,* vol. 26, 2011, pp. 919-924.

[38] A. Yeboah-ofori, J.-D. Abdulai, and F. Katsriku, "Cybercrime and risks for cyber physical systems: A review," 2018.

[39] V. Gunes, S. Peter, T. Givargis, and F. Vahid, "A survey on concepts, applications, and challenges in cyber-physical systems," *KSII Transactions on Internet and Information Systems (TIIS),* vol. 8, 2014, pp. 4242-4268.

[40] C. Sonntag, "Towards Enhanced EU-US ICT Pre-competitive Collaboration-Opportunity Report of the EU Project PICASSO.[Hrsg.] C. Sonntag und S. Engell. Revised version V1. 0.1 (March 19, 2017). 2017," ed.

[41] K. Ashton, "That 'internet of things' thing," *RFID journal,* vol. 22, 2009, pp. 97-114.

[42] I. Stojmenovic, and F. Zhang, "Inaugural issue of 'cyber-physical systems'," Cyber-Physical Systems, 1(1), 2015, pp.1-4.

[43] E. A. Lee, "The past, present and future of cyber-physical systems: A focus on models," *Sensors,* vol. 15, 2015, pp. 4837-4869.

[44] S. Sierla, B. M. O'Halloran, T. Karhela, N. Papakonstantinou, and I. Y. Tumer, "Common cause failure analysis of cyber–physical systems situated in constructed environments," *Research in Engineering Design,* vol. 24, 2013, pp. 375-394.

[45] T. Sanislav, G. Mois, S. Folea, L. Miclea, G. Gambardella, and P. Prinetto, "A cloud-based Cyber-Physical System for environmental monitoring," in *2014 3rd Mediterranean Conference on Embedded Computing (MECO)*, 2014, pp. 6-9.

[46] Y. Zhou, Q. Xiao, Z. Mo, S. Chen and Y. Yin, "Privacy-preserving point-to-point transportation traffic measurement through bit array masking in intelligent cyber-physical road systems," in *2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*, pp. 826-833.

[47] J. M. Bradley and E. M. Atkins, "Optimization and control of cyber-physical vehicle systems," *Sensors*, *15*(9), 2015, pp. 23020-23049.

Is Deep Learning on Tabular Data Enough? An Assessment

Sheikh Amir Fayaz[1]
Research Scholar
Department of Computer Sciences
University of Kashmir, J&K, India-190006

Sameer Kaul[3]
Department of Computer Sciences
University of Kashmir
Srinagar, J&K, India-190006

Majid Zaman[2]
Directorate of IT & SS
University of Kashmir
Srinagar, J&K, India-190006

Muheet Ahmed Butt[4]
Department of Computer Sciences
University of Kashmir
Srinagar, J&K, India-190006

***Abstract*—It is critical to select the model that best fits the situation while analyzing the data. Many scholars on classification and regression issues have offered ensemble techniques on tabular data, as well as other approaches to classification and regression problems (Like Boosting and Logistic Model tree ensembles). Furthermore, various deep learning algorithms have recently been implemented on tabular data, with the authors claiming that deep models outperform Boosting and Model tree approaches. On a range of datasets including historical geographical data, this study compares the new deep models (TabNet, NODE, and DNF-net) against the boosting model (XGBoost) to see if they should be regarded a preferred choice for tabular data. We look at how much tweaking and computation they require, as well as how well they perform based on the metrics evaluation and statistical significance test. According to our study, XGBoost outperforms these deep models across all datasets, including the datasets used in the journals that presented the deep models. We further show that, when compared to deep models, XGBoost requires considerably less tweaking. In addition, we can also confirm that a combination of deep models with XGBoost outperforms XGBoost alone on almost all datasets.**

Keywords*—*Deep learning; XGBoost; NODE; TabNet; DNF-net; statistical significance test; tabular geographical data

I. Introduction

Deep learning has gained popularity in a variety of fields in recent years, including medicine, engineering, and agriculture. The exponential growth of data is most likely to blame. Deep learning algorithms have shown to be effective in a variety of domains, including audio [1], images [2], and text data [3]. Many architectures exist in these domains that are capable of converting raw data into meaningful exemplifications. Because the most common type of data is in tabular format, which consists of rows and columns with a variety of parameters, These types of data are used in real-world applications in a variety of fields, including medicine, agriculture, academia, and geography. Traditional and ensemble machine learning approaches, such as Logistic model tree (LMT), Decision tree (DT), Random forest (RF), Gradient Boosted decision tree (GBDT), and others, are used to process these tabular datasets, and these models still outperform deep learning on tabular data. When using a deep learning model on tabular data, there are a number of issues to consider, including missing data, data integrity i.e., mixed data (nominal, numerical, and categorical), data imbalance, data overfitting, and a lack of specific knowledge about the dataset's structure. When tabular data is taken into account, boosting machine-learning algorithms like XGBoost perform better, according to the "no free lunch" (NFL) theorem [4] [5]. Since then, the authors [6] [7] have implemented deep learning on the tabular dataset in their research, and it has been demonstrated that the deep learning model outperforms GBDT. However, because each study was conducted on different datasets, one of the major flaws in their approach is that there was no benchmark dataset [8] [9]. So, based on these papers alone, it's difficult to claim that deep learning always outperforms traditional and ensemble algorithms like GBDT when dealing with tabular data [10].

Since the number of research studies using deep learning on tabular data is growing, there is no standard benchmark model in deep learning from which we can conclude that deep learning always outperforms traditional machine learning on tabular data. As a result, the main goal of this paper is to see if any deep learning model is a good fit for these types of tabular dataset problems. Furthermore, in this paper, we attempt to evaluate the proposed deep learning models on tabular datasets, as well as implement XGBoost on various algorithms, with a focus on a historical geographical dataset from India's Kashmir province [11].

This paper is structured as: Section 2 provides a basic background of deep learning and ensemble models on the tabular data. Next, Section 3 presents the experimental setup where dataset descriptions are presented and furthermore this section defines the implementation details with optimization parameters and statistical significance test. Section 4 defines the experimental results and model evaluation. Section 5 defines the overall working of the paper. Finally, the conclusion and future strategies have been suggested in Section 6.

II. Review of Literature

In this section, we present studies that used deep learning approaches and ensemble approaches to predict rainfall using a tabular geographical dataset. This section is divided into two subsections: Section 1 contains several studies that use deep learning models on tabular datasets, and Section 2 contains some model ensemble approaches that use the same tabular geographical dataset and record individual performances.

A. Deep Learning on Tabular Geographical Dataset

Salman et al. [12] (2015) use a variety of deep learning techniques, including recurrence neural networks (RNN), convolutional neural networks (CNN), and conditional restricted Boltzmann machines (CRBM), to look for hidden patterns in the dataset. These techniques were used in the Indonesian region, with data collected from the national weather service center for environmental forecasting (NOAA). This study used a dataset that spanned 35 years, from 1973 to 2009. Initially, RNN was applied to a dataset containing ESNO variables. RNN produces results with a higher level of accuracy, according to the findings.

Emiley et al [13] (2016) present a deep learning architecture-based accumulated daily rainfall prediction. This research employs auto encoders to reduce non-linear attribute relationships and a multi-layer perceptron (MLP) for prediction. This hybrid architecture was then compared to previously implemented techniques, and it was discovered that the model performs better for daily rainfall prediction when using root mean square error (RMSE) and mean squared error (MSE) statistical approaches. This research was carried out in the Colombian city of Manziles, where the data was grouped into a daily time series spanning the years 2002 to 2013.

Devi et al. [14] (2017) propose an artificial neural network (ANN) model for a reliable forecast mechanism. This method was used to analyze spatial and temporal data from the Nilgiris district in Tamil Nadu, India. Performance was measured using a variety of statistical parameters such as correlation coefficient, MSE, and so on. When compared to time delay neural network (NN) and other ANN models, the best model in this study is a wavelet Elman model. This research also develops a system for early landslide warnings based on the wavelet Elman model.

According to Geetha et al. [15] (2018), using deep learning techniques for meteorological purposes on a time series dataset will significantly improve accuracy precision. This research uses deep learning architectures such as LSTM and ConvNet to analyze time series data from 468 months in various locations. Later, it was discovered that increasing the number of hidden layers improves the model's performance for daily rainfall prediction when using RMSE and MAPE statistical approaches.

Yen et al. [16] (2019) proposed using Echo state network (ESN) and deep Echo state network (DeepESN) algorithms to apply deep learning models to rainfall prediction. This research uses hourly rainfall data from southern Taiwan from 2002 to 2014, spanning a period of 12 years. When the DeepESN algorithm's correlation coefficient was compared to ESN and commercial neuronal network algorithms like BPNN and SVR, the study concluded that it is a reliable algorithm. It was suggested that DeepESN could be used globally on larger sets of data to predict rainfall based on the results obtained.

Manoj et al. [17] (2020) proposed a hybrid deep learning model (BLSTM-GRU), for the monthly prediction of rainfall. The experiment was conducted using data obtained from Bhutan's National Center of Hydrology and Meteorology Department (NCHM). To test the data's predictive capability, various NN algorithms such as LSTM, CNN, BLSTM, and GRU were used. LSTM outperforms the other techniques with a MSE score of 0.0128, but the hybrid model BLSTM-GRU outperforms LSTM by approximately 41% with a MSE score of 0.0075.

Zeelan et al. [18] (2020) claimed that deep learning models can learn from nonlinear data with less error. The Multi-layer perceptron (MLP) and Auto-encoder NN are used in this study to predict the rainfall. The accuracy parameters used in this study were RMSE and MSE, and these implemented models were later compared with other machine learning models on the same set of data, with the study concluding that MLP and Auto-encoder NN perform significantly and can be used as a solution to all available approaches.

Ari Yari et al. [19] (2021) present a rainfall prediction comparative analysis study. The authors use deep learning (DL) models and simple rainfall estimation approaches based on traditional machine learning algorithms. The study was conducted in five major cities across the United Kingdom (UK), with data collected spanning roughly 20 years (2000-2020). The bidirectional LSTM network and stacked LSTM with two hidden layers performed best after the proposed model was evaluated. One of the study's major flaws was the model's inability to generalize the data. That is, the model over-fits the training data in most cases, which makes it difficult to record accurate, predicts in the testing and validation sets.

Razeef et al. [20][21] (2020,2022) proposed a neural network approach to predict the rainfall on the time series data of UT of J&K, India. Rainfall was predicted using a Grey Wolf-based neural network model. The data in this study spans 30 years, from 1990 to 2020, and includes variables such as maximum temperature, humidity, minimum temperature, wind, vapor pressure, and others. When using RMSE, PRD values, and MSE statistical approaches, it was discovered that the model performs better for daily rainfall prediction. This model was later compared to non-linear autoregressive models with exogenous inputs (NARX), and the study concluded that when both models are used together, non-linear time series data would perform better.

According to the literature reviewed in this study, many deep learning models have been utilized for various time-series prediction applications, but they have yet to become a standard algorithm in the artificial intelligence arena. We must analyze the performance of these models using varied threshold datasets, and these techniques must be re-evaluated as a result.

B. Model Ensembles on Tabular Geographical Dataset

Zaman et al. [22] (2019) use an ensemble distributed decision tree (DDT) approach to improve classification accuracy on a historical geographical dataset. The experiment was conducted on a tabular dataset containing approximately 6000 records with five different parameters. When the DDT approach was used, there was no performance improvisation, according to this study.

Patil et al. [23] (2020) use machine learning algorithms to forecast rainfall based on a variety of variables such as temperature, humidity, wind speed, and rainfall. These algorithms include linear regression and NN, and the type of data fed to it, according to the study, determines the accuracy of the algorithm. That is, when the dataset of different structures is used, we may get different accuracies and require some modifications. Furthermore, the accuracy of DT's was found to be superior to other techniques used on the same type of data in this study.

Sheikh et al [24] (2021) proposed a stepwise machine learning approach on the discrete data collected from the Indian Meteorological Department (IMD), Pune India. The implemented model, known as LMT, employs logistic regression functions at the DT's leaf nodes. The logistic functions on the leaf nodes combine the final output of the constructed DT into linear models, which were examined and revealed a significant improvement in accuracy performance. The accuracy of the constructed DT on the same set of data is 66 percent, but when the logistic functions are applied to the leaf nodes, the accuracy jumps to 87 percent. The dataset used in this study was from J&K's Kashmir province, and it covered the years 2012 to 2017, with around 6000 data rows.

Since there are other ensemble [25-29] and deep learning approaches such as NODE, TabNet, DNF-Net, and Boosting (XGBoost, CatBoost, GBDT) [30-33]. These models perform better on larger datasets, and we use the entire training dataset to train the model.

III. Experimental Setup

A. Dataset Description

In this study, we employed a variety of tabular datasets from diverse fields which are used in various classification and regression problems. Some of these datasets have heterogeneous features, while others have just homogeneous features. There are approximately seven tabular datasets that have already been used by various academics in their publications, and we have used one additional dataset that has yet to be used by any researcher. In the experimental operations, roughly 80000 samples were taken, and the datasets range from 7 to 1600 parameters. The seven datasets are obtained from TabNet, NODE, and DNF-Net studies, and each dataset has been well-trained and preprocessed in its respective paper. These datasets include Blastchar [34] (Source: Kaggle), Higgs Boson [35] (Source: Kaggle), Microsoft MSLR [36](Source: MSLR-WEB10K), Forest Cover Type [37] (Source: Kaggle), Epsilon [38] (Source: PASCAL Challenge 2008), YearPrediction [39] (Source: Million Song Dataset) and Gas concentrations (Source: OpenML) [40]. These datasets have also been adjusted and relative values were calculated, resulting in a standardized data with a zero mean value and unit variance. As a result, we won't go into detail about these datasets in this study; instead, we'll just establish the historical geographical dataset that we will be implementing latter. The historical geographical dataset has been collected from three different locations in Jammu and Kashmir's UT. These three locations are in the province of Kashmir, but they are quite far apart. The data spans five years, from 2012 to 2017. At these locations, the average annual rainfall is around 1700 mm. The data consists of 5491 records with a total of 9 explanatory characteristics, including minimum temperature (°C), maximum temperature (°C), station ID, season, year, humidity at various intervals, and the target parameter rainfall, which shows the quantum of rainfall measured in millimeters [41-43].

In Fig. 1, the reader can find a brief description of the data. It has gone through an ETL (Extract, Transform, and Load) process to achieve data integrity, normalization, and standardization.

To normalize the dataset's range of parameters, we use the function (1), as given below:

$$X = \frac{X - \mu}{\sigma} \quad (1)$$

The data was scaled using the R tool's built-in function 'Scale.' We also use relative values of each attribute to normalize the training data. The tabular (Table I) and graphical (Fig. 2) representation of the dataset is shown.

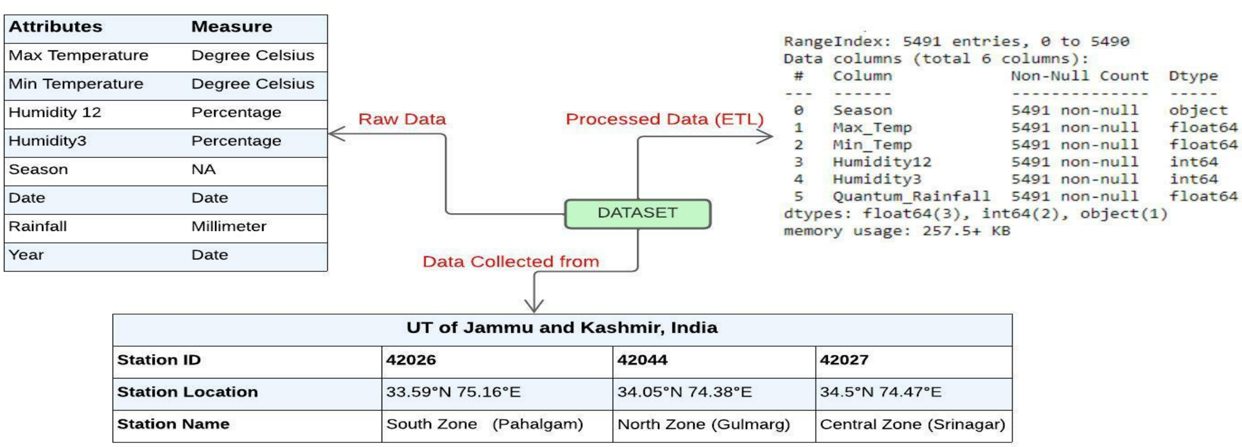


Fig. 1. Historical Geographical Dataset Description.

TABLE I. TABULAR REPRESENTATION OF GEOGRAPHICAL DATASET WITH RELATIVE VALUES

MAX	Relative Frequency (MAX)	MIN	Relative Frequency (MIN)	Hum 12	Relative Frequency (Hum12)	Hum 3	Relative Frequency (Hum3)	Rf	Relative Frequency (Rf)
20	0.010562739	8	0.00910581	60	0.024767802	93	0.032052449	0	0.685303223
19	0.008195229	1.6	0.008923693	50	0.024039337	92	0.032052449	0.4	0.014751411
27.5	0.008013112	11.2	0.008741577	53	0.023675105	88	0.029867055	0.6	0.012930249
29	0.00764888	9	0.008741577	51	0.023128756	91	0.029684939	0.2	0.012019669
14	0.007102531	0.2	0.008741577	55	0.022764524	89	0.029684939	1.2	0.011291204
22	0.006920415	0.4	0.008559461	54	0.022764524	87	0.029502823	0.8	0.009287926
16	0.006738299	10.2	0.008377345	59	0.022218175	85	0.027317429	1.4	0.008195229

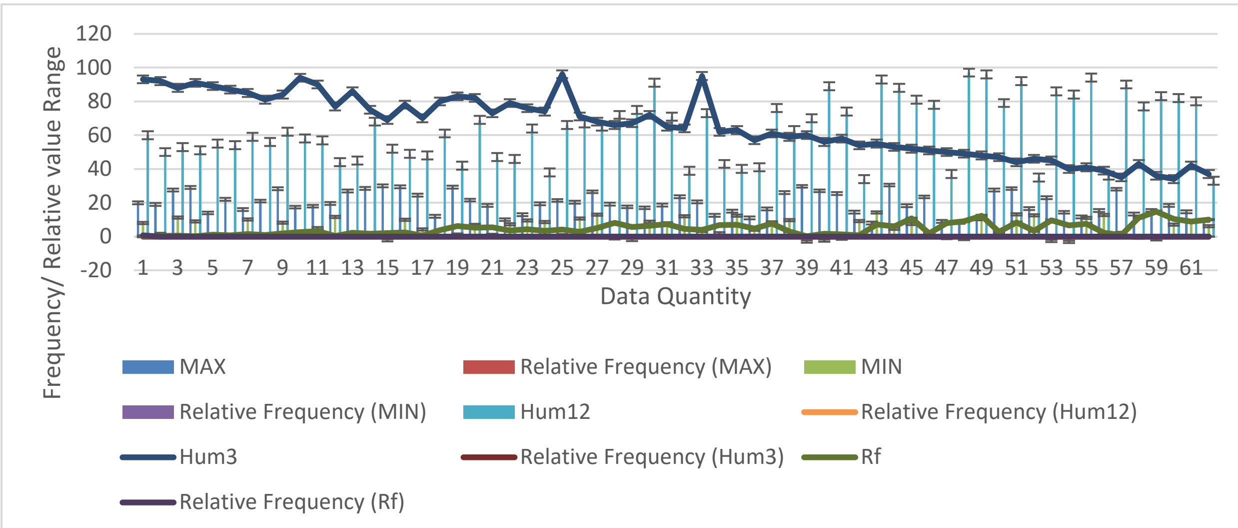


Fig. 2. Graphical Representation of Geographical Dataset with Relative Values.

A total of 70% of the data is used for training, while 15% is used for validation and testing, i.e. 3844 samples were randomly selected for training, 823 samples for validation, and the remaining 823 samples were selected for testing.

Thus, the overall description of the tabular datasets used in this paper is shown in Table II.

B. Implementation Details

1) Optimization process: To pick the model hyperparameters during the optimization phase, we used the HyperOpt parameter-tuning package. To optimize the results on the validation set, this technique first uses Bayesian optimization, followed by hyperparameter search on each dataset utilized in this study. There were around 7-9 main hyperparameters, which in the case of a deep learning model include the number of nodes, layers, and, most importantly, the learning rate.

To optimize the hyperparameters all the datasets used in this study were initially divided into three individual splits, which include training split, testing split and validation split. In partitioning process, we use stratified random sampling partitioning to split the data. The below tabular representation (Table III) shows the individual splits of the datasets in order to optimize the model.

Around 1000 steps of search were performed on each set of data in order to maximize the validation set's findings, and only the set of hyperparameters with the smallest loss for the final configuration were chosen.

2) Metrics evaluation and statistical significance test: In the case of classification issues using discretized data, we simply utilize cross-entropy loss metrics to evaluate the datasets. It calculates the impurity at each stage of the data and the total entropy loss in the end. Furthermore, when the data is continuous in nature, such as in regression situations, statistical parameters such as RMSE, mean signed difference (MSD), and MAE are used. We reported the performance of each dataset on their respective test sets based on these metrics. We also have Friedman's testing for statistical significance in addition to these cross entropy and RMSE, MAE measurements. Friedman's testing has the advantage of assuming that data is not evenly distributed. Using Friedman's hypothesis, we compare all of the classifiers to the baseline classifier. The null hypothesis is rejected at a certain level of confidence (90 percent in this study) if the p-value for any model pair is less than 0.05; otherwise, the hypothesis is not rejected.

TABLE II. TABULAR DATASETS DESCRIPTION

Dataset	Parameters	Records (Approx.)	Source	Research Paper	Hyperparameters and search spaces used for configuring each algorithms implemented	Ref
Blastchar	20	7000	Kaggle	DNF-Net	• Discrete uniform distribution for n. formulas and Number of layers. • Learning rate with log-distribution. • Batch size with Uniform choice.	[34]
Higgs Boson	30	80000	Kaggle	TabNet	• Log-Uniform distribution for Learning rate • Discrete uniform distribution for feature dimensions, n steps and output dimensions. • Uniform distribution for relaxation factor, Batch size and bn epsilon.	[35]
Microsoft(MSLR)	126	75000	MSLA-WEB10K	NODE	• Log-Uniform distribution for learning rate. • Discrete uniform distribution for Num Layers, tree output dimensions and tree depth. • Uniform choice for Batch size.	[36]
Forest Cover Type	50	55000	Kaggle	TabNet	• Log-Uniform distribution for Learning rate • Discrete uniform distribution for feature dimensions, n steps and output dimensions. • Uniform distribution for relaxation factor, Batch size and bn epsilon.	[37]
Epsilon	1700	50000	PASCAL Challenge 2008	NODE	• Log-Uniform distribution for learning rate. • Discrete uniform distribution for Num Layers, tree output dimensions and tree depth. • Uniform choice for Batch size.	[38]
Year Prediction	90	51500	Million Song Dataset	NODE	• Log-Uniform distribution for learning rate. • Discrete uniform distribution for Num Layers, tree output dimensions and tree depth. • Uniform choice for Batch size.	[39]
Gas Concentration	129	13900	OpenML	DNF-Net	• Discrete uniform distribution for n. formulas and Number of layers. • Learning rate with log-distribution. • Batch size with Uniform choice.	[40]
Historical geographical Dataset	9	5491	IMD	New Dataset	---	[41-43]

TABLE III. TABULAR REPRESENTATION WITH TRAINING TESTING AND VALIDATION SPLITS

Dataset	Records	Training	Testing	Validation
Blastchar	7000 (100%)	5600 (80%)	---	1400 (20%)
Higgs Boson	80000 (100%)	50000 (62%)	---	30000(38%)
Microsoft(MSLR)	75000 (100%)	60000 (80%)		15000 (20%)
Forest Cover Type	55000 (100%)	38500 (70%)	8250 (15%)	8250 (15%)
Epsilon	50000 (100%)	40000 (80%)	---	10000 (20%)
Year Prediction	51500 (100%)	41200 (80%)	---	10300 (20%)
Gas Concentration	13900 (100%)	9730 (70%)	2780 (20%)	1390 (10%)
Historical geographical Dataset	5491 (100%)	3844 (70%)	823 (15%)	823 (15%)

IV. EXPERIMENTAL RESULTS

A. How Effectively can Deep Learning Models Generalize to other Datasets?

The performance of deep learning models on the aforementioned datasets is proposed in this study, and the individual outcomes are compared to the XGBoost technique. The performance of each algorithm on each dataset is presented in the table below. The mean and standard error of each model's performance on the datasets are shown in the Table IV. The best performance of the dataset is presented for each model, and it was discovered that the model with the lowest value is considered to have the best performance. Friedman's testing was utilized to perform a statistical significance test between the models with a 90% confidence level.

TABLE IV. RESULTS AND PERFORMANCE OF EACH TABULAR DATASET BASED ON EACH MODEL USED IN THIS STUDY. FOR YEARPREDICTION MSE IS USED AND CROSS ENTROPY IS USED FOR ALL OTHER DATASETS. THE VALUES WITH LOWER VALUE IS BETTER AND THESE VALUES ARE THE AVERAGES OF DIFFERENT TRAINING RUNS WITH STANDARD ERROR OF MEAN (SEM)

Model	**Blastchar**	**Higgs Boson**	**Microsoft (MSLR)**	**Forest Cover Type**	**Epsilon**	**Year Prediction**	**Gas Concentration**	**Historical Geographical Dataset**
NODE	21.36 ± 0.23	21.21 ± 0.67	54.62 ± 3e-2	4.25 ± 0.17	**10.26 ± 1e-2**	76.88 ± 0.16	2.25 ± 0.22	14.76 ± 0.12
DNF-Net	27.91 ± 0.18	23.71 ± 0.88	55.78 ± 3e-2	4.01 ± 0.09	12.42 ± 4e-2	82.06 ± 0.15	**1.45 ± 0.08**	15.36 ± 0.18
TabNet	23.66 ± 0.16	**21.15 ± 0.22**	55.09 ± 2e-2	**3.02 ± 0.15**	11.96 ± 3e-2	82.89 ± 0.11	1.86 ± 0.10	14.62 ± 0.16
XGBoost	20.41 ± 0.23	21.83 ± 0.34	54.39 ± 2e-2	3.21 ± 0.11	11.23 ± 2e-2	75.68 ± 0.08	2.06 ± 0.32	13.45 ± 0.19
Simple Ensemble	21.22 ± 0.15	22.49 ± 0.41	54.44 ± 3e-2	4.15 ± 0.16	11.38 ± 4e-2	78.65 ± 0.16	2.41 ± 0.18	13.67 ± 0.15
Deep Ensemble & XGBoost	**20.13 ± 0.16**	22.36 ± 0.51	**54.21 ± 1e-2**	2.86 ± 0.05	11.35 ± 1e-2	**75.01 ± 0.22**	1.66 ± 0.06	**12.13 ± 0.15**
Deep Ensemble w/o XGBoost	24.36 ± 0.31	22.45 ± 0.55	55.53 ± 3e-2	3.57 ± 0.11	10.88 ± 1e-2	79.01 ± 0.17	1.91 ± 0.17	14.15 ± 0.14

There are some observations based on the results, as given in the table. To begin with, the models almost outperform unknown datasets on original datasets. On each dataset, the XGBoost model nearly outperformed all deep learning models such as NODE, DNF-Net, and TabNet. As we can see, the XGBoost model outperforms deep learning models in 5 of the 8 datasets, and these datasets had significant p-values (< 0.05), indicating that the results were significant. We can also see that the deep learning model has not consistently performed. The authors claimed in their study that deep learning models outperform other models, but this was only true for the datasets included in their study. As a result, when distinct datasets are involved, this conclusion is unjustifiable. We can also observe that the Deep ensemble and XGBoost model beats individual models in the majority of cases, i.e. it outperforms 5 individual models out of 8, and the p-value in these 5 cases was substantially less than 0.05, indicating that the null hypothesis is rejected.

Now, in order to evaluate these models and see which one is better for a given dataset, we compared the relative performance of each model (NODE [44], TabNet [45], DNF-Net [46][47], and so on) to the best model for that dataset. For example, assume we used the historical geographical dataset in table (Table IV) and compared the relative performance of the models to choose the model with the best performance (Deep Ensemble & XGBoost in this case). We discovered that Deep Ensemble & XGBoost had the best relative value gain of 2.46 percent, with XGBoost coming in second with 3.86 percent, TabNet with 8.67 percent, DNF-Net with 10.55 percent, and NODE with 13.23 percent. The tabular representation of average relative performance deterioration on unseen datasets is shown (Table V).

With these findings, we discovered that deep learning does not always outperform other methods. When compared to XGBoost, Deep Ensemble, and XGBoost, the deep learning model performs the worst when trained on datasets other than those used in the original studies. Only two choices exist for the lowest performance results. Either there is a selection bias or there is a difference in hyperparameter optimization. Furthermore, the results in the original papers reflect the results that we have reported, excluding the possibility that implementation errors were the cause of our observation.

B. *Model Evaluation: Is it required to apply both the XGBoost and Deep Models in Combination?*

In this section, we will see which model performs better in all scenarios when compared to other models. We employed four types of models in table (Table V), including deep models (TabNet, DNF-Net, and NODE), XGBoost, Deep ensemble with XGBoost, and Deep ensemble without XGBoost. When comparing the performance of deep learning models to XGBoost and combined Deep ensemble & XGBoost Models on various data types, we discovered that deep learning models perform poorly in most circumstances. The question now is whether we require a combined XGBoost and Deep model. To answer this, we can see that in 6 of the 8 examples, the combined ensemble and XGBoost show significant results. Simple ensemble did not produce any improvised results, although competing with deep learning model results. Furthermore, when we look at the Deep ensemble models without XGBoost, we can observe that it did not do well in any situation when compared to any other model. As a result of this analysis, we can conclude that for tabular datasets, we require both deep ensemble and XGBoost in combination.

TABLE V. AVERAGE VALUES OF ALL THE MODELS IMPLEMENTED ON EACH DATASET WITH LOWER VALUE TREATED AS BEST

Model	**Average Relative performance (%)**
Deep Ensemble w/o XGBoost	7.10 %
XGBoost	3.86%
Deep Ensemble & XGBoost	2.46%
TabNet	8.67%
DNF-Net	10.55%
NODE	13.23%
Simple Ensemble	4.23%

In real-world situations, time and resources are limited when it comes to training a model for a new dataset and optimizing its hyperparameters. As a result, it's fascinating to learn how difficult it is to do so for each model. Calculating the number of computations required by the model is one way to assess this. Floating point operations per second (FLOPS) is a common unit of measurement. However, because each parameter set has a different FLOPS number, comparing various models in this way when optimizing model parameters has become impossible [47].

V. Discussion

This study was based on deep models that had already been deployed by several academics on a tabular dataset [12-14]. Deep models were applied to tabular datasets (Forest CoverType, Higgs, Gas Concentration, Epsilon [30], MSLR [31], Year Prediction [32], Blastchar [33], and so on) by the authors in their publications, and they argued that deep models exhibit some promising outcomes. However, their research was limited to a single dataset. We used one more tabular dataset (Geographical dataset) in this research and attempted to construct all of the deep learning and ensemble models. On all of the datasets utilized in this study, we also investigated various possible tradeoffs that are required in real-time applications, such as hyperparameter tuning, metrics evaluation, and Statistical Significance test. Our results reveal that the performance is similar to what the authors have shown in their respected publications, but when we tried to compare the performance of different datasets on the models used by the authors in their study, the deep learning results were not as good as the original datasets. We next looked at XGBoost and ensembles of deep models with XGBoost and without XGBoost, and discovered that the XGBoost model outperforms deep models. However, as seen in the table, the ensemble of XGBoost models with Deep models outperforms the XGBoost model alone. Furthermore, optimizing a new dataset using deep models is a difficult procedure, whereas optimizing a new dataset using ensemble models with XGBoost is quite simple [48].

VI. Conclusion and Future Strategies

This research demonstrates that using various deep learning algorithms on tabular data does not improve performance. We also used XGBoost on these datasets, which produced some promising results when compared to deep models, and we used ensemble deep learning with and without XGBoost to see how it affected the performance of each dataset. On these tabular datasets, an ensemble of XGBoost models without deep learning never performed well, but when we looked at the overall performance using an ensemble of deep models with XGBoost, the results were astounding. This ensemble deep model with XGBoost beats all previous models, and our enhanced models pave the way for future study on tabular datasets in terms of comparing performance and assisting researchers in determining the best technique for optimizing hyperparameters. Our findings will also aid in the development of new models (such as CatBoost, where learning rates are uniformly distributed) that are simple to optimize and can compete with the performance of ensemble deep models such as XGBoost and many others.

References

[1] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: NAACL, 2019.

[2] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[3] van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, K. Kavukcuoglu, WaveNet: A generative model for raw audio, in: Proc. 9th ISCA Workshop on Speech Synthesis Workshop (SSW 9), 2016, p. 125.

[4] Adam, Stavros P., Stamatios-Aggelos N. Alexandropoulos, Panos M. Pardalos, and Michael N. Vrahatis. "No free lunch theorem: A review." Approximation and optimization (2019): 57-82.

[5] Popov, Sergei, Stanislav Morozov, and Artem Babenko. "Neural oblivious decision ensembles for deep learning on tabular data." arXiv preprint arXiv:1909.06312 (2019).

[6] Gorishniy, Yury, Ivan Rubachev, Valentin Khrulkov, and Artem Babenko. "Revisiting deep learning models for tabular data." Advances in Neural Information Processing Systems 34 (2021).

[7] Marais, Jan André. "Deep learning for tabular data: an exploratory study." PhD diss., Stellenbosch: Stellenbosch University, 2019.

[8] Shwartz-Ziv, Ravid, and Amitai Armon. "Tabular data: Deep learning is not all you need." Information Fusion 81 (2022): 84-90.

[9] Richman, Ronald, and Mario V. Wuthrich. "LocalGLMnet: interpretable deep learning for tabular data." Available at SSRN 3892015 (2021).

[10] Luo, Renqian, Xu Tan, Rui Wang, Tao Qin, Enhong Chen, and Tie-Yan Liu. "Neural architecture search with gbdt." (2020).

[11] Zaman, Majid, Sameer Kaul, and Muheet Ahmed. "Analytical comparison between the information gain and Gini index using historical geographical data." Int. J. Adv. Comput. Sci. Appl 11, no. 5 (2020): 429-440.

[12] Salman, Afan Galih, Bayu Kanigoro, and Yaya Heryadi. "Weather forecasting using deep learning techniques." In 2015 international conference on advanced computer science and information systems (ICACSIS), pp. 281-285. Ieee, 2015.

[13] Hernández, Emilcy, Victor Sanchez-Anguix, Vicente Julian, Javier Palanca, and Néstor Duque. "Rainfall prediction: A deep learning approach." In International Conference on Hybrid Artificial Intelligence Systems, pp. 151-162. Springer, Cham, 2016.

[14] Renuga Devi S., Arulmozhivarman P., Venkatesh C. (2017) ANN Based Rainfall Prediction—A Tool for Developing a Landslide Early Warning System. In: Mikoš M., Arbanas Ž., Yin Y., Sassa K. (eds) Advancing Culture of Living with Landslides. WLF 2017. Springer, Cham. https://doi.org/10.1007/978-3-319-53487-9_20.

[15] Aswin, S., P. Geetha, and R. Vinayakumar. "Deep learning models for the prediction of rainfall." In 2018 International Conference on Communication and Signal Processing (ICCSP), pp. 0657-0661. IEEE, 2018.

[16] Yen, Meng-Hua, Ding-Wei Liu, Yi-Chia Hsin, Chu-En Lin, and Chii-Chang Chen. "Application of the deep learning for the prediction of rainfall in Southern Taiwan." Scientific reports 9, no. 1 (2019): 1-9.

[17] Chhetri, Manoj, Sudhanshu Kumar, Partha Pratim Roy, and Byung-Gyu Kim. "Deep BLSTM-GRU model for monthly rainfall prediction: A case study of Simtokha, Bhutan." Remote sensing 12, no. 19 (2020): 3174.

[18] Basha, Cmak Zeelan, Nagulla Bhavana, Ponduru Bhavya, and V. Sowmya. "Rainfall prediction using machine learning & deep learning techniques." In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 92-97. IEEE, 2020.

[19] Barrera-Animas, Ari Yair, Lukumon O. Oyedele, Muhammad Bilal, Taofeek Dolapo Akinosho, Juan Manuel Davila Delgado, and Lukman Adewale Akanbi. "Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting." Machine Learning with Applications 7 (2022): 100204.

[20] Mohd, Razeef, Muheet Ahmed Butt, and Majid Zaman Baba. "Grey Wolf-Based Linear Regression Model for Rainfall Prediction."

International Journal of Information Technologies and Systems Approach (IJITSA) 15, no. 1 (2022): 1-18.

[21] Mohd, Razeef, Muheet Ahmed Butt, and Majid Zaman Baba. "GWLM–NARX: Grey Wolf Levenberg–Marquardt-based neural network for rainfall prediction." Data Technologies and Applications (2020).

[22] Fayaz, Sheikh Amir, Majid Zaman, and Muheet Ahmed Butt. "To ameliorate classification accuracy using ensemble distributed decision tree (DDT) vote approach: An empirical discourse of geographical data mining." Procedia Computer Science 184 (2021): 935-940.

[23] Patil, Deepali, Shakib Badarpura, Abhishek Jain, U. G. Student, and Aniket Gupta. "Rainfall Prediction using Linear approach & Neural Networks and Crop Recommendation based on Decision Tree."

[24] Fayaz, Sheikh Amir, Majid Zaman, and Muheet Ahmed Butt. "An application of logistic model tree (LMT) algorithm to ameliorate Prediction accuracy of meteorological data." International Journal of Advanced Technology and Engineering Exploration 8, no. 84 (2021): 1424.

[25] Altaf, Ifra, Muheet Ahmed Butt, and Majid Zaman. "A Pragmatic Comparison of Supervised Machine Learning Classifiers for Disease Diagnosis." In 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 1515-1520. IEEE, 2021.

[26] Fayaz, Sheikh Amir, Ifra Altaf, Aaqib Nazir Khan, and Zahid Hussain Wani. "A possible solution to grid security issue using authentication: an overview." J. Web Eng. Technol 5, no. 3 (2019): 10-14.

[27] Zaman, Majid, and Muheet Ahmed Butt. "Information translation: a practitioners approach." In World Congress on Engineering and Computer Science (WCECS). 2012.

[28] Ashraf, Mudasir, Syed Mudasir Ahmad, Nazir Ahmad Ganai, Riaz Ahmad Shah, Majid Zaman, Sameer Ahmad Khan, and Aftab Aalam Shah. "Prediction of cardiovascular disease through cutting-edge deep learning technologies: an empirical study based on TENSORFLOW, PYTORCH and KERAS." In International Conference on Innovative Computing and Communications, pp. 239-255. Springer, Singapore, 2021.

[29] Ashraf, Mudasir, Majid Zaman, and Muheet Ahmed. "Performance analysis and different subject combinations: an empirical and analytical discourse of educational data mining." In 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence), pp. 287-292. IEEE, 2018.

[30] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, LightGBM: A highly efficient gradient boosting decision tree, in: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems, Vol. 30, Curran Associates, Inc., 2017, URL https://proceedings.neurips.cc/paper/2017/file/ 6449f44a102fde848669bdd9eb6b76fa-Paper.pdf.

[31] L. Prokhorenkova, G. Gusev, A. Vorobev, A.V. Dorogush, A. Gulin, Catboost: Unbiased boosting with categorical features, in: Proceedings of the 32nd International Conference on Neural Information Processing Systems, in: NIPS'18, Curran Associates Inc., Red Hook, NY, USA, 2018, pp. 6639–6649.

[32] Y. Zhao, G. Chetty, D. Tran, Deep learning with XGBoost for real estate appraisal, in: 2019 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, 2019, pp. 1396–1401.

[33] S. Ramraj, N. Uzir, R. Sunil, S. Banerjee, Experimenting XGBoost algorithm for prediction and classification of different datasets, Int. J. Control Theory Appl. 9 (2016) 651–662.

[34] D. Dua, C. Graff, UCI machine learning repository, 2017, URL http://archive.ics.uci.edu/ml.

[35] Kaggle, Rossmann store sales, 2019, URL https://www.kaggle.com/c/ rossmannstore-sales.

[36] J. Vanschoren, J.N. Van Rijn, B. Bischl, L. Torgo, OpenML: networked science in machine learning, ACM SIGKDD Explor. Newsl. 15 (2) (2014) 49–60.

[37] T. Qin, T. Liu, Introducing LETOR 4.0 datasets, 2013, CoRR http://arxiv.org/abs/1306.2597.

[38] Pascal, Pascal large scale learning challenge, 2008, URL https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary.html.

[39] Kaggle, Churn modelling, 2019, URL https://www.kaggle.com/shruti mechlearn/churn-modelling.

[40] IBM, Telco customer churn, 2019, https://community.ibm.com/ community/user/businessanalytics/blogs/steven-macko/2019/07/11/telco -customer-churn-1113.

[41] Fayaz, Sheikh Amir, Majid Zaman, and Muheet Ahmed Butt. "Performance Evaluation of GINI Index and Information Gain Criteria on Geographical Data: An Empirical Study Based on JAVA and Python." In International Conference on Innovative Computing and Communications, pp. 249-265. Springer, Singapore, 2022.

[42] Fayaz, Sheikh Amir, Majid Zaman, and Muheet Ahmed Butt. "Knowledge Discovery in Geographical Sciences—A Systematic Survey of Various Machine Learning Algorithms for Rainfall Prediction." In International Conference on Innovative Computing and Communications, pp. 593-608. Springer, Singapore, 2022.

[43] Sidiq, S. Jahangeer, Majid Zaman, and Muheet Ahmed. "How Machine Learning is Redefining Geographical Science: A Review of Literature." (2019).

[44] S. Arik, T. Pfister, Tabnet: Attentive interpretable tabular learning, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 35, (8) 2021, pp. 6679–6687, URL https://ojs.aaai.org/index.php/AAAI /article/view/16826

[45] L. Katzir, G. Elidan, R. El-Yaniv, Net-DNF: Effective deep modeling of tabular data, in: 9th International Conference on Learning Representations, ICLR 2021,Virtual Event, Austria, May 3-7, 2021, OpenReview.net, 2021, URL https://openreview.net/forum?id=73WTG s96kho.

[46] S. Popov, S. Morozov, A. Babenko, Neural oblivious decision ensembles for deep learning on tabular data, in: 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020, OpenReview.net, 2020, URL https://openreview.net/forum?id=r1eiu2VtwH.

[47] Kaul, S., Fayaz, S.A., Zaman, M., Butt, M.A. (2022). Is decision tree obsolete in its original form? A Burning debate. Revue d'Intelligence Artificielle, Vol. 36, No. 1, 105-113. https://doi.org/10.18280/ria. 360112.

[48] Fayaz, S. A., Zaman, M., & Butt, M. A. (2022). Numerical and Experimental Investigation of Meteorological Data Using Adaptive Linear M5 Model Tree for the Prediction of Rainfall. Review of Computer Engineering Research, 9(1), 1-12.

Smart Home Automation by Internet-of-Things Edge Computing Platform

Zubair Sharif[1], Low Tang Jung[2], Muhammad Ayaz[3], Mazlaini Yahya[4], Dodo Khan[5]
Computer and Information Sciences Department (CISD), UTP Perak, Malaysia[1, 2, 5]
Sensor Networks and Cellular Systems (SNCS) Research University of Tabuk, Tabuk 71491, Saudi Arabia[3]
Head IoT Automation, Petronas, Malaysia[4]

***Abstract*—Internet of Things (IoTs) offer significant benefits to various applications, including homes automation (HA), environmental monitoring, healthcare, homeland security, agriculture, and many others. Consequently, the trend of IoTs is rapidly evolving in many new sectors leading to higher comfort, better quality of life, and conveniences that can offer optimum consumption of valuable resources for the users. This paper presents a low-cost and flexible HA system that uses different sensors and other resources to control commonly used home appliances and connected devices by establishing IP connectivity to access, manage, and monitor them remotely. To enable remote access, an android-based application was developed to monitor and control the home devices. Edge computing (EC) platform is used in the proposed system to enhance reliability and robustness when the computation offloading is required. The system operates automatically according to the environmental conditions in the home based on the home occupants' requirements. The outcomes of the proposed system reveal that it greatly supports the concept of smart HA and capable to significantly reduces the wastage of electricity via optimum utilization.**

Keywords—Smart home; Internet of Things (IoTs); home automation; android application; raspberry pi; edge computing

I. Introduction

Recent advancement in smart phones and inexpensive open-source hardware platforms has empowered the development of low-cost home automation (HA) systems using Internet of Things (IoTs). This decade has seen an enormous attention and interest in sensors and sensor networks. Billions of devices are connected to internet already and further increasing sharply with the passage of time. Considering the current progress and the importance of IoTs, 31 billion devices have been connected in 2020 and this figure can reach to 75 billion by 2025 [1-3] as detailed facts are represented in Fig. 1. The devices include smart appliances (lights, fans, refrigerators, air conditioners, ovens, washers etc.), security and safety systems (PIR motion detector sensors, monitoring cameras, smoke and fire detectors sensors etc.), and other smart home energy equipment, like smart lighting, thermostats, and actuators.

HA is a way of life by utilizing wireless sensors and other modern smart technologies that can make a home to accomplish different sets of operations or tasks automatically and remotely whenever they are required. Based on recent success of IoTs for HA system, it is becoming reality in improving the quality of living by making it more comfortable due to many embedding intelligences into sensors and actuators and connecting them intelligently for smart homes. The IoT market has generated revenue around $300 billion in 2020, and the value of Smart Home market worldwide has reached to $43 Billion. This is nearly three times more than that in 2014 [4-6].

Homes and offices chores can be automated by using the internet to control and monitor the household devices by considering the environmental conditions and the human comfort. Many vendors have introduced and producing a range of HA devices to regulate and control the indoor ventilation or environment. For this a variety of wireless sensors and home controlling products including home energy monitor, smart plugs, smart temperature thermostats, outlets, sockets, and smart vents etc. are available. Remote controlling or monitoring is not only the facility of HA system but more importantly it helps to minimize operating costs and saves electricity or other resources. Moreover, it can be ideal for the aged and disabled people [7].

The edge computing (EC) platform has a lot of advantages and useful applications related to smart HA. Its prominent benefits are minimizing the latency issues, providing location awareness and mobility support, real time response, less cost for data processing in terms of less bandwidth consumption, home inside environment control without the home resident's intervention, timely weather predictions and many more. In IoTs, each equipment can consume a huge amount of energy if its communication is not optimized [8, 9]. Further, many applications are response sensitive so to satisfy time critical applications, EC can be greatly helpful. In addition, for the smart devices, local computation is cheaper in terms of energy consumption due to less physical distance for data communication because devices may generate data at every single moment and required processing on the collected data. During their data transmission, they required more energy if the transmission time is high, and they utilize more communication channels as well. For instance, when CCTV cameras' data is stored in the cloud, it means far away from the end devices that it requires more bandwidth and more time for data transmission comparatively to the EC. So, in the case of EC, due to less distance it takes less time and consumes less bandwidth for data transmission. In this example EC is therefore saving resources in terms of energy, time, and communication channel cost etc. [10].

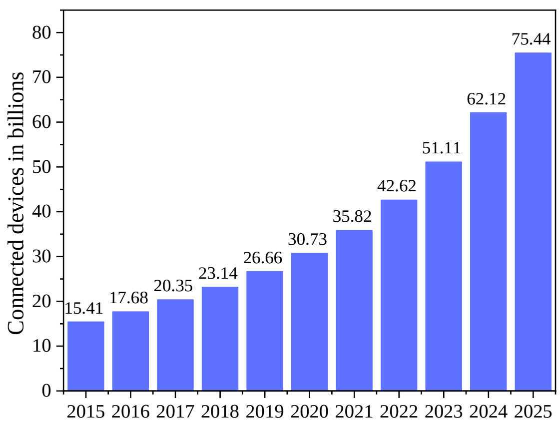


Fig. 1. The Current and Expected Number of Devices that will be connected to Internet on Yearly basis.

In this paper we present a HA system using IoTs, android application, that involving the EC platform. The significance of our designed system is that there is no need to change or replace the current household appliances, as it offers options to connect these devices through a control panel. The main component in the control panel is Raspberry pi which controls all the home appliances. Several sensors are connected/integrated to Raspberry pi for it to receive various measurements like temperature, humidity, light intensity etc., for which it can control the appliances in considering the resident requirements. The whole system consists of commonly used home appliances (lights, fans, switches, power plugs, air conditioners etc.) different sensors (light sensor, PIR sensor, temperature sensor etc.), Raspberry pi 4, power relays, and a touch screen panel as shown in Fig. 5. The effectiveness of the designed system (i.e., for integrated sensors with the home devices and switches), reveals that it greatly achieves the concept of smart HA.

A. Some Major Applications of HA

Some major applications are discussed here to support the concept of HA. With the help of IoTs, it can recognize or notice the resident's certain habits or activities e.g. exercising, showering, sleeping times, hence IoT can regulate the home interior environment accordingly [11] such as changing the temperature or environment of a house, by smartly controlling windows/door entry points. Doors can be operated automatically only when residents enter or leave the room. Same concept can be applied to the car garage. Windows shutters can be adjusted according to the outside temperature and light intensity, or closed fully in case of raining [12, 13].

Security or safety is always the prime concern of any house. It can be greatly improved by IoT concepts. For example, human presence or any moving object, particularly at the room entrances and the main gate, can be detected by using PIR motion detection sensor, hence enhancing the security. The safety of home residents can certainly be attained efficiently by installing some sensors for smoke or fire detection, sensing the odd materials in water, and storms prediction etc. An automated safety alarm can mitigate the loss or harm based on the received information. Security and safety systems can be beneficial to elderly and disabled people for them to control or react appropriately to the home appliances/devices during critical conditions [14].

Home gardens, lawns, indoor plants can be made smarter by IoT. Sensor(s) can be fixed in root of plant or in the soil to identify the moisture level to trigger a motor pump to irrigate/water the plants [15]. Other than that home devices can be controlled according to an individual routines and behaviors. For example, music can be turned on/off or its volume can be tuned depending on the routines or schedules of the home residents. Likewise, it is possible to define morning or evening routines to adjust the bathroom water heater or windows shatters to operate in considering the weather conditions [16, 17]. The aforementioned and some other major examples of IoTs are illustrated in Fig. 2.

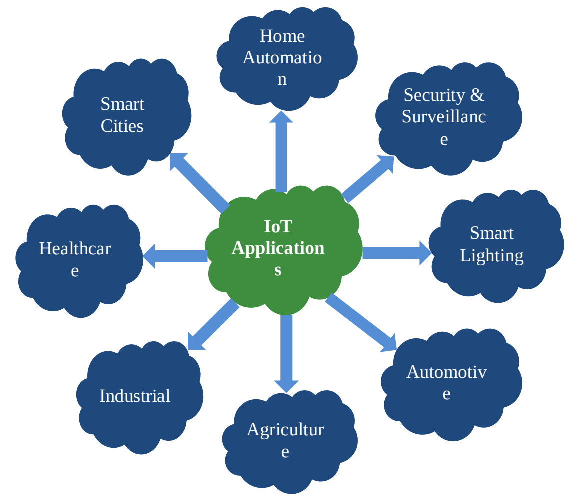


Fig. 2. Major IoT Application Domains.

II. RELATED WORK

The use of IoTs to control the home appliances to achieve the concept of smart home keep attracting the researchers and engineers [18, 19]. Adding IoT based facilities to home environment can enhance the quality of life for home resident and proving blessing for the elderly and disabled people that depends on caregivers or institutional care [20]. There are various wireless technologies suitable for remote data transfer such as Wi-Fi, RFID, and cellular connectivity has been utilized to embed various levels of intelligence in the HA. Further, by introducing the IoTs, the research and implementation of HA are getting more and more attention from the research community.

An IoT based system was proposed in [21] to reduce the cost and power consumption by utilizing the solo computing platform. The system used Arduino and Android smart phones through virtual buttons and sliding switches, user could control and observe the home devices using text and voice commands. A similar work was presented in [12] to provide the facility for monitoring smart home and was especially useful for the disabled persons. The proposed system was customizable to allow changes be made according to the needs of future requirements.

In [22], the Bluetooth 4.0 protocol was used to provide the communication among the users and smart home devices.

Users could control the home devices with their mobile phones or tablets. The drawback of the system was the control was restricted to a short range due to Bluetooth technology. To conserve the home energy, the authors have designed a system by using different sensors like temperature sensors, motion detection sensors and luminance sensors to automate the home appliances like air conditioners, fans, lights, etc. and the proposed system can save electricity usage as well [23].

A reliable and low cost HA system using Arduino microcontroller and various other sensors was offered in [24] to access and control the home devices (fans, lights, TV, etc.) with the help of smart phone application or web-browser. Similarly, another HA system was presented in [25] which used a micro-web server at home side with the IP connectivity allowing user to interact home devices with the smart phone. Different sensors were used like current sensor and temperature sensor to automate the home appliances. Further, authors proposed [26] a system to control the home appliances for HA. In the proposed system four types of sensors were considered i.e., temperature sensor, PIR sensor, smoke and gas sensor, and ultrasonic sensor to automate the home environment and intrusion detection.

For the home security purposes, many systems have been proposed, few of them are discussed here. In [27], authors designed a system to improve the home security by utilizing the cameras to capture pictures to share the information on who entered the house or room. Another system was presented in [28], to enhance the home security by using some sensors, and the advantage of this system was cost effectiveness. Data could be transferred between the devices and controllers by using one or more communicating technologies like Wi-Fi, GSM, ZigBee or Ethernet. ZigBee and Bluetooth were used for data sending among the sensors and CPU. These data sharing technologies were used due to easy implementation and low power usage.

Involving the edge/fog computing in smart home system can be very effective and useful for computation offloading to minimize the latency and gives good real time response. So, researchers have integrated edge/fog computing in their smart HA systems. In [29], authors presented a framework for HA system by involving the fog computing to achieve the concept of energy optimization. With their proposed system, the network traffic and delays were minimized for smart homes. In addition, by involving the EC platform, a system for electrical demand prediction for the smart homes was presented in [30]. An intelligent edge node was used for their system that stored heterogeneous data, but the analysis and processing were done at the cloud node. The system can collect data by the IoT devices from the environment, and provide a better quality of service and greater scalability of computing resources through the EC.

Further in [31], the authors proposed an IoT based system by involving the EC for the smart buildings and homes. In the designed system, the facilities like temperature control, energy consumption monitoring, safety, comfort level were considered. The system addressed issues like scalability and interoperability. A brief comparison of characteristics of IoT, edge and cloud computing are presented in Table I.

TABLE I. CHARACTERISTICS OF IOT, EDGE AND CLOUD COMPUTING [32]

Characteristics	**IoT**	**Edge**	**Cloud**
Deployment	Distributed	Distributed	Centralized
Components	Physical devices	Edge nodes	Virtual resources
Computational	Limited	Limited	Unlimited
Storage	Small	Limited	Unlimited
Response time	NA	Fast	Slow
Big data	Source	Process	Process

Many literatures have made significant contributions to the design and development of HA systems. However, most of these efforts mainly focused on switching and controlling the home appliances rather than remotely monitoring the home environment. The system provided in this article carried a unique architecture for IoT based HA with the help of an android application which saves energy via optimum utilization and reduces the latency when the data is required for computation offloading, hence can be beneficial irrespective of regional or location dependencies.

III. PROPOSED SYSTEM

Considering the factors and challenges discussed in literature review, we designed a low cost, flexible and standalone home monitoring and controlling system. The developed system contains different controllers, relay modules, Raspberry Pi with Wi-Fi support, and various sensors. The main equipment and devices used are: light sensor, air sensor, PIR (passive infrared sensor), smoke detecting sensor, hall effect sensor, Raspberry Pi, display panels (touch screen), and Android application.

To manage and control the home appliances or against the data measurements by various sensors, a smart interface is designed with the help of an Android-based application to enhance the system usability. The significant feature of the system is that it is customizable according to various requirements and scenarios with little configurations and minimum re-coding. The major advantage of our system is that there is no need for replacement of the current household appliances, the existing devices can be connected to a control panel forming part of the smart HA system.

In the designed system, we utilize the Hall Effect sensor and passive infrared (PIR) sensor to observe and monitor moving activities when these are required. The mentioned sensors are positioned or installed at the entry and exits points to and from the rooms and relevant home space. In addition, these activity monitoring sensors can be useful in dealing with suspicious or intrusion activities. To identify the light intensity, light sensor is utilized so that system can decide when the room or house lights will be turned on/off. Measuring the air temperature and its intensity helps to make auto switched on/off of air conditioners and fans or making decisions regarding air control in the smart home. For smoke detection, smoke detector sensor is installed to make timely alarm warnings to escape from fire disaster. Various other sensors can be equipped or considered according to the needs or requirements of the system or the home residents.

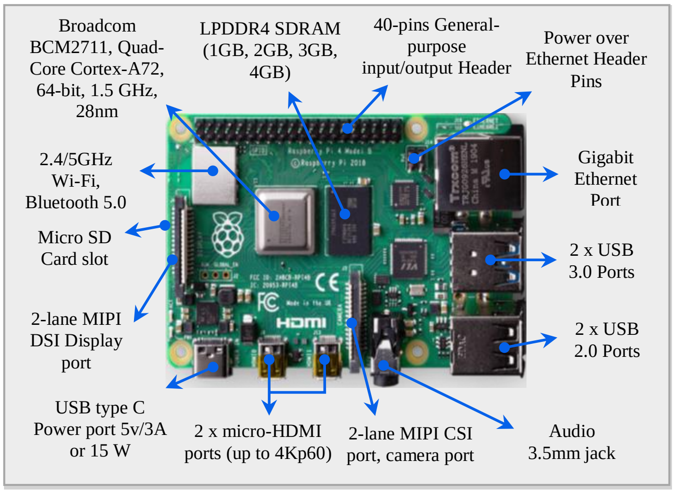


Fig. 3. Raspberry Pi 4 Model B Overview.

Raspberry pi plays the role of main controlling unit where all the sensors pass their measurements to it for taking further actions. Raspberry pi has many features like Wi-Fi connectivity, Bluetooth facility, and it can be connected to a display panel. In case of extra memory requirements, external memory storage can be used [33]. For utilizing the Wi-Fi or wired internet connection facility, it can be connected to the internet where we can easily control its operations through the android application. In the case of unavailability of internet, Raspberry pi can be connected via Bluetooth. To make it functional we have written a code for it, so that it could perform various tasks for our proposed system. Raspberry pi receives instructions as inputs from different sensors or from residents then it sends command(s) to home appliances via power relays. Various components and specifications of Raspberry Pi 4 model B are shown in Fig. 3.

Fig. 4 (block diagram) illustrates the basic components for the proposed system, where raspberry pi is connected to smart phone, to a display screen, and to an internet connection. It is coupled with different sensors and power relays and these power relays perform the operations of switched on/off the home appliances according to the given instructions of Raspberry pi. In the diagram, it shows how the proposed HA system components interact with each other.

In the proposed system we have considered eight power relays, but these can be extended according to the needs or number of appliances. The power relays are used to switch on/off the appliances according to the instructions from Raspberry pi. As mentioned, the Raspberry pi is connected to the internet via wired or Wi-Fi connection, so Android-based smart phone or smartwatch can be used to control the operations of Raspberry pi. It can receive the instructions as an input from the smartphone or smartwatch or can send the notifications to them for the HA system. Further, a display panel is attached to Raspberry pi where user can see the operations accomplished or by touching the digital buttons on the screen to control the home appliances. Fig. 5 illustrates example of equipment and appliances that can be setup for the proposed HA system.

The designed system can automatically open or close the doors or garage when detecting the human presence. The Raspberry pi can also receive the current weather forecasts from internet then it can automatically open/close the windows or curtains, or on/off the required appliances and home temperature can be regulated accordingly.

If the number of cameras is higher and they record a huge volume of data, then the produced data can be stored in the EC where later on features extraction or processing tasks can be done at this platform (i.e., EC). So, for the computation offloading, EC is added which makes our proposed system more robust and more reliable especially when the task is time stringent or having more data size for computation.

If the resident doesn't have mobile phone and wants to control something at home, then it can regulate the home appliances by using touch screen which as displayed in Fig. 6. But this is not the main motive of our designed system, although we can say it is a facility whenever in the case of unavailability of mobile phone.

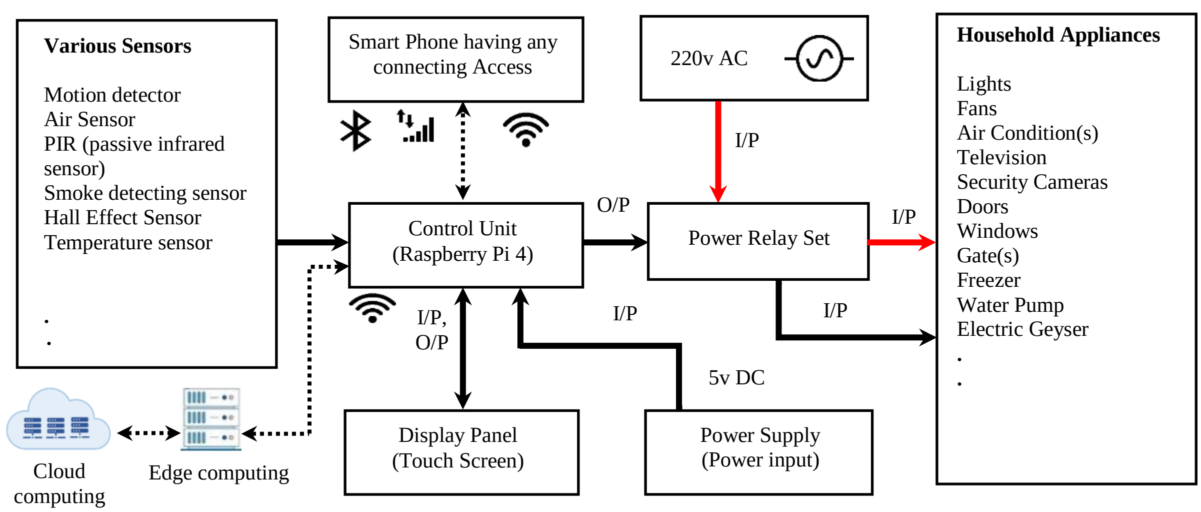


Fig. 4. Block Diagram for the Designed System.

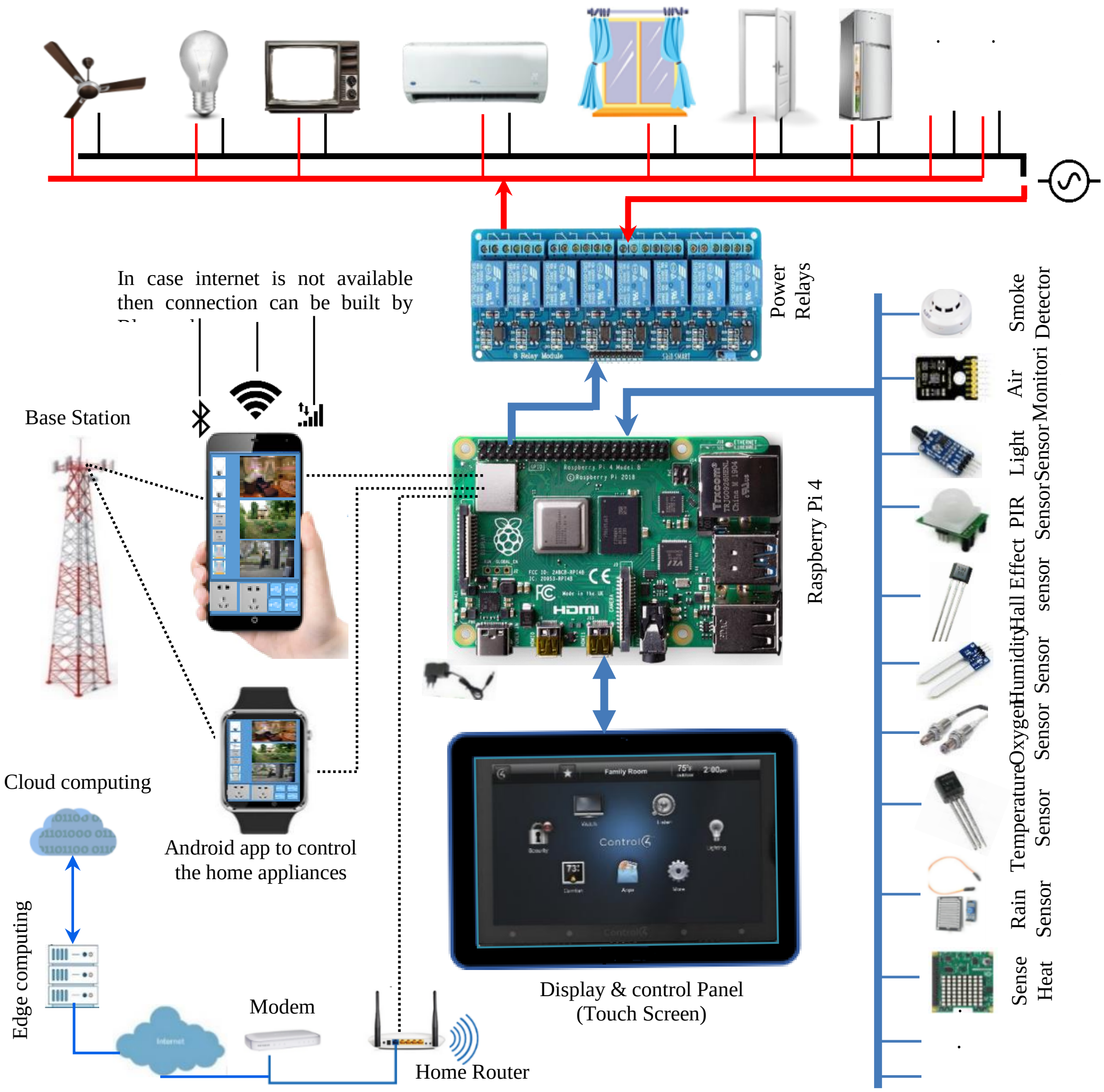


Fig. 5. Example of Equipment and Appliances Connected for the Proposed HA System.

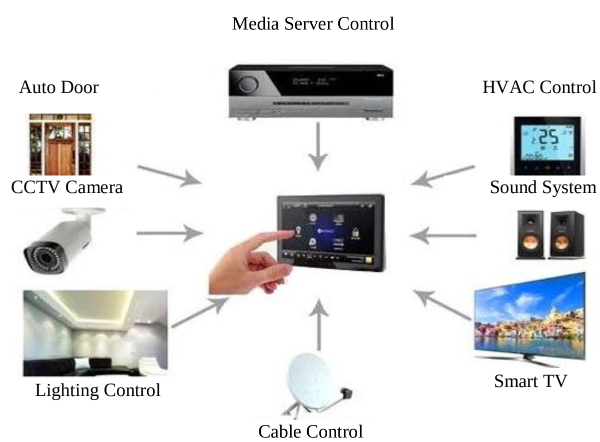


Fig. 6. Illustrates when the Mobile Phone is Unavailable, Touch Panel can be Utilized.

To use the designed system, the user needs to install the developed application. After checking the internet connectivity, user will use its user id and password for login to access the home appliances. Afterward login access, the system shows the available rooms, where a specific room will be selected for switching on/off the device(s).

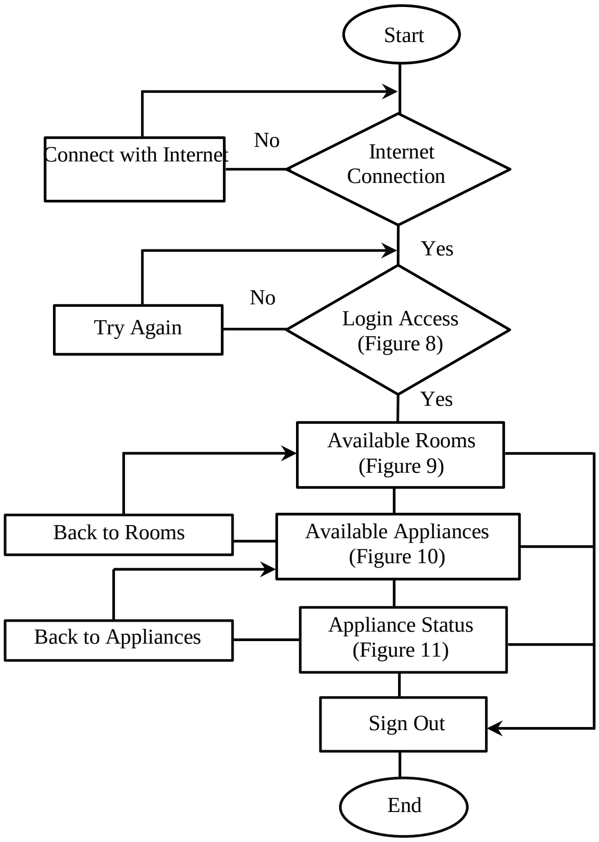


Fig. 7. Process Flow for the Developed System.

Algorithm 1: Smart HA System

1: **Input:** Instruction from Android App, 220-volt input to home appliances, inputs from different sensors to Raspberry Pi 4
2: **if** Internet connection available **then**
3: **if** Login Access granted (Figure 8) **then**
4: **show** Available Rooms (Figure 9) || Sign out
5: **if** select Available Room (Figure 9), **then**
6: **show** Available Appliances (Figure 10) || Sign out
7: **if** select Available Appliances (Figure 10), **then**
8: **show** Appliance Status (Figure 11) || Back to Room || Sign out
9: **if** select Back to Rooms, t**hen**
10: goto line number 4
11: **if** select Appliance Status (Figure 11) **then**
12: Modify the applicane status (i.e., ON/OFF etc.) || Back to Appliances || Sign out
13: **if** select Back to Appliances, t**hen**
14: goto line number 6
15: **end if**
16: **else** Try Again
17: **end if**
18: **else** Connect with Internet
19: **end if**
20: **Output:** Home appliances are operated automatically according to the residents' desires

Next, the screen will display the information of available appliances in that room. By clicking on the desired appliance user can see its status like either it is on or off. By choosing the options on the display, resident can on/off the household devices or manage the desired operation. Fig. 7 (flowchart) and the Algorithm 1 present the whole mechanism and working behavior of the proposed system.

A. Android Application

After installing the HA system application, the user can login by using theuser ID and password or can create an account as shown in Fig. 8. After successful login, Fig. 9 shall appear as the main interface screen. The information about room environment and weather conditions like temperature will be available on this screen. Further this main screen has a list of all available rooms. User will select the desired room for controlling the appliances, for example if the user selects bedroom 2, then the next screen is opened as in Fig. 10 showing the list of all available appliances for the selected room. Here, the user can control or navigate the desired appliance for the selected room and also the user can see the current status of a particular appliance.

User can select any appliance from the list. For instance, if the desired appliance is air conditioner, then user can switch on/off this appliance like in Fig. 11 (screen), and can change its functions/intensity like temperature or fans speed, etc. An additional feature of this application is that user can set the timings for different appliances like when the lights to be turned on/off. The system is capable for reminding the owner by sending the notification(s) about the current weather and home environment conditions. It can send the alert(s) when motion sensors detect suspicious activities or when someone trying to access application as unauthorized user.

B. Challenges during System Designing and Application Development

Installation of different types of sensors in home appliances particularly when these are designed by different vendors, can be a major challenge in developing a HA system [34, 35]. Hence, there must be an appropriate mechanism for data controlling and processing capabilities to manage the entire system effectively. Although availability of reliable internet connectivity plays a key role in many cases, at certain locations they may not be feasible. Further, professionals are needed to develop the system by considering the various home situations and the dependability of sensors on the home devices. Another issue is that the confidential information of a person in house can be leaked during the data transmission, so proper data security mechanisms must be well defined for home data. To achieve the goal of "HA or connected home", several professions like electrician, builders, locksmith, gasman, etc., may be involved for their skills, expertise, and businesses according to market demands. Often the mentioned occupations may not have entire knowledge about the HA and there must be some predefined standards when they are installing the home devices especially when these devices are from different vendors.

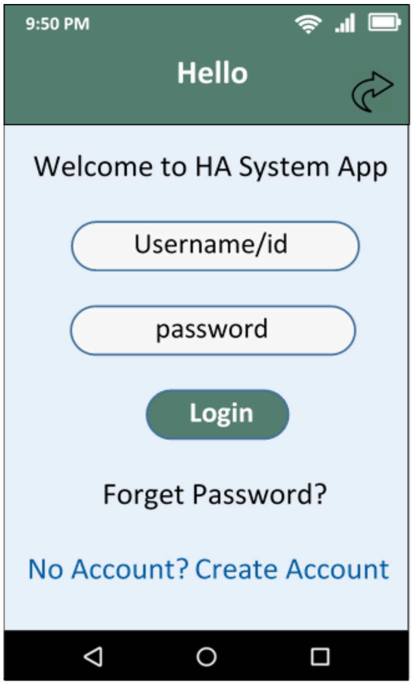


Fig. 8. Sample Screenshot of the Developed Application for Login Access.

Fig. 9. Sample Screenshot of the Developed Application for Selecting the Desired Room.

It is usual that smart home automation systems come with regular upgrades or patches, so customized configurations or alterations must be supported in such system.

It is normal that appliances or devices are of different types and manufactured by multiple manufactures. So, there should be one application that can deal with all of them. As such, developing a generic application for HA system not only comes with a variety of technological problems but security is also a major challenge that must be considered. Risk factors are often involved when sending or exchanging a home's confidential information on the internet. It is important to examine all the technicalities of home appliances or devices and security matters when sharing the private information like presence of location to someone whom are unauthorized. Our developed application effectively deals with the above-mentioned situations.

Fig. 10. Sample Screenshot of the Developed Application for Selecting the Desired Appliances.

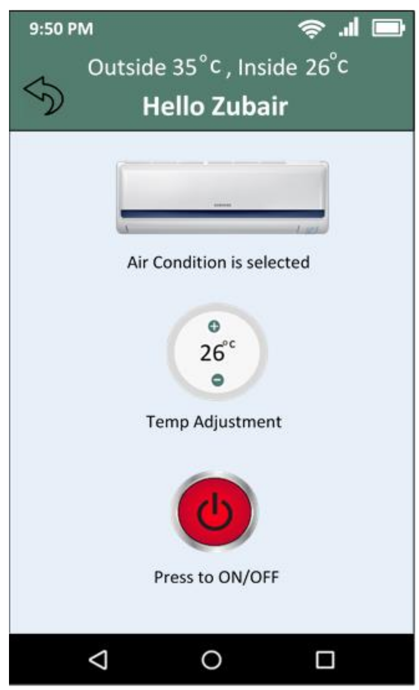


Fig. 11. Sample Screenshot of the Developed Application for Adjusting the Functions of Desired Appliances.

IV. Result and Discussion

Our design has considered an average size of house consisting of two bedrooms, one guest room, one TV lounge, one dining hall, one kitchen, and one car garage. It has appliances including three air conditioners, 3 room heaters, one fridge, one water heater, seven fans, one microwave oven, one washing machine, and 12 lights. We assumed that it is in Multan city, Pakistan where the weather conditions are normally around 5°C to 20°C in winter and 30°C to 45°C during the summer. Electricity bills are higher in summer comparatively to the winter due to more air conditioners usage.

A. Electricity Consumption of Existing System against the Designed HA System

For electricity consumption analysis ‘current and voltage clock meter’ are utilized, to capture key energy consumption information. Data from these meters, allow us to calculate the electricity usage in the aspects of cooling, heating, and lighting of the house. Three methods were used to measure the electricity utilization i.e., day-wised, monthly, and annually. In the first case, power consumption is calculated for a single day by comparing the existing electricity system with the designed HA system and this is demonstrated in Fig. 12. By this method, owner(s) know the home power consumptions at different times of a day. So, with this data, resident(s) can plan at which time to save the electricity or the timings when the wastage of electricity can be best avoided.

The electricity consumption is measured in per Kilowatt hour (kWh). That is, one unit of electricity or kWh can be calculated as follows:

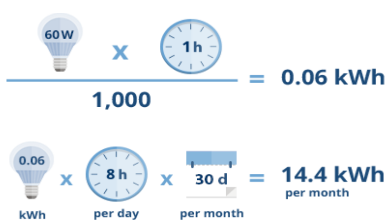
$$\text{Energy consumed} = \text{Power} \times \text{Time}$$

$$\text{Electricity unit (kWh)} = \frac{\text{Power (in watts)} \times \text{Time (hours)}}{1000}$$

A lightbulb of 60 watts used for 8 hours per day (240 hours per month) will consume electricity in one month as follows:

(watts/1000 = .06 kilowatts)

06 kW * 8(hours) * 30(days) = 14.4 kWh



In the second case, these calculations were perfomed on monthly-wise. The months in which more cooling are required are May to August and are considered as the peak of hot weather in Pakistan. To attain the comfort level of 20°C-27°C (68-80°F), a normal size home consumes average of 700 to 900 kWh per month of electricity in these months. Whereas to maintain the comfort level for the winter months which required more heating, i.e., December to February, it required nromally 500 to 600 kWh. In considering the same house environemnt but with the involvement of the designed HA system, the power consumption was around 100 to 130 kWh per month lesser than the existing system. This comparison of the electricity consumption shows that the proposed HA system significantly lowered the power consumption. The developed HA system, helps to save the electricity by 15%-20% per month. Fig. 13 presents the comparison, and clearly showing that the developed system consumed less electricity than the existing system.

To attain the comfort level of 20°C - 27°C (68-80°F), one house can consume around 7180 kWh per year when it is not using the designed HA system, but when it is integrated with the desigend HA system then the consumption drops to around 6100 kWh annually. Comparison of these power consumptions annualy are depicted in Fig. 14.

The proposed HA system saves on average of 1080 kWh of electricity per year for one house. It is beneficial to mentioned that this amount can be greater when considering to integrate this system in larger number of houses. Conservatively, for a housing society having 8000 houses, the saved electricity can be reaching to 8,640,000 kWh per year. This amount is saved within one year.

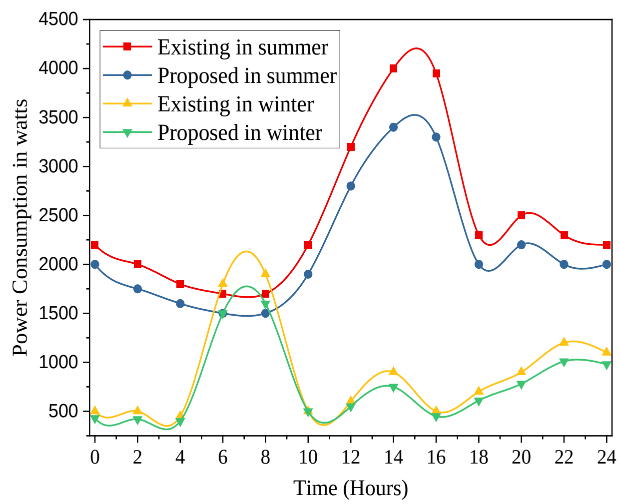


Fig. 12. Hourly Energy Consumption Graph for One Day, Comparing the Existing System to the Designed HA System.

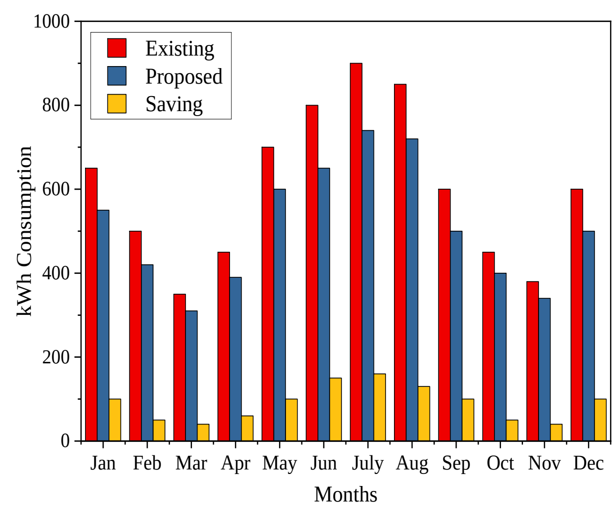


Fig. 13. Comparison of Monthly kWh Consumptions between Existing System and the Developed HA System.

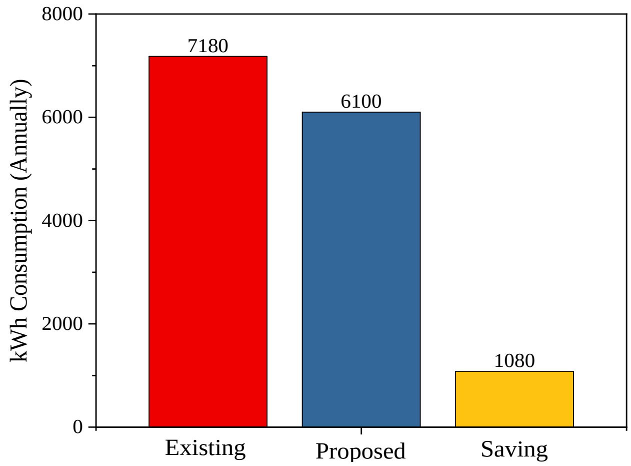


Fig. 14. Yearly Comparison of kWh Consumptions.

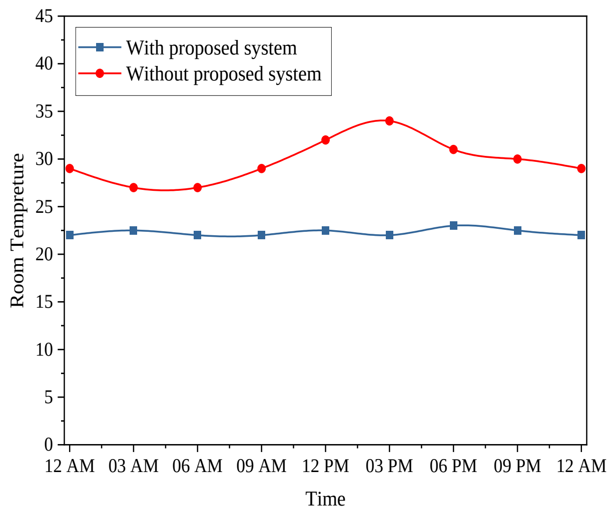


Fig. 15. Room Temperature in Summer Season with and without the Involvement of the Designed HA System.

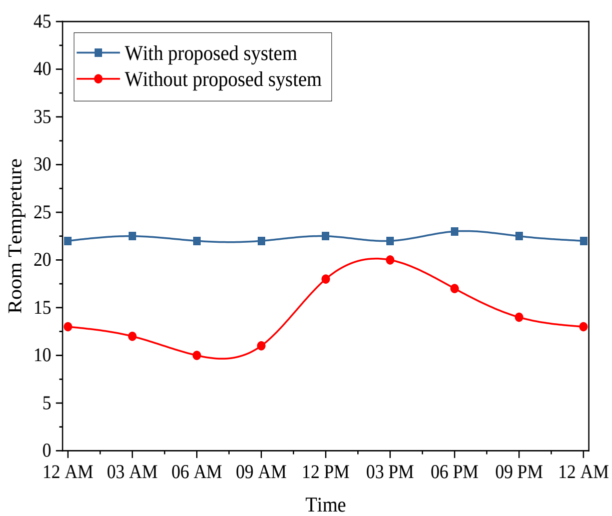


Fig. 16. Room Temperature in Winter Season with and without the Involvement of the Developed HA System.

The comparison of the room temperature in summer and winter seasons, with and without the involvement of the developed system is illustrated in Fig. 15 and Fig. 16, respectively. The temperature was measured with respect to hourly time for a single day. It can be seen that the designed HA system is able to maintain the desired room temperature without major fluctuations in both seasons (summer and winter). The temperature data was taken by using temperature sensor. With the available data, the residents of a home can analyze the home environment from anywhere and can take the necessary actions to save electricity. It is clear that without the involvement of the proposed system, there are large variations in room temperature. So, the concept of our smart HA is greatly beneficial to consumers.

B. *What have been Achieved?*

- Smart use and control of home devices or appliances by the proposed HA system.
- Designed HA system saves energy because appliances operate only when they are required, while automatically switched off when they are not needed, thus reduces the wastage of electricity (around 18%).
- No need to replace the current appliances because changes are made and required only for the control unit.
- Connectivity with internet makes HA system more robust and provide intelligent work flow according to the prediction on home appliances.
- HA system provides home alarms and warnings features to prevent any loss of home appliances/properties.
- Generate timely notification for voltage fluctuations that helps to save home appliances from potential faults or problems.
- The system is good enough and capable to control a large size of house with the variety of electronics and electrical devices easily.
- EC is integrated in the proposed system to improve its reliability and making it more robust for the computation offloading.
- The motion monitoring sensors can be greatly useful in sustaining the home security by sending timely alert to the resident or house owner.
- It is an Android app that can be easily installed and used by users with Android smart phone.

The developed HA system can control how and when a home device will respond and act to attain desired home environment. It offers a fully controlled and convenient environment and helps to reduce the energy expenditures. It can assist the users and residents to control any mishap and unpredicted incidents that may happen at home.

V. Conclusion

In this paper, an energy efficient and flexible HA system is presented for controlling and monitoring the home appliances using the smart phone-based Android application. Edge computing is used in the system to enhance reliability and robustness in computation tasks offloading. The system uses IoT-based networking functions to remotely control and monitor the home appliances. The proposed application allows home user/resident to conveniently control the commonly used appliances in a home. The application and the overall hardware architecture for controlling are flexible for easy modification when new appliances or new features in application to be added or remove to support new home requirements. The system operates automatically according to the home environmental conditions set by the home occupants' requirements. The developed HA system enhances the automation, security, remotely controlled, user friendly, more home comfort by saving around 15% to 20 % electricity through the properly configured HA system.

In future work, we are planning to enhance the system's features and its coverage area, as well as adding more functions and options to extend its robustness. The enhanced system can cater for garden and lawn. Adding in extra intelligence to the future system to capture the behavior of the home residents, thus setting the home environment more naturally.

Acknowledgment

This research was conducted in Universiti Teknologi PETRONAS (UTP) under the YUTP grant scheme with the reference code of #RG2022-0754, the cost center of 015LC0-413, and the project title of "Edge Computing Oriented IoT Operation and IoT Resources Optimization".

References

[1] S. Shome and R. Bera, "Narrowband-IoT base station development for green communication," in Advances in greener energy technologies: Springer, 2020, pp. 475-487.

[2] Y. Deng, Z. Chen, X. Yao, S. Hassan, and A. M. Ibrahim, "Parallel offloading in green and sustainable mobile edge computing for delay-constrained iot system," IEEE Transactions on Vehicular Technology, vol. 68, no. 12, pp. 12202-12214, 2019.

[3] Z. Sharif, L. T. Jung, I. Razzak, and M. J. I. I. o. T. J. Alazab, "Adaptive and Priority-based Resource Allocation for Efficient Resources Utilization in Mobile Edge Computing," 2021.

[4] A. Bujari, M. Furini, F. Mandreoli, R. Martoglia, M. Montangero, and D. Ronzani, "Standards, security and business models: key challenges for the IoT scenario," Mobile Networks and Applications, vol. 23, no. 1, pp. 147-154, 2018.

[5] D. C. Khedekar, A. C. Truco, D. A. Oteyza, and G. F. Huertas, "Home Automation−A Fast - Expanding Market," Thunderbird International Business Review, vol. 59, no. 1, pp. 79-91, 2017.

[6] T. Chakraborty and S. K. Datta, "Home automation using edge computing and internet of things," in 2017 IEEE International Symposium on Consumer Electronics (ISCE), 2017, pp. 47-49: IEEE.

[7] R. Sokullu, M. A. Akkaş, and E. Demir, "IoT supported smart home for the elderly," Internet of Things, vol. 11, p. 100239, 2020.

[8] J. Ren, H. Wang, T. Hou, S. Zheng, and C. Tang, "Federated learning-based computation offloading optimization in edge computing-supported internet of things," IEEE Access, vol. 7, pp. 69194-69201, 2019.

[9] Z. Sharif, L. T. Jung, and M. Ayaz, "Priority-based Resource Allocation Scheme for Mobile Edge Computing," in 2022 2nd International Conference on Computing and Information Technology (ICCIT), 2022, pp. 138-143: IEEE.

[10] V. C. Leung, X. Wang, F. R. Yu, D. Niyato, T. Taleb, and S. Pack, "Guest Editorial: Special Issue on Blockchain and Edge Computing Techniques for Emerging IoT Applications," IEEE Internet of Things Journal, vol. 8, no. 4, pp. 2082-2086, 2021.

[11] X. Guo, Z. Shen, Y. Zhang, and T. Wu, "Review on the application of artificial intelligence in smart homes," Smart Cities, vol. 2, no. 3, pp. 402-420, 2019.

[12] S. K. Vishwakarma, P. Upadhyaya, B. Kumari, and A. K. Mishra, "Smart energy efficient home automation system using iot," in 2019 4th international conference on internet of things: Smart innovation and usages (IoT-SIU), 2019, pp. 1-4: IEEE.

[13] Z. Sharif, "Internet-of-Things based Home Automation System using Smart Phone," Sir Syed University Research Journal of Engineering Technology, no. 2, 2021.

[14] D. Mocrii, Y. Chen, and P. Musilek, "IoT-based smart homes: A review of system architecture, software, communications, privacy and security," Internet of Things, vol. 1, pp. 81-98, 2018.

[15] A. A. Zaidan et al., "A survey on communication components for IoT-based technologies in smart homes," Telecommunication Systems, vol. 69, no. 1, pp. 1-25, 2018.

[16] H. Ghayvat, S. Mukhopadhyay, B. Shenjie, A. Chouhan, and W. Chen, "Smart home based ambient assisted living: Recognition of anomaly in the activity of daily living for an elderly living alone," in 2018 IEEE international instrumentation and measurement technology conference (I2MTC), 2018, pp. 1-5: IEEE.

[17] N. Ahmad, Z. Sharif, S. Bukhari, and O. J. I. J. o. A. R. i. B. Aziz, "Insights Into Functional and Structural Impacts of nsSNPs in XPA-DNA Repairing Gene," vol. 12, no. 1, pp. 1-12, 2022.

[18] B. Hafidh, H. Al Osman, J. S. Arteaga-Falconi, H. Dong, and A. El Saddik, "SITE: the simple Internet of Things enabler for smart homes," IEEE Access, vol. 5, pp. 2034-2049, 2017.

[19] M. Uddin et al., "Cloud-connected flying edge computing for smart agriculture," vol. 14, no. 6, pp. 3405-3415, 2021.

[20] G. A. A. de Oliveira, R. W. de Bettio, and A. P. Freire, "Accessibility of the smart home for users with visual disabilities: an evaluation of open source mobile applications for home automation," in Proceedings of the 15th Brazilian Symposium on Human Factors in Computing Systems, 2016, pp. 1-10.

[21] O. Hamdan, H. Shanableh, I. Zaki, A. Al-Ali, and T. Shanableh, "IoT-based interactive dual mode smart home automation," in 2019 IEEE international conference on consumer electronics (ICCE), 2019, pp. 1-2: IEEE.

[22] S. M. Patel and S. J. Pasha, "Home automation system (HAS) using android for mobile phone," International Journal Of Scientific Engeneering and Technology Research,, ISSN, pp. 2319-8885, 2015.

[23] V. Lohan and R. P. Singh, "Home Automation using Internet of Things," in Advances in Data and Information Sciences: Springer, 2019, pp. 293-301.

[24] L. M. Satapathy, S. K. Bastia, and N. Mohanty, "Arduino based home automation using Internet of things (IoT)," International Journal of Pure and Applied Mathematics, vol. 118, no. 17, pp. 769-778, 2018.

[25] R. Piyare, "Internet of things: ubiquitous home control and monitoring system using android based smart phone," International journal of Internet of Things, vol. 2, no. 1, pp. 5-11, 2013.

[26] T. S. Gunawan, I. Rahmithul, H. Yaldi, M. Kartiwi, and N. Ismail, "Prototype design of smart home system using internet of things," Indonesian Journal of Electrical Engineering and Computer Science, vol. 7, no. 1, pp. 107-115, 2017.

[27] S. Brundha, P. Lakshmi, and S. Santhanalakshmi, "Home automation in client-server approach with user notification along with efficient security alerting system," in 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon), 2017, pp. 596-601: IEEE.

[28] N. Vikram, K. Harish, M. Nihaal, R. Umesh, and S. A. A. Kumar, "A low cost home automation system using Wi-Fi based wireless sensor network incorporating Internet of Things (IoT)," in 2017 IEEE 7th International Advance Computing Conference (IACC), 2017, pp. 174-178: IEEE.

[29] B. R. Stojkoska and K. Trivodaliev, "Enabling internet of things for smart homes through fog computing," in 2017 25th Telecommunication Forum (TELFOR), 2017, pp. 1-4: IEEE.

[30] S. Zhou and L. Zhang, "Smart home electricity demand forecasting system based on edge computing," in 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), 2018, pp. 164-167: IEEE.

[31] F.-J. Ferrández-Pastor, H. Mora, A. Jimeno-Morenilla, and B. Volckaert, "Deployment of IoT edge and fog computing technologies to develop smart building services," Sustainability, vol. 10, no. 11, p. 3832, 2018.

[32] W. Yu et al., "A survey on the edge computing for the Internet of Things," IEEE access, vol. 6, pp. 6900-6919, 2017.

[33] P.-J. Chen, T.-H. Hu, and M.-S. Wang, "Raspberry Pi-Based Sleep Posture Recognition System Using AIoT Technique," in Healthcare, 2022, vol. 10, no. 3, p. 513: MDPI.

[34] A. Hong, C. Nam, and S. Kim, "What will be the possible barriers to consumers' adoption of smart home services?," Telecommunications Policy, vol. 44, no. 2, p. 101867, 2020.

[35] M. Ayaz, M. Ammad-Uddin, Z. Sharif, A. Mansour, and E.-H. M. Aggoune, "Internet-of-Things (IoT)-based smart agriculture: Toward making the fields talk," IEEE Access, vol. 7, pp. 129551-129583, 2019.

An Integrated Approach to Research Paper and Expertise Recommendation in Academic Research

Charles N. Mabude[1], Iyabo O. Awoyelu[2], Bodunde O. Akinyemi[3], Ganiyu A. Aderounmu[4]
Department of Information Systems and Technology, Kings University, Nigeria[1]
Department of Computer Science and Engineering, Obafemi Awolowo University, Nigeria[2, 3, 4]

***Abstract*—Research papers and expertise recommendation systems have been operating independently of one another, resulting in avoidable search queries and unsatisfactory recommendations. This study developed an integrated recommendation system to harmonize research papers and expertise recommendations in the academic domain for enhanced research collaboration. It aims to address issues related to the isolated existence of these systems. A recommendation algorithm was developed to synergize the research paper and expertise recommendation models. The Cosine similarity function between user query and available research papers as well as experts, was combined with selected criteria to achieve recommendation. The synergized model was simulated and evaluated using Precision, Recall, F-measure and Root Mean Square Error as performance metrics. The findings showed that the harmonization of research paper and expertise recommendation approaches provides a holistic and enhanced approach towards research paper and expertise recommendation. Thus, academic researchers now have a reliable way to recommend experts and research papers, which will lead to more collaborative research activities.**

Keywords—Research paper; expertise; recommendation systems; academic

I. Introduction

The ever-increasing growth of data has resulted in the discovery of new research areas within the computer science field, among which is recommender systems. As the amount of available data grows, the problem of managing the information becomes more difficult, which culminates in information overload. The need to address the challenges associated with information overload led to research on information filtering. The Recommender system, as a subclass of the Information Filtering System, is a completely automated system that analyzes users' preferences and predicts users' behaviors. The research interest in recommender systems is still very high, perhaps due to the practical significance of recommendation tasks. Recommender systems collect various kinds of data to create their recommendations [1].

An Expertise Recommender (ER) system offers a means by which scarce resources, in the form of human experts, can be identified and accessed [2]. It is a system designed to facilitate identifying individuals who have the necessary expertise to solve a specific problem [3] [4]. In systems for recommendation of human expertise, the interest is "expert" instead of items [5] [6] like movies or books, as in the case of other recommender systems. Mobile agent technology is one of the approaches that have been implemented for expertise recommendation [7] [8] [9]. Available literature has shown recommendation systems for human expertise to be a valuable instrument incredibly, crucial in scholarly research for managing knowledge.

Research Paper Recommender (RPR) systems have emerged to ease the problem of finding publications relating to researchers' areas of interest [10] and are designed to offer the right publication to the right researcher in the right way. Thus, some users may be interested in the works (research papers or publications) of the expert (researcher) rather than the expert. This is reflected in the search for a research paper (i.e., documented knowledge of the expert) without necessarily being interested in the researcher (expert).

These two aspects of recommender systems are utilized within the academic research domain. The research papers and expertise recommender systems have greatly contributed to improving research collaborations in academia by bridging the knowledge and physical distance among researchers. However, it is mostly the case that a researcher in a particular domain must have either directly or indirectly interacted with one or more researchers in the research domain or certainly perused many research publications in order to elicit knowledge about a particular research project. These publications might be from the researcher (s) that has been directly interacted with or from different researchers.

The success of the recommender system in academics as a tool for collaboration in research is evident in the literature. In recent years, a detached approach to the issues of a research paper and expertise recommendation has existed. However, the academic research concept requires that a researcher be exposed to both direct and indirect contributions of fellow researchers for acceptable research output. The existing expertise recommender systems have not considered using explanations for more details about the researchers' expertise, while the research paper counterpart considered concept similarity of articles without the methods, which resulted in a misleading and unsatisfactory recommendation. Furthermore, both systems experience sparse ratings and function independently, thus, providing partial information about the same academic domain, thereby causing avoidable search queries. There is a need for a holistic approach rather than the current separate approach to expertise and research paper recommendations in academics, and also a system that will harmonize the concept or method similarity of research papers and explanation-based expertise recommendations for an enhanced recommendation.

Thus, in this paper, an attempt is made to concurrently address expertise recommendation and research paper recommendation in a unified recommender system using an integrated recommendation concept. The remaining sections of the paper are organized as follows: Section II discusses the related works while the methodologies applied to achieve the desired results are presented in Section III. The results obtained as well as the interpretations are presented in Section IV. The conclusion of the study is contained in Section V and future work is presented in Section VI.

II. Related Work

There are numerous works in the areas of expertise and research paper recommendations. A survey of available literature on research expertise recommendations shows that academic research expertise systems are less common. However, the increase in available data on academics and their publications has sustained research in the area.

In several researches, a recommender system was developed based on expert and item category by extending recommendations beyond the frequency of rating from users [11] to using trust-aware social web service recommendations mechanisms [12], a triangulated approach[2], bi-directional feedback approach [13] [14], ontology based approach [15], social network-based approach [16], [17], [18], Clustering analysis approach, expertise modeling [19], Knowledge-based approach [20], graph-based approach [21] Citation network graph approach [22] [23] [24], random walk graph approach [25], Content Based Filtering and Collaborative Filtering concepts [26], and association rule approach [27] with the aim of improving the accuracy of the expert recommendation predictions. The factors of topic relevance, expert quality, and researcher connectivity are considered useful in recommending experts in the science-based research community. For example, researcher modeling approach was applied to recommend experts in scientific communities in [28], a co-author recommendation based on powerful and similar peers that suggests future co-authors for scientific article writing was also proposed in [29]. However, these approaches did not address rating sparsity and cold start problem efficiently.

On the other hand, research paper recommendations are another area that is well explored. Nowadays, much of the world’s new knowledge is largely captured in digital form and archived within a digital library system. Also, a huge number of academic papers are constantly being published through several conferences and journals. These trends lead to information overload, where users find an overwhelming number of publications that match their search queries but are largely irrelevant to their latent information needs [30]. Thus, most researchers rely on key-based search or browsing through proceedings of top conferences and journals to find their articles of interest [31]. However, according to [10], the challenge is not just to provide researchers with very rich publications at any time, any place, and in any form, but to also offer the right publication to the right researcher in the right way. To achieve this crucial objective, several researchers have initiated some novel approaches depending on their perspectives. Some of the recommendation approaches in the literatures are set theory and graph theory concepts [32] [33] [34], user’s preferences, user's requirements and opinion mining approach [30], [35], [36], [37], content-based approach [38], [39], collaborative filtering approach[10] [31] [40], source-independent framework [41], tagging approach [42] and algorithms like an unsupervised algorithm called Paragraph Vector [43], shortest paths algorithm [44], that provide personalized research papers recommendations regardless of the research field and regardless of the user's expertise.

It was observed in all of these approaches that an atomistic approach was being used to address issues related to the research paper and expertise recommendation in the same academic domain, whereas the academic system requires a researcher to be exposed to both direct and indirect contributions of fellow researchers in order to produce acceptable research output. The need to alleviate the challenges of the current isolated approach to research paper and expertise recommendation necessitated this study.

III. Methodology

In this study, the focus was to improve the existing expertise and research paper recommendation models and, as well, synergize them for better recommendations in academics. The details provided in the expertise recommendation part of the integrated system will assist expertise seekers by providing more details about the available experts to guide their choice among recommended experts. Also, in addition to the existing concept semantics-based research paper recommendation, the study will reduce the chances of recommending the wrong research paper to users by combining the concept semantics with the method semantics of research publications to build similarity between research papers for improved recommendation.

In this study, the following research questions were addressed:

- How could the expertise recommendation be enhanced by incorporating explanation to provide more details about the experts’ knowledge?
- How could the research paper recommendation be enhanced by combining semantic relationships between concepts and methods used in a research paper?
- How could recommendations be enhanced by synergizing (i) and (ii) above, since both deal with recommendations for the same academic domain?

A. Conceptual Model

The developed model was conceptualized on the basis of building a synergy between two independent but related recommender systems. The improvements are presented as follows.

1) Proposed expertise recommendation model: The existing ER with user classification [13] is being improved by incorporating explainable recommendations to provide further details about the depth of knowledge and research interests of the experts for more effective and persuasive recommendations. The conceptual view of the proposed expertise recommendation approach, presented in Fig. 1, demonstrates the search query as the key input to the system. The search query gets processed through matching with the available experts in the database based on relevance. This is carried out in the "Query/Expertise Matching" section by searching through the database of experts. However, in the database of experts, the experts are classified according to their academic qualifications, which help to differentiate their level of knowledge. An explanation which contains a summary of available information about the knowledge depth and research interests of each expert is generated for each expert using the recommendation logic while the experts are clustered using the tag-based K-Means clustering algorithm. The "Recommendation" segment then presents a list of experts with explanations arising from the calculations in the "Recommendation Logic" segment through computation of the likeness cosine linking the user query with the available data in the target database.

2) Proposed research paper recommendation model: The existing concept-based research paper recommendation system [37] is also being improved by incorporating similarity between methods applied by researchers to solve problems as published in journals. This is because it considers that each published research article consists of two major parts – the concept and the method (s) employed to actualize the concept. Thus, the deficiency of the existing system is addressed by incorporating the similarity in the methods used to solve the problem identified in the research publications into the existing concept-based approach as shown in Fig. 2. The figure shows the search query section where the user inputs keywords relating to the desired research paper. The user's query is then processed by matching it with the available research papers in the database. The recommendation logic is then applied to generate a list of suitable research papers using the tag-based K-Means clustering algorithm. The user then views the recommendation list.

3) The developed integrated recommendation framework: It was observed that though the systems for research papers and expertise recommendation currently exist independently, the dataset utilized in either case is about the same domain, i.e., the academic domain. For instance, expertise recommender systems make use of datasets of published research papers to build the expertise domain of the experts by extracting relevant texts from the publications. Therefore, the two recommended systems are synergized as presented in Fig. 3, which shows that the dataset is pre-processed and loaded into the database. The user profile is created during the first access to the system. Subsequent access requires only the login information of the user, which is entered through "User Info". The "User Identifier" authenticates the user using available user profile information in the database. Once the user is authenticated, the user can then search for the desired information through the "User Query" section. The user query is then processed by applying the cosine similarity function to determine the available experts and research papers in the desired area as specified in the user query. If the user query is available, then the "Recommendation Logic" segment is activated to perform tag-based K-Means clustering of experts and research papers, matrix factorization, and generation of details for the recommendation. The result of this segment is then pushed to the "Recommendation" segment where the list of recommended experts with details and the list of recommended research papers are displayed.

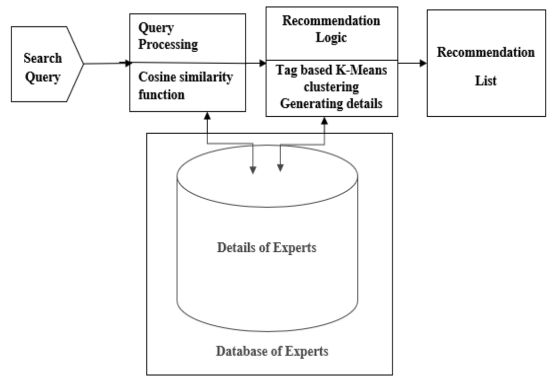


Fig. 1. Improved Expertise Recommendation System.

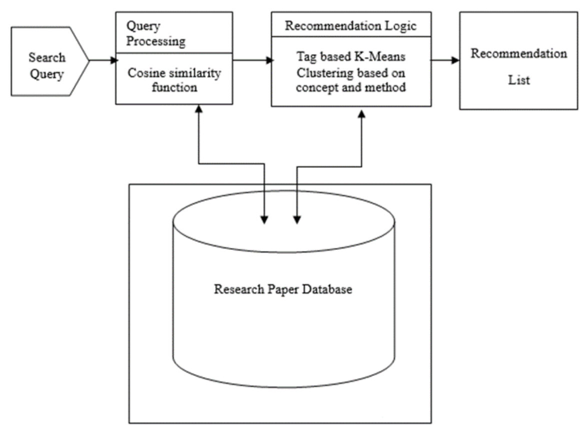


Fig. 2. Improved Research Paper Recommendation System.

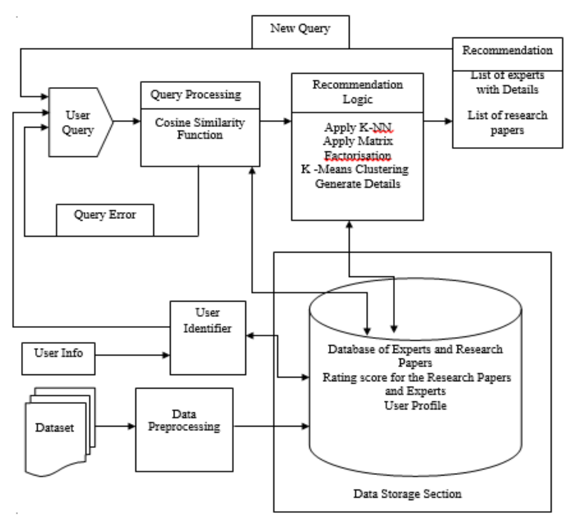


Fig. 3. Proposed Integrated Expert and Research Paper Recommender System.

B. Mathematical Formulation of the Developed Model

The developed model was formulated on the basis of building a synergy between two related sets to achieve a common goal. The related sets considered in this study are a set of experts, i.e., researchers, and a set of research papers, as presented in Equation 1 and Equation 2, respectively. The reviewed literature shows that there are certain features and criteria for expertise and research paper recommendation. A few of the attributes include quantity of publications for target expert, number of citations for target research paper, and user rating, while the criteria for recommendation constitute a combination of selected features.

$$E = \{e_1, e_2, e_3, \dots, e_n\} \tag{1}$$

$$R = \{r_1, r_2, r_3, \dots, r_n\} \tag{2}$$

where

E represents the set of experts and *R* represents the set of n distinct research papers $r_1\, to\, r_n$ while $e_1\, to\, e_n$ represent distinct experts.

Since the dataset of published articles is also utilized for research paper recommendation, in which case, keywords or key phrases from the publications are extracted for the purpose of recommendation, this study therefore considers the possibility of bridging the two independent but related recommendation systems and approaches in order to synergize both systems for enhanced recommendation. To achieve the synergy, the information theory of synergy and the set theory are applied. The set of experts and the set of research papers are then combined into a unified set as presented in Equation 3 and Equation 4, respectively.

$$D = E \otimes R \tag{3}$$

This implies that:

$$D = \{(x, y) \mid for\ all\ x \in E, y \in R\} \tag{4}$$

Therefore:

$$D = \{(e_1, r_1), (e_2, r_2), \dots, (e_n, r_n)\} \tag{5}$$

$E,\ R, e_1\ to\ e_n$ and $r_1\ to\ r_n$, retain their previous meanings while *D* represents the unified set.

Applying the information theory of synergy, the set of experts, "E", can be viewed as variable X_1 while the set of research papers "R" is variable X_2. These two related sets are then combined to form the unified set "D" of research papers and experts, which is represented by the third variable "Y".

Furthermore, the recommendation for the experts is achieved by calculating the impact of the expert in the chosen research area using Equation 6 while incorporating the rating function presented in Equation 7. The impact of an academic researcher in a particular domain obtained through research publications could be helpful for profiling the research expertise or experience of tertiary institutions and other academic institutes, either for the purpose of appointment, collaboration or consultancy.

$$E_{impact} = \sum_{i=1}^{n} \frac{1}{i}\left(\frac{1}{T_p}\left(N_{(cp)^I} + N_{(jp)^I} + N_{c^I}\right) + (A_{re})^I\right) \tag{6}$$

such that;

$$(A_{re})^I = \frac{R_e, i}{N_u, i} \tag{7}$$

where;

In Equation 6, the number of years of experience through research publication is represented by "i", which ranges from 1 to n, where the value of "n" is determined by the user. N_{cp} signifies the quantity of conference articles the expert has published, while N_{jp} signifies the quantity of journal articles the expert has published within the interval of years specified by the user. T_P is the total number of publications by the expert within the specified period, which is the addition of N_{cp} and N_{jp}. N_c represents the number of citations for the papers published by the expert, while T_c represents the total citations of all publications in the area of research publication. $(A_{re})^I$ represents the average rating for each expert over a given year interval. Also, in Equation 7, A_{re} retains its meaning in Equation 6, R_e is the total rating for the experts and N_u is number of users that rated the expert.

1) Matrix factorisation: The research paper recommender system and the expertise recommender system are usually associated with rating sparsity, which was addressed using matrix factorisation. In the matrix factorization approach, every item denoted by *i* is linked with a vector $x_i \in K^f$ and every user denoted by *u* is linked with a vector $y_u \in K^f$. The item *i* refers to expert or research paper while K^f represents the features. Given any item denoted by *i*, the components x_i determine the degree to which positive or negative factors are contained in the item. Given any user *u*, the elements of y_u determine the degree of the user's interest in items with high matching positive factors or negative, as the case may be. The resultant dot product, $x_i^T y_u$, reflects the interface between the user denoted by *u* and the item denoted by *i*— the general interest of the user in the features of the item. This sums up the rating for the item denoted by *i* provided by the user denoted by *u*, which is presented as $\hat{r}_{ui}$, resulting in an approximation.

$$\hat{r}_{ui} \cong x_I \cdot y_u \tag{8}$$

2) Model formulation algorithms: Algorithm 1 shows the procedure for developing an integrated research paper and expert recommendation. The algorithm for predicting user ratings using matrix factorization is presented in Algorithm 2, while the algorithm for the K-Means Clustering Algorithm is presented in Algorithm 3.

IV. Result and Discussion

The developed model was simulated and evaluated in Anaconda with Python version 3.7 environment. The Jupyter Notebook was deployed to provide the Graphical User Interface (GUI) for visualizing the data and the results of the simulation process, and Pandas for data manipulation and analysis. Furthermore, Surprise, which is a Python *scikit* for building and analyzing recommender systems that deal with explicit rating data, was also utilized to apply matrix factorization and cosine similarity functions to the dataset. The detailed results are presented as follows.

A. Analysis of the Dataset

The dataset, sourced from the Scopus database used in this study, contained two thousand (2000) documents while the papers with ratings of one (1) and above were one thousand, four hundred and fifty-one (1451) documents. The cross-validation method was used to validate the dataset using the train_test_split of the scikit-learn library. The data was analyzed and divided into two sets- Eighty percent (80%) training and Twenty percent (20%) testing. The number of records in the training dataset was one thousand, one hundred and sixty (1160) while the number of records in the test dataset was two hundred and ninety-one (291). The analysis of the actual dataset used in the study is presented in Fig. 4 and 5. Fig. 4 shows the distribution of ratings over the dataset, where the rating value of 4 (four) appeared most common among the rated publications, while many of the publications had a zero (0) rating. Fig. 5 shows the distribution of articles over the conference proceedings. It shows that articles are more than conference.

The articles are one thousand and two hundred (1,200) while the conference papers are Eight hundred (800).

B. Simulation Results

In order to ascertain the effectiveness of the developed model using the dataset, the Matrix Factorization algorithm was first applied to address sparse ratings. Then the K-Means algorithm was applied to cluster the publications according to their domain of relevance using the ratings. The Cosine similarity algorithm was further applied to generate similarities for the top one hundred (100) publications, considering each publication as a user case as shown in Fig. 6. The time taken to complete the computation of the similarities was 0:00:00.614037 seconds, which implies that the system is capable of delivering recommendations fast. The detailed results are as follows:

1) Simulation results of the existing expertise and research paper recommender systems: Tables I and II show the results of the selected existing models for expertise recommendation [13] and research paper recommendation [37] systems. The table shows the True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) values over ten (10) iterations, where;

- TP implies true positive and represents relevant experts and research papers recommended.
- FN implies false negative and represents irrelevant experts and research papers recommended.
- FP implies false positive and represents relevant experts and research papers not recommended; and
- TN implies false negative and represents irrelevant experts and research papers not recommended.

In Table I, the values for the TP, TN, and FN fluctuated whereas the values for the FP remained constant for all iterations, which suggest that all recommended items were relevant. In Table II, the values for the TP, TN, FP and FN fluctuated, which suggests that the precision is not maximum.

Algorithm 1 Research Paper and Expert Recommendation Pseudocode

Input: $\vec{R}_u$, *User Query*
D, Dataset of Experts & Research papers
Return: *RL, list of recommended research papers & experts*
1. *Begin*
2. *RL* $\leftarrow \varnothing$;
3. $\vec{R}_u \leftarrow \varnothing$;
4. $\vec{R}_v \leftarrow D$;
5. *Prompt* $\vec{R}_u$;
6. *Compute* $Cos\_Sim(u, v) = \frac{\vec{R}_u \bullet \vec{R}_v}{|\vec{R}_u| \bullet |\vec{R}_v|}$;
7. *IF* $\exists \vec{R}_u \notin D$;
8. *RL* $\leftarrow$ *Not available, enter new search;*
9. *IF* $\exists \vec{R}_u \in D$;
10. *Search;*
11. *Run K-Means Algorithm;*
12. *Run Matrix Factorisation and K-NN;*
13. *For each expert in D identified in Line 9; do*
14. *Calculate* $E_{impact} = \sum_{i=1}^{n} \frac{1}{i}\left(\frac{1}{T_p}\left(N_{(cp)^I} + N_{(jp)^I} + N_{c^I}\right) + (A_{re})^I\right)$;
15. *Generate details - "the expert is recommended because he has: "* $+N_{cp}+N_{jp}$ *"conference & Journal publications cited "* $+N_c$ *"with average rating"* $+A_{re}$ *"over the past three years";*
16. *For each research paper in D identified in Line 9; do*
17. *Generate* $(A_{rr})i = \frac{R_r, i}{N_u, i}$ *and* $(N_c)i$;

    Where A_{rr}, *is the average rating for each research paper,* R_r *is the total rating for the research paper and* N_u *is number of users that rated the research paper and i represents interval in years in all cases.*
18. *Return RL;*
19. *End*

Algorithm 2: Rating Prediction Using Matrix Factorization

Input: training matrix Z, the number of features K, Regularization Parameter λ, Learning Rate ε
Output: row related model matrix X and Column related Model Matrix Y

1. *Initialize X, Y to uniform real* $(0, \frac{1}{\sqrt{k}})$
2. *Repeat*
3. *For random* $z_{ij} \epsilon Z$ *do*
4. *Error* $\leftarrow X_i * Y_j - V_{ij}$
5. $X_i \leftarrow X_i - \varepsilon(Error.\boldsymbol{Y}_j^T + \lambda X_i)$
6. $Hj \leftarrow Hj - \varepsilon(Error.\boldsymbol{X}_j^T + \lambda Y_j)$
7. *End for*
8. *Until convergence*

Algorithm 3: Tag Based K-Means Clustering Algorithm

Input: Target number of clusters K, Dataset of research papers D, Set of tags R_c *and* R_m,
Output: An assignment of research paper to clusters using tags.

1. *Start*
2. *Initialization: arbitrarily choose K midpoints of cluster C1,C2,C3,…..,Ck*
3. *Calculate: estimate the space connecting every data position and group midpoints*
4. *Affectation: allocate the data position to the group midpoint whose space from the group midpoint is least of all the group midpoints.*
5. *Bring up to date / calculate: Redo calculation of the latest cluster midpoint using* $\boldsymbol{V_i} = \frac{1}{\alpha_i}\sum_{j=1}^{\alpha_i \in \mathbb{N}} X_i$ *where, '*α_i*' signifies the amount of data positions in* i^{th} *group and* $\mathbb{N}$ *is a set of natural numbers. Redo calculation of the space connecting every data position and latest obtained group midpoints.*
6. *Discontinue condition: do again the Affectation steep pending when no data position was assigned again.*
7. *Stop*

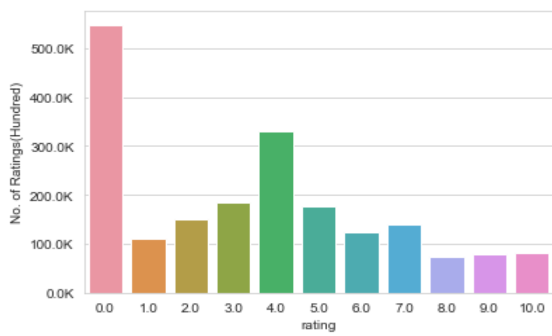


Fig. 4. Distribution of Ratings over Dataset.

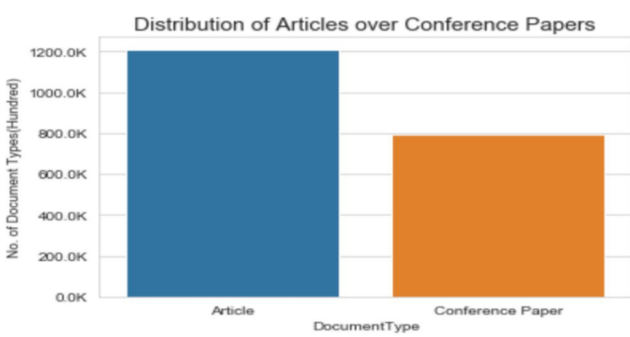


Fig. 5. Distribution of Journal Articles and Conference Proceedings.

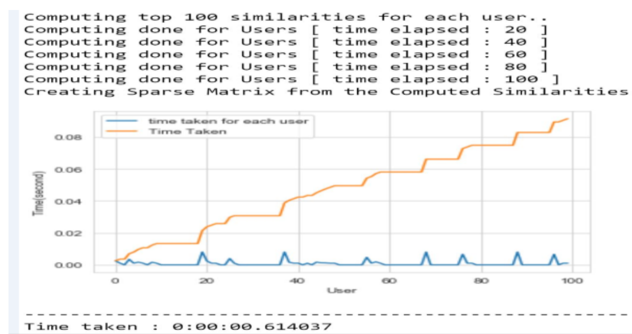


Fig. 6. Cosine Similarity Computation Time for Top 100 Users.

TABLE I. RECOMMENDATION RESULTS OF THE EXISTING EXPERTISE SYSTEM

Iterations	TP	FP	TN	FN
1	93	0	183	15
2	37	0	184	70
3	59	0	199	33
4	89	0	149	53
5	55	0	191	45
6	47	0	189	55
7	68	0	196	27
8	69	0	204	18
9	57	0	186	48
10	46	0	212	33

TABLE II. RECOMMENDATION RESULTS OF THE EXISTING RESEARCH PAPER SYSTEM

Iterations	TP	FP	TN	FN
1	31	27	164	69
2	58	13	178	42
3	33	48	143	67
4	46	43	148	54
5	54	13	178	46
6	86	18	173	14
7	54	23	168	46
8	87	6	185	13
9	88	12	179	12
10	76	5	186	24

The values showed that the existing systems were able to recommend available experts and research papers from the database based on the dataset and the user's query.

2) Simulation results of the developed expertise and research paper recommender systems: Tables III and IV show the results of the developed system for expertise recommendation over ten (10) iterations. The values for the TP and FN fluctuated, whereas the values for the FP and TN remained constant. The results show that the respective developed systems were able to recommend available experts and research papers from the database based on the dataset and the user's query.

3) Simulation results of the developed integrated system: The results of the developed system for the integrated system of Expertise and Research paper recommendation over ten (10) iterations are presented in Table V. The values show the capability of the developed integrated research paper and expertise recommendation system to generate useful sets of research papers and experts according to the user's query. This is in line with the set theory because a collection of expertise and research papers were defined as a set and when the recommendation is zero (0), it implies an empty set. The values are also in line with choice theory because each recommended set of expertise and research papers was based on an assumed user's choice of query using the available keywords in the database.

TABLE III. RECOMMENDATION RESULTS OF THE DEVELOPED EXPERTISE SYSTEM

Iterations	TP	FP	TN	FN
1	93	0	191	7
2	63	0	191	37
3	67	0	191	33
4	45	0	191	55
5	86	0	191	14
6	77	0	191	23
7	92	0	191	8
8	91	0	191	9
9	95	0	191	5
10	73	0	191	27

TABLE IV. RECOMMENDATION RESULTS OF THE DEVELOPED RESEARCH PAPER SYSTEM

Iterations	TP	FP	TN	FN
1	71	0	191	29
2	89	0	191	11
3	87	0	191	13
4	78	0	191	22
5	77	0	191	23
6	92	0	191	8
7	68	0	191	32
8	95	0	191	5
9	94	0	191	6
10	69	0	191	112

TABLE V. RECOMMENDATION RESULTS OF THE DEVELOPED INTEGRATED SYSTEM

Iterations	TP	FP	TN	FN
1	57	7	212	15
2	66	4	161	60
3	42	3	226	20
4	78	3	180	30
5	75	13	147	56
6	97	18	152	24
7	92	23	145	31
8	90	6	172	23
9	89	12	178	12
10	86	5	170	30

C. Evaluation Results

The performances of the existing and developed expertise and research papers recommendation systems were evaluated using precision, recall, F-measure, accuracy, and Root Mean Square Error (RMSE) as follows: The detailed results are as follows:

- Precision shows the recommended system's capacity for showing only relevant experts, while trying to minimize mixing them with irrelevant ones. This is calculated as follows:

$$\text{Precision (P)} = \frac{\text{TP}}{\text{TP+FP}} \quad (9)$$

- Recall represents the coverage of relevant experts and research papers that the recommender system can generate. In other words, it measures the capacity of the system to generate all the relevant experts and research papers present in the database. This is calculated as follows:

$$\text{Recall (R)} = \frac{\text{TP}}{\text{TP+FN}} \quad (10)$$

- The recommendation accuracy, which indicates the correctness of the recommendation, is calculated using Equation 11.

$$\text{Accuracy} = \frac{\text{(TP+FN)}}{\text{(TP+TN+FP+FN)}} \quad (11)$$

- The F-Measure, i.e., the harmonic mean, gives a combined output of both precision and recall for each system. This is calculated as follows:

$$F_1 = 2\frac{X.Y}{X+Y} \quad (12)$$

where X = Precision and Y = Recall.

- The Root Mean Square Error (RMSE) is used to measure the error in the predicted ratings of the recommender system. It is an error in predicting user ratings to address rating sparsity. This is calculated as follows:

$$RMSE = \frac{\sqrt{\sum_{n=1}^{N}(\hat{r}_n - r_n)^2}}{N} \quad (13)$$

Where: $\hat{r}_n$ means the predicted rating while r_n means the true rating. N is the number of rating prediction pairs between the actual data and prediction result.

1) Evaluation of the existing and developed expertise recommendation system: Table VI and Fig. 7, Table VII and Fig. 8 show the evaluation results of the existing and the developed Expertise recommender systems respectively. Both systems gave a constant average precision value of 1.00, which implies that all the recommended experts were relevant. Also, the developed system gave an increase of 17% in average recall value, an increase of 7% in accuracy, an increase of 13% in F-measure, and a decrease of 0.35% in the root mean square error (RMSE) over the existing expertise recommender system. The decreased RMSE value of the developed system implies that there is a wide margin between the minimum and maximum values of the user ratings as contained in the dataset.

TABLE VI. EVALUATION RESULTS OF THE EXISTING EXPERTISE RECOMMENDATION SYSTEM

Iterations	Precision	Recall	Accuracy
1	1	0.86	0.95
2	1	0.35	0.76
3	1	0.64	0.87
4	1	0.63	0.82
5	1	0.55	0.85
6	1	0.46	0.81
7	1	0.72	0.91
8	1	0.79	0.94
9	1	0.54	0.84
10	1	0.58	0.89
Average	**1**	**0.61**	**0.86**
F-measure			**0.75**
RMSE			**2.06**

TABLE VII. EVALUATION RESULTS OF THE DEVELOPED EXPERTISE RECOMMENDATION SYSTEM

Iterations	Precision	Recall	Accuracy
1	1	0.93	0.98
2	1	0.63	0.87
3	1	0.67	0.89
4	1	0.45	0.81
5	1	0.86	0.95
6	1	0.77	0.92
7	1	0.92	0.97
8	1	0.91	0.97
9	1	0.95	0.98
10	1	0.73	0.91
Average	**1**	**0.78**	**0.93**
F-measure			**0.88**
RMSE			**1.71**

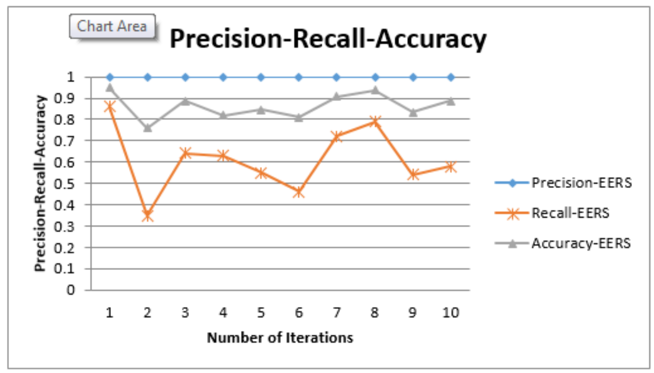


Fig. 7. Evaluation Graph for the Existing Expertise Recommendation System (EERS).

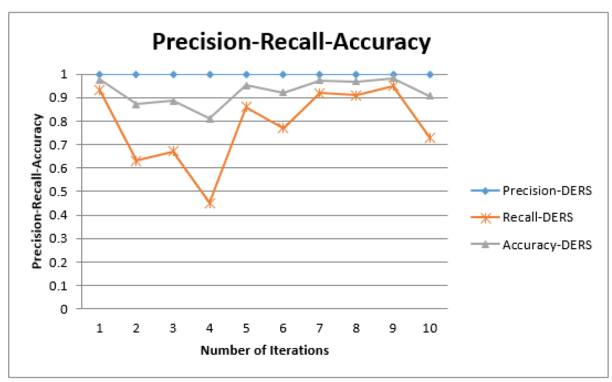


Fig. 8. Evaluation Graph for the Developed Expertise Recommendation System (DERS).

2) Evaluation of the existing and developed research paper recommender systems: Table VIII and Fig. 9, Table IX and Fig. 10, show the evaluation results of the existing and the developed paper recommender system, respectively. The result of the developed system gave an increase of 26% in precision value, an increase of 18% in average recall value, an increase of 12% in accuracy, an increase of 22% in F-measure, and a decrease of 0.47% in the root mean square error (RMSE) over the existing paper recommender system. The decreased RMSE value of the developed system implies that there is a wide margin between the minimum and maximum values of the user ratings as contained in the dataset.

3) Evaluation results of the developed integrated system: Table X and Fig. 11 show the evaluation results of the developed integrated system over ten (10) iterations. The resulted precision value implied that 90% of all the recommended research papers and experts were relevant, which is a very satisfactory performance. The resulted recall value implied that 73% of all relevant research papers and experts were recommended. The resulted accuracy value implied that the correctness of the recommendation was 86%. The resulted F-measure value implied that the combined effect of precision and recall is 81%, which shows that the capability of the developed research paper and expertise recommender system to recommend relevant research papers is very satisfactory. The lower value of 1.69 root mean square error is the effect of matrix factorisation to address rating sparsity.

The results showed that the proposed integrated system performed well based on Precision, Recall, and F-Score and RMSE values. The proposed expertise recommendation component was consistent with the existing system with an average precision value of 1.00, which is an implicit improvement with respect to the recommendation conditions. However, it performed better than the existing system in terms of recall, with average values of 0.61 and 0.76 for the existing and proposed systems respectively, implying a 15% improvement. Furthermore, the proposed research paper recommendation component outperformed the existing system based on precision values by 26% and recall values by 18%. Collectively, the integrated system proved to be reliable by recording appreciable average precision and recall values of ninety percent (90%) and seventy-three percent (73%), respectively, in addition to eighty-one percent (81%) harmonic mean.

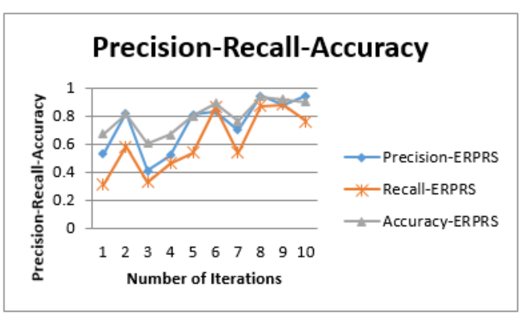


Fig. 9. Evaluation Graph of the Existing Research Paper Recommendation System (ERPRS).

TABLE VIII. EVALUATION RESULTS OF THE EXISTING RESEARCH PAPER RECOMMENDER SYSTEM

Iterations	Precision	Recall	Accuracy
1	0.53	0.31	0.67
2	0.82	0.58	0.81
3	0.41	0.33	0.6
4	0.52	0.46	0.67
5	0.81	0.54	0.8
6	0.83	0.86	0.89
7	0.7	0.54	0.76
8	0.94	0.87	0.93
9	0.88	0.88	0.92
10	0.94	0.76	0.9
Average	**0.74**	**0.61**	**0.80**
F-measure			**0.66**
RMSE			**1.87**

TABLE IX. EVALUATION RESULTS OF THE DEVELOPED RESEARCH PAPER RECOMMENDER SYSTEM

Iterations	Precision	Recall	Accuracy
1	1	0.71	0.9
2	1	0.89	0.96
3	1	0.87	0.96
4	1	0.78	0.92
5	1	0.77	0.92
6	1	0.92	0.97
7	1	0.68	0.89
8	1	0.95	0.98
9	1	0.94	0.98
10	1	0.38	0.7
Average	**1**	**0.79**	**0.92**
F-measure			**0.88**
RMSE			**1.70**

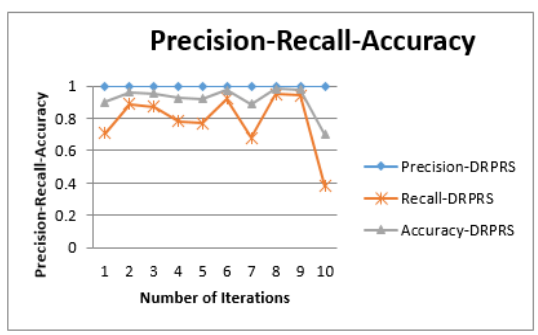


Fig. 10. Evaluation Graph of the Developed Research Paper Recommendation System (DRPRS).

TABLE X. EVALUATION RESULTS OF THE DEVELOPED INTEGRATED SYSTEM

Iterations	Precision	Recall	Accuracy
1	0.89	0.79	0.92
2	0.94	0.52	0.78
3	0.93	0.68	0.92
4	0.96	0.72	0.89
5	0.85	0.57	0.76
6	0.84	0.8	0.86
7	0.8	0.75	0.81
8	0.94	0.8	0.9
9	0.88	0.88	0.92
10	0.95	0.74	0.88
Average	**0.9**	**0.73**	**0.86**
F-measure			**0.81**
RMSE			**1.69**

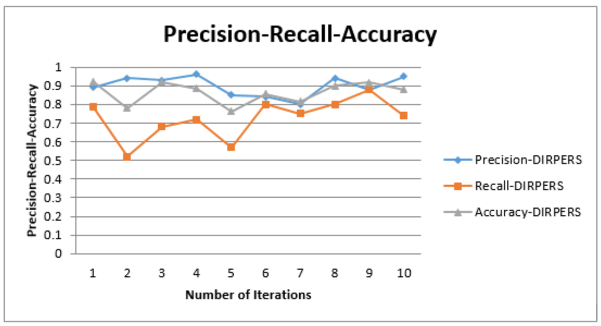


Fig. 11. Precision, Recall and Accuracy Graph for the Developed Integrated System.

V. CONCLUSION

This study has introduced an invaluable means of ensuring reliable recommendation of expertise and research papers in academic research, which would translate into enhanced collaborative research activities among academics. The study concludes that incorporating the similarity of keywords of methods used in carrying out research as reflected in the published research paper to the similarity of keywords of the concept of the research greatly improved the recommendation of relevant research papers. Furthermore, evaluating the expertise of the researchers (experts) using the expert's impact criteria and proving more details about them, also enhanced the recommendation of the relevant experts to the expertise seeker. Finally, the synergized research paper and expertise recommendation system greatly facilitated the simultaneous finding of relevant research papers and experts using a single query, thereby reducing too many avoidable and unnecessary queries, and saving time. Thus, this combined approach to research papers and expertise recommendations is a positive dimension towards achieving great research experience and collaboration among researchers in an academic environment.

VI. FUTURE WORK

The cold start problem continues to surface even as research for a better solution continues. Further studies would consider how to incorporate a reviewer's feedback as an implicit rating to alleviate this challenge in and enhance the results obtained from the matrix factorization method in the research paper recommendation component. The evaluation criteria for expertise will be expanded to make them more robust, while efforts will also be made to create a context-sensitive recommendation for the integration system.

ACKNOWLEDGMENT

This Research was funded by the TETFund Research Fund" and Africa Centre of Excellence OAK-Park.

REFERENCES

[1] H. M. Aghdam M. Analoui, and P Kabiri, "A Novel Non-negative Matrix Factorization Method for Recommender Systems," International Journal of Applied Mathematics and Information Sciences, Vol. 9, 2015, pp. 2721-2732.

[2] D. Richards, M., Taylor, and P. Busch, "Expertise Recommendation: A triangulated Approach," International Journal on Advances in Intelligent Systems, Vol. 2, No.1, 2009, pp. 12-25.

[3] D. W. McDonald, "Evaluating Expertise Recommendations," In ACM International Conference on Supporting Group Work (GROUP01), Colorado, USA. 2001.

[4] J. S. Iyilade, G. A. Aderounmu and M. O. Adigun., "An Agent-Based Approach for Finding a Supervisor in an Academic Environment" In proceedings of the 3rd International Conference on Education and Information Systems: Technologies and Applications, 2005, pp 71-75.

[5] A. T. Olaniran, I.O., Awoyelu, A.O. Amoo and B.O. Akinyemi, "A Hybrid Recommendation Architecture for Nigerian Online Stores" British Journal of Applied Science and Technology, Vol. 12, No. 5, 2016, pp. 1-9. http://dx.doi.org/10.9734/BJAST/2016/21466.

[6] A. T. Olaniran, I.O., Awoyelu, A.O., Amoo, B.O. Akinyemi and R.O. Abimbola., "An Enhanced Hybrid Item Recommender Model for Nigerian Online Stores" International Journal of Applied Information Systems, Vol. 10, No. 1, pp. 31-42, (2015http://dx.doi.org/10.5120/ ijais2015451459, USA.

[7] G. A. Aderounmu, I.O. Awoyelu and A. O. Adetoye, "Agent-based approach to Selecting Views and indexes in a Data Warehouse. Science Focus: An International Journal of Biological and Physical Sciences, Vol. 3, 2003, pp. 61-68.

[8] B. O Akinyemi, O.D., Akinwuyi and G.A. Aderounmu, "A Collaborative Agent - Based Valuation System: Nigeria in Context" Journal of computer science and its application, Vol.22, No. 2, 2015, pp. 93-103.

[9] A. O. Ajayi, G.A. Aderounmu, M.O. Adigun, "Web-based Expert System for Property Valuation Using Mobile Agent Technology" In proceedings of the International Conference on Artificial Intelligence and Soft Computing (IASTED-04), November 22-24, St Thomas US Virgin Islands, 2004, pp. 26-31.

[10] K. Haruna I.M., Akmar D., Damiasih J., Sutopo and T., Herawan, "A collaborative approach for research paper recommender system," PLoS ONE, Vol. 12, No. 10, 2017, pp. 1-17.

[11] T. Phuksenga and S. Sodseeb, "Recommender System Based on Expert and Item Category," Engineering Journal, Vol. 22, No. 2, 2018, pp. 157-168.

[12] A. Kalaï, C., Zayani I., Amous and F. Sèdes, "Expertise and Trust-Aware Social Web Service Recommendation," 14th International Conference on Service Oriented Computing, Oct 2016, Banff, Canada, 2016, pp. 517-533.

[13] C. N. Mabude and I. O. Awoyelu, “Performance Evaluation of a Bi-directional Feedback Approach for Expertise Recommendation in Academic Research,” World Journal of Computer Application and Technology; Vol. 3, No. 2, 2015, pp.27-39.

[14] C. N. Mabude, I.O., Awoyelu and A.O. Amoo, “Design of a Hybrid Recommender System with Dial Feedback Structure,” In proceedings of the 8th International Conference on Application of Information and Communication Technologies to Teaching, Research and Administration (AICTTRA 2013), Nigeria, 2013.

[15] X. H., Pham, T.T., Nguyen, J.J. Jung and N.T. Nguyen, “Spear: A New Method for Expert Based Recommendation Systems,” Cybernetics and Systems: An International Journal, Vol. 45, No. 2, 2014, pp. 165-179.

[16] E., Davoodi, M., Afsharchi, and K. Kianmehr, “A Social Network-based Approach to Expert Recommendation System”, 2013, Available at: http://www.znu.ac.ir/members /afsharchim/pub/HAIS12-GT-2105.pdf. Date Accessed: 10th May, 2016.

[17] Y. Xu, X., Guo, J., Hao, J., Ma, R.Y.K., Lau, and W. Xu, “Combining Social Network and Semantic Concept Analysis for Personalized Academic Researcher Recommendation,” Elsevier: Decision Support Systems, Vol. 54, 2012, pp.564–573.

[18] A. Datta, J.T., Yong, and A. Ventresque, “ T-RecS: Team recommendation system through expertise and cohesiveness,” In Proceedings of the 20th ACM International Conference Companion on World Wide Web, Hyderabad, India, 2011, pp. 201 – 204.

[19] S. Budalakoti, D., DeAngelis and K.S. Barber, “Expertise Modeling and Recommendation in Online Question and Answer Forums,” In Proceedings of IEEE 2009 International Conference on Computational Science and Engineering, pp. 481-488.

[20] M.M., Afsar, T., Crump, and B.H. Far, “An exploration on-demand article recommender system for cancer patients’ information provisioning”. In: E. Bell, F. Keshtkar (eds.) Proceedings of the Thirty-Fourth International Florida Artificial Intelligence Research Society Conference, North Miami Beach, Florida, USA, May 17-19, 2021. https://doi.org/10.32473/flairs.v34i1.128339.

[21] M. Bereczki, “Graph neural networks for article recommendation based on implicit user feedback and content” Master’s thesis, KTH, School of Electrical Engineering and Computer Science (EECS), 2021.

[22] Z., Ali, G., Qi, K., Muhammad, B., Ali, and W.A., Abro, “Paper recommendation based on heterogeneous network embedding” Knowledge Based Systems 210, 106438, 2020, pp.0950-7051. https://doi. org/10.1016/j.knosys.2020.106438.

[23] A., Chaudhuri, N., Sinhababu, M., Sarma, and D., Samanta, “Hidden features identification for designing an efficient research article recommendation system” International Journal of Digital Library, Vol. 22, No. 2, 2021, pp.233–249. https://doi.org/10.1007/ s00799-021-00301-2.

[24] Y., Kang, A., Hou, Z., Zhao, and D., Gan, “A hybrid approach for paper recommendation.” IEICE Transactions on Information and Systems E104.D(8), 1222–1231 (2021). DOI 10.1587/transinf.2020BDP0008.

[25] G., Manju, P., Abhinaya, M.R., Hemalatha, G., Manju-Ganesh and G.G., Manju, “Cold start problem alleviation in a research paper recommendation system using the random walk approach on a heterogeneous user-paper graph” International Journal of Intelligent Information Technologies Vol. 16, No. 2, 2020, pp.24–48. https://doi.org/10.4018/IJIIT.2020040102.

[26] N., Sakib, R.B., Ahmad, M., Ahsan, M.A., Based, K., Haruna, J., Haider, and S., Gurusamy, “A hybrid personalized scientific paper recommendation approach integrating public contextual metadata” IEEE Access 9, 2021, pp.83080–83091. https://doi.org/10.1109/ACCESS. 2021. 3086964.

[27] L., Ahmedi, E., Rexhepi, and E., Bytyci, ”Using association rule mining to enrich user profiles with research paper recommendation. International Journal of Computing and Digital Systems, Vol. 11, No. 1, 2021, pp.1141-1146. http://dx.doi.org/10.12785/ijcds/110192.

[28] J. Sun, J., Ma, X., Cheng, Z., Liu and X. Cao, “Finding an Expert: A Model Recommendation System,” In Proceedings of Knowledge Management and Business Intelligence: 34th International Conference on Information Systems, 2012, pp. 1-10.

[29] R .L. L Sie, H., Drachsler, M. Bitter-Rijpkema and P. Sloep, “To whom and why should I connect? Co-author recommendation based on powerful and similar peers,” International Journal of Technology Enhanced Learning, Vol. 4, No. 1/2, 2012, pp. 121-137.

[30] K. Sugiyama, and M. Kan, “Scholarly Paper Recommendation via User’s Recent Research Interests,” ACM Joint Conference on Digital Learning, 2009, pp. 29-38.

[31] J. Lee, K., Lee, and J. G. Kim, “Personalized Academic Research Paper Recommendation System,” arXiv:1304.5457v1 [cs.IR], 2013, pp. 1-8. Available at: https://arxiv.org/pdf/1304.5457.pdf.

[32] B. M. Maake, S. O., Ojo, and T. Zuva, “A Serendipitous Research Paper Recommender System,” International Journal of Business and Management Studies, Vol. 11, No.1, 2019, pp. 38 – 53.

[33] A. Kanakia, Z., Shen, D., Eide, and K. Wang, “A Scalable Hybrid Research Paper Recommender System for Microsoft Academic, “In Proceedings of the 2019 World Wide Web Conference, May 13–17, 2019, San Francisco, 2019, pp. 2893 – 2899.

[34] P. Zhao, J., Ma, Z., Hua and. S. Fang, “A Novel Personalized Academic Knowledge Sharing System in Online Social Network,” in Proceedings of the 51st Hawaii International Conference on System Sciences, 2018, pp. 989-998.

[35] H. Sahijwani and S. Dasgupta, “User Profile Based Research Paper Recommendation,” 2017, Available Online: https://arxiv.org/pdf/1704. 07757.pdf.

[36] A. Sardar, J., Ferzund, M. A., Suryani, and M. Shoaib, “ Recommender System for Journal Articles using Opinion Mining and Semantics,” International Journal of Advanced Computer Science and Applications, Vol. 8, No. 12, 2017, pp. 213-220.

[37] R., Sharma, D., Gopalani, and Y. Meena, “Concept-Based Approach for Research Paper Recommendation,” Springer International Publishing, LNCS, Vol. 10597, 2017, pp. 687–692.

[38] N. Akhtar, “A Model- Based Research Material Recommendation System for Individual Users,” Transactions on Machine Learning and Artificial Intelligence, Vol. 5, No.2, 2017, pp 1-8.

[39] S., Philip, P.B. Shola, and A. O. John, “Application of Content-Based Approach in Research Paper Recommendation System for a Digital Library,” International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 5, No.10, 2014, pp.37-40.

[40] B. Madhushree, “A Novel Research Paper Recommendation System,” International Journal of Advanced Research in Engineering and Technology; Vol. 7, No.1, 2016, pp. 7 – 16.

[41] C. Nascimento, A. H. F., Laender, A. S., Da Silva, and M. A. Gonçalves, “A Source Independent Framework for Research Paper Recommendation,” ACM Joint Conference on Digital Learning, June 13–17, Ottawa, Ontario, Canada, 2011, pp. 297-306.

[42] W. Choochaiwattana, “Usage of Tagging for Research Paper Recommendation,” 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE), Vol.2, 2010, pp. 439-442.

[43] Q.V., Le, and T. Mikolov, “Distributed representations of sentences and documents,” In proceedings of the 31st International Conference on Machine Learning, Beijing, China, June 21–June 26., Vol. 14, 2014, pp. 1188–1196.

[44] W. Zhao, R., Wu, and H. Liu, “Paper recommendation based on the knowledge gap between a researcher’s background knowledge and research target,” Elsevier: Information Processing and Management, 2016, Available Online: http://dacemirror.sci- hub.tw/journal article/f3e12ba212c04513bfde1906237b439d/zhao2016.pdf.

Organizational Architecture and Service Delivery Re-Alignment based on ITIL and TOGAF: Case Study of the Provincial Development Bank

Asti Amalia Nur Fajrillah, Muharman Lubis, Irmayanti Syam
Department of Information System, Telkom University
Bandung, Indonesia

***Abstract*—The operations function area is the core function areas of the development bank to serve its customer related to the financial needs. Interestingly, the total scope of services provided in this case can be categorized as small or not optimal compared to the total population of its coverage area in the West Java and Banten. The lack of customer confidence in the services offered can be said as one of the reason cause by the unpreparedness of organization to adopt business agility and technological innovation as their alignment framework. Thus, as the beginning, IT planning in the function area can be utilized as the solution to be implemented to increase the service delivery performance through strengthening the organizational architecture. To support this, Enterprise Architecture (EA) should align business and IT with mapping ITIL best practice as a foundation and practical direction to bring the company operational services to have sustainability in growth, profit and satisfaction. This study delivers the roadmap design using the TOGAF framework to identify the current state of the company and the desired IT architecture with business strategies in the area of operations functions.**

***Keywords*—Organization; service; innovation; alignment; ITIL; TOGAF; alignment**

I. Introduction

Information Technology (IT) is one of the functional aspects that is needed to support the operational performance of a company or organization to achieve the objective based on the designated of vision and mission. Indeed, IT can be a differentiating factor between companies especially to provide insight of innovation within organizational architecture and ensuring the service delivery of core competence to the market segment. Therefore, the IT application has become the main focus in helping companies achieve success and excellence amid increasingly fierce business competition, which in the further lead to the creation of competitive advantages that bring uniqueness of the company. The rapid progress of IT is expected to be able to facilitate every processes related to company information, from collecting, processing, to reporting information. Actually, many important issues for bank that should be considered due to the emergence of advanced technology that drag the business process to be shifted dramatically that influence on how to perceive towards the importance of architecture such as personalization security and policy, universality of interaction and communication, leveraging access with clear separation and easy integration, service component development as well as disruptive innovation of financial technology [1, 2, 3, 4]. Thus, the key to successful execution related to the time when the deployment of IT can efficiently support the achievement of strategies, goals, and business needs. Thus, the alignment between IT strategy and business strategy can provide solutions to organizations that rigorously and frequently face pressure from competitor or even the supplier. The development of the role of IT and the increasing value of high investment have led to increasing demands for the added value that mostly the company go to IT to provide the answer.

Actually, organizational architecture composed of strategy, structure, systems, skills and culture that provide room for operation to satisfy the customers in delivering the product or service to certain extent that allow sustainability. In this era of transformation, many business processes are successful due to the utilization of the existing technology to create competitive advantage. However, achieving these benefits is not easy with numerous problems faced by the company that often occur as the business environment change both from within and outside the company. Often, it occurs continuously that force the companies to adapt to the new environment, which of course required a lot of resources if they do not want to fail. Other problems that arise turned out to be from internal namely declining employee performance, lack of controlling from the leadership in responding to dynamic market or even the lack of achievement in the company targets or incomplete task on schedule lead to huge impact to the growth, profit and of course customer satisfaction.

The complexity of the existing problems with the development of IT is directly proportional and has penetrated various industrial sectors. One of the clearest examples is the use of IT in the banking industry, wherein the utilization of IT related products can provide a variety of services to consumers with fewer workers. Therefore, the use can be called as not optimal due to several restrictions that bank cannot do by the regulation such as that the investment or the ownership to real sector. As with any business, banks need to be aware of wise spending due to limitation to convert its profit into real assets. In addition, the banking industry today also is facing a series of new situations that specifically support the need for efficiency with the changing customer preferences and expectations, new competition and new technologies influence the nature of banking services. By moving to a technology-based digital model while retaining an important aspect of traditional interpersonal business models are extremely difficult because the regulation has set the categories of business that bank can do or cannot do. Of course, to stay competitive, banks need to invest in technology, marketing, automation, and self-service capabilities, as well as improve traditional investments in traditional branches and systems. All of these changes are occurring in an industrial environment that is witnessing narrowing margins, slowing deposit growth, and the potential situation for a recession with optimization of customer driven strategy. Thus, the basic premise of solution for this collection of problem by realignment of organizational architecture and service delivery through restructuring and withdrawing a low-margin business line to migrate with a business line that is essentially more cost-effective and makes the bank more profitable. By setting this direction as the priority, bank can take a strong approach to strategic planning, assessing the minimum commitment of resources needed to compete in a particular business area, and identifying opportunities to differentiate themselves from their competitors. In many cases, this means that traditional banks may choose to terminate the unconventional businesses such as professional financing and payment processing.

On the contrary, these strategic shifts may require banks to increase their investment and costs in the short term to improve profitability and efficiency in the long run. In this case, development bank in west java has compiled key performance indicators (KPI) to provide periodic evaluations for productivity and profitability, which engaged directly in financial services by applying IT governance. Based on the current strategic plan, there are several processes that have not met their targets, both in terms of business, data, applications, and technology. This study wants to explore the realignment process with formulating the roadmap by using the combination of TOGAF as EA framework for organizational architecture and ITIL as service delivery model. Implementation of information systems (IS) that are aligned with the organization requires the existence of a roadmap to define major goal in desired goal of company to establish business architecture, IS and technology that can align business strategies with IT. As provincial development bank in Indonesia, it has the objective to support the master plan and government policy through distributing credits as well have role for intermediacy and restructurization of business, especially in the era of pandemic. Thus, the risk should be identified and controlled concusively to maintain the profitability in the above average for return of equity and sharings.

II. Literature Review

A. Organizational Perspectives

To succeed in today's competitive business environment, companies need a clear business strategy that is supported by other organizational strategies. By defining coordination as a strategic consensus or organizational fit, it is suggested that the strategic priorities be adjusted at the operational level to support the business unit level [5], [6]. Organizational resilience is portrayed as an important ability of the organization to cope with unprecedented changes in the business environment, fierce market competition, etc., and prosperity to achieve global success. To deal well with volatile external and internal changes, flexibility provides organizations a fast and easy way to improve business processes [5], [7]. Therefore, organizational resilience can be defined as a business need related to rapid response or the ability of an organization to detect unexpected changes in customer demand, competitors' strategy, business environment, etc. This means that agile organizations can respond quickly and efficiently to changes in the market. With regard to the relationship between strategic integrity of IT business and organizational agility, external factors such as environmental uncertainty are one of the issues that should be considered a mitigating role. When the external environment is uncertain, companies need to anticipate imminent changes and adapt their organizational strategies to environmental changes [8].

Process-level values are a major factor in enabling IT-based strategic design. This affects the performance of IT resources and creates unique resources that cannot be easily replicated as a job. They facilitate the functional activity of the agility or operational capabilities of an organization, they facilitate the transfer of dynamic capabilities from the diversity between companies, or the way in which their functional processes are carried out, which are complementary resources and capabilities [9]. Many factors are said to explain the conditions under which labor value is created, that is, organizational, industrial and macroeconomic factors, i.e. shortage of qualified personnel and training, can adversely affect the success implementation [10]. Information technology poses a difficult dilemma for current management, which proper investment can provide a competitive advantage over competitors while the IT budget allocation with the increased burden of measuring yield, the result is not good at best [11]. Through critical analysis of the relationship between the six aspects of the overall value of e-government, which are the improvement of management efficiency, open government (OG) function, and improvement of ethical and professional behavior have been closely linked to improve government management [12].

Other studies described cost, time, convenience, personalization, proper communication, ease of information retrieval, trust, adequate information, and participation in decision-making to explain most of the overall e-government perception [13], [14]. Organizations are implementing changes to adapt to rapid context within turbulent environments; these changes often affect business and IT. In most cases, changes affect organizational elements which do not clearly define adaptation and exclude elements that could lead to inconsistencies or misalignment [15], [16]. It is the responsibility and duty of officials or managers in the face of environmental uncertainty to understand important events and their changes to prepare the proper response towards how they affect the organization. Agility allows organizations to provide survival in such environments quickly, innovatively and creatively. It requires rapid changes in company structure and composition, which can be implemented in various strategic planning activities [17].

B. Technological Innovations

Interestingly, there is a productivity paradox that investment in information technology does not ultimately affect productivity growth [18]. Despite apparent efforts by the scientific community to articulate the vision of strategic alignment of IT and business, IT and business are inevitably very dynamic, so the application of the organization to the real world still suffers from a gap. Contributions on the topic have grown very rapidly, but the relationship between assumptions and conclusions is inconsistent and inconsistent. The main reason is that the proposed model is highly conceptual and detached from the actual reality of the organization [19], [20], which is related to distorted vision of myopia as another paradox in this area. Meanwhile, organizations also question the return on investment in formal security awareness strategies intertwined with training, campaigns, and reward systems [21]. This can improve ethical and unethical perceptions of IT in users, but these perceptions are often vague, making it difficult to monitor their effectiveness.

The use of technology must address the quality of information such as accuracy, completeness, timeliness and reliability related to the data subject that should be available to the public where innovation should be properly designed [22], [23]. However, keep in mind that innovation alone is not enough to sustain a work life without being integrated with corporate strategy, readiness, employee motivation, and business vision. Therefore, learning from the failures of today's enterprises that have lost control of the market can provide general insights into how innovation works [24]. Various infrastructure network organizations bring direct efficiency to data delivery and processing, enabling networks to be recognized as a data flow model with the technological elements underlying communication between computer systems [25]. To improve profits, growth, satisfaction, performance, business strategy from competition and supplier pressure, companies can take full advantage of information technology and take standard operating steps to maintain service. You need to help your organization achieve its goals [26], [27].

C. Business Objectives

Complex business processes designed on the standards of each product and component can make process automation, customization, and customization difficult. Therefore, the realization of new projects is the creation of business case study documents that are responsible for identifying current customer requests and finding potential clients, maximizing business value and improving corporate governance structure and begins with a feasibility study [28]. It is necessary to build an information system that helps you create fast and efficient services within the business goals associated with your organization's vision and mission [29]. Competition between companies has become a major trend towards business service delivery where certain standard such as ITSM (Information Technology Service Management) has been utilized as a source of practical guidance for creating process improvements [30].

On the other hand, maintaining the integrity of business and information technology has four perspectives: strategy implementation, technology transformation, competitiveness, and service level. In supporting the business vision, one of the primary functions that greatly support the services a company provides is the logistics and maintenance functions. Second, to support your company's business operations, you need to use the right IT applications and devices to optimally manage your family and provide your company requirements to serve your customers. Scalability is the most important feature of an IT architecture and has a major impact on all aspects of a project, so it is important to incorporate it into the enterprise architecture design [31]. Therefore, it is best to design a business evaluation before providing a business service. In this case, outcome questions can influence service design and provide information that can help you set realistic expectations. You may need to collect data before you can provide the service, but you can get a successful evaluation after providing the service. Some evaluation questions can be overcome with current employee data, accounting, or company performance such as costs of keeping turnover. Moreover, new data such as changes in skills, supervisors, and customer satisfaction are needed to compare performance over different time periods and interpret business service outcomes [32].

By emphasizing the level of business and IT integration that represents the administrative aspect of the organization and the value that IT contributes to, which represents the complexity aspect of the system in business. Assessing the degree and level of consistency can deliver specific detail on the state of business objective. These can reflect the coordination profile of the organization, which ultimately consists of an operational resource profile, a strategic resource, and a weapon profile [33]. Top management should provide a technical view to better clarify the logic and options associated with the IT strategy that support the chosen business strategy. Therefore, the role of the IT manager should be that of the technology engineer by efficiently and effectively designing and implementing the required IT infrastructure that is compatible with the external components of an IT strategy [34]. Every business unit has its own business requirements, and while IT departments need to make standards cost-effective, they need to be analyzed [35]. IT governance enables both business and IT personnel to fulfill their responsibilities to support business and IT integration and create commercial value from IT-supported business investments. Business strategy drives the design of information system infrastructure and organization through new technology capabilities [36], [37].

III. Methodology

The TOGAF ADM framework and ITIL best practices have essentially the same cycle. For the difference, TOGAF focuses on developing business architecture and ITIL's scope of IT development is effective and efficient. TOGAF does not cover runtime development and maintenance, namely how services are produced to be delivered. To build an EA with the TOGAF approach in meeting market needs in a sustainable cycle, it is necessary to add additional phases to ITIL best practice, namely service operation, so as to improve operational services. The advantages possessed by ITIL are the general concept and a series of integrated best practices that can help companies meet market needs in a sustainable cycle. These best practices provide a competitive advantage through value creation and agile change. Companies that adopt ITIL can be optimized in service provision, because service levels have been agreed upon so that it is easy to provide consistent service. Apart from the advantages offered by ITIL, there are also disadvantages, namely the large cost because the concept is comprehensive. ITIL is holistic, covering the entire IT governance framework. In addition, in its utilization, it requires special training by employees, so it can be said that ITIL is less flexible as can be seen in Fig. 1.

If the attention to alignment is greatly recognized, its implementation remains very limited. Therefore, this means that while the actors in the organization cannot distinguish between alignment and imbalances, the lack of a method for assessing alignment makes the task at the decision level very difficult [38]. Today, building a robust information system is not enough. To enable companies to perform, compete and evolve, business information systems and processes must be permanently coordinated and fully aligned with their strategies. In today's dynamic business environment, TOGAF can be effectively applied to any business enterprise due to its nature as open consortium framework. This is a series of operations aimed at developing software to meet customer requirements and obtaining solutions through cross-functional teams. Prior to ITIL and TOGAF combination, the mapping process to understand the overall knowledge, perspectives and characteristics of both methodologies and to coordinate strategies makes appropriate transition based on workflow, continuous learning and organizational culture. ITIL can interact with IT service providers and the business world to facilitate compliance with TOGAF principles regarding business requirements and ongoing business communications. From an ITIL perspective, this process provides a comprehensive understanding of the company's initial requirements. TOGAF will then help you prioritize your requirements as can be seen in Fig. 2. The concept of service level management complements the management of business relationships related to non-functional requirements.

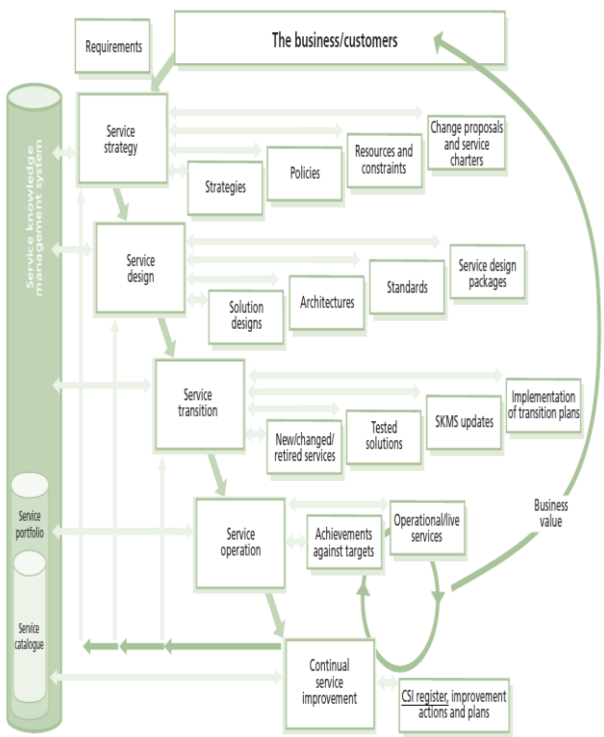


Fig. 1. Lifecycle of ITIL Integration.

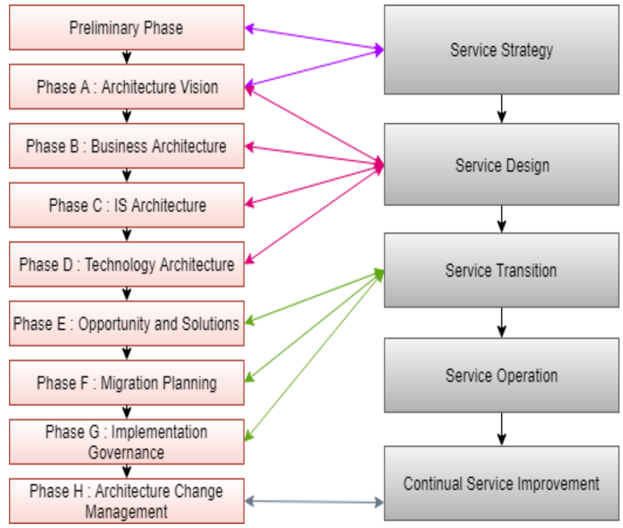


Fig. 2. TOGAF and ITIL Mapping.

IV. Result and Analysis

A. Mapping the Function and Structure

In carrying out their respective functions, it is necessary to map the organizational structure into a diagram. The organizational structure is a picture that describes the type of organization, the position of the organization's department, and the type of authority of the official, the field and work relations, the line of command and responsibility, the span of control and the organizational leadership system. The organizational structure of a company displays a hierarchy that contains the components that make up the company, which clearly describes the position, function, rights and obligations of each position within the scope of the company, as well as the work relationship between individuals. With the existence of an organizational structure, it can be seen the work relationship between units and roles within the company so that it is able to build good communication in running the company's business. Another goal is that each component in the company can function optimally and be able to move the wheels of the company efficiently and effectively as they should. In carrying out their respective functions, it is necessary to map the organizational structure into a diagram. The organizational structure is a picture that describes the type of organization, the position of the organization's department, and the type of authority of the official, the field and work relations, the line of command and responsibility, the span of control and the organizational leadership system.

Executives need to set the direction for IT initiatives. They need to establish policies regarding obtaining, using, and disposing of company information assets. Business priorities are set when the value is expected to be realized only by the business executives, as the sponsors or protagonists can drive the realization of value from IT related projects. Information technology itself cannot provide value. Therefore, business policies must be translated into IT organization priorities and projects. It is important to have this partnership to ensure that the right IT priorities are set. Means of the governance process for these include steering committees, IT business coordinators, resource and budget allocation processes, IT organizations, and evaluation. However, IT executives must play their part in prioritizing workloads that are rated as the biggest hurdles by non-IT executives [39].

B. Service Strategy

IT is an important facilitator for today's organizations to function properly by allowing IT companies to change the way they organize their business processes, communicate with their customers, and provide services. A consistent description of the different focus areas in EA can provide insights, enable communication between stakeholders and guide complex transformation processes. At the same time, the emergence of new products and services is often separated by many companies or enterprises' IT such as IT and IT products that support business and management components, or those embedded in products and industrial automation. It should be tightly integrated into what supports. One of the potential benefits of such an integration is the ease of accessing the data collected by a large number of instances of the product's IT during operation. Enterprise architecture is likely to be a vehicle to support both the ongoing coordination of business and information technology, product-specific IT integration and enterprise IT [40]. Service strategies provide a direction for organizations to define their business strategies and are effectively supported by IT strategies through several key activities such as market definition.

A sufficient set of EA focal areas allow you to design your organization in a "sufficiently appropriate" way to minimize resource use within EA modeling. In this case, it may be appropriate to reduce the number of typical focus areas. A possible way to do this is to integrate people and network focus areas into the organizational structure and integrate jobs and time into business processes. An enterprise architecture that supports and improves enterprise operations, aligns business with integrated enterprise information systems, responds to change, and helps organizations present their business strategies [41]. Additionally, the expert advisor can be used as a means to coordinate business and IT strategies, developing innovative ways to rethink the scope of business processes using management-related information and IT resources to drive effectiveness. [42]. However, the competitive advantage depends on customer satisfaction, process life cycle, resource management, task allocation and scheduling, estimated cost [43], and core competencies related to maintainability and sustainability.

As the environment changes and business needs change rapidly, executives need to be updated every time to manage their business more effectively. Companies need to make information available not only at the business unit level, but also at the organizational level. To improve information at the organizational level, you need to incorporate all the information held by your business unit. Computers are now the most important thing in data processing, as computer media is now used by almost everyone in all disciplines and it increases the power of computers in terms of physical form, equipment, software, and hardware. The increasing need for data and information in the business functions of higher education institutions is the driving force behind the use of information systems [43]. Today's IT development is growing rapidly and has a huge impact on the endless business competition between companies with information technology plays an important role in business activities.

Simplification of the business processes can increase the productivity and profits. As a result, many companies want to use better technology as a basis for competing with their companies and other competitors. The better the technology you use, the more competitive your company will be and the greater your ability to achieve your goals. It can also improve the service with customer care, production and business innovation [44]. In the information system engineering stage, the data to be achieved and the applications to be built are combined to meet architectural principles and visions and enable business engineering visualization. Therefore, all companies that want to improve the effectiveness of their business operations and want to be superior to other companies in the commercial competition strive to implement information technology in their business [45], [46]. Unfortunately, the vast majority of organizations try to create synergies, but ultimately they are separated in a piecemeal fashion that makes it difficult to find and manage results. Creating synergies requires more than just concepts and strategies. An enterprise value proposition outlines a strategy for creating value through alignment, but does not explain how to achieve it. The alignment strategy should be complemented by the alignment process. The amendment process, as well as budget preparation, should be part of the annual governance cycle. When plans change at the company or business unit level, executives may need to readjust their organization in new directions.

C. Service Design

Service design provides guidelines for developing services and service management processes. It covers design principles and methods for transforming strategic goals into a portfolio of services and service assets. To achieve a competitive advantage, technology development requires good planning. EA often links a company's technical systems to elements of strategic management. It is also a technical and management practice that aims to improve the performance and quality of a company by helping to design technical resources, business processes, and strategic directions to help the company achieve its goals. In addition, the corporate architecture also aims to create a suitable environment for the company so that it can properly implement the business operations according to the corporate strategy. EA frameworks typically consist of business engineering, information engineering, application systems engineering, and infrastructure technology engineering. This is an important way to the company's success and plays a major role in increasing its demands for speed, agility, efficiency and quality [47], [48]. Dynamic business requires more and more data flows to departments to support decision making, timely and efficient sourcing of spare parts, inventory management, accounting, human resources, and product distribution. It gets more complicated with functional units [48].

Consistency between business strategy and information technology is required by the fact that most organizations grow organically and that company design is primarily based on intuition rather than well-defined and executed principles. Actually, the solution can be offered is through business process improvement or reengineering through analyzing existing processes develops the ability to increase the managerial decision-making, allowing organizations to better understand IT functions by linking business functions with existing IT resources, and is expected to be able to do so. The target business process should accommodate the simplification in the sense not only allowing the automation but also the flexibility and customization in the activity flow [49].

An overview of the fundamental vision of Architecture and Architecture is provided through areas that include business, data, applications, and technology for business applications, technology infrastructure and enterprises. Defining the architectural vision is an important step in an organization's value chain analysis [50]. These companies must make a high level of effort to extract value from their IT infrastructure, align them with strategic objectives, and gain their advantage over their competitors. Service migration improves functionality and improves new or modified services by identifying service requirements when designing services to be effective and fulfilled in the service process and controlling the risk of failure or outage by clarifying the guidelines for moving to operational.

All aspects of the company influence the effectiveness and efficiency of information systems or information technology applications in business processes [51]. EA can provide insights into the current use of IT in business operations. It can also provide insight into the future use of IT in business operations. Finally, the roadmap for the evolution of the IT overview from the current situation to the future, along with the temporary state between them, can support the company's direction in operations and delivery services [52]. Therefore, support activities can add value to the main products a company provides, such as a company's infrastructure, to support the system and support the daily core activities of the company.

Human resource management is related to employee management such as recruitment, training and compensation. It has to work for the identification purpose for the sake of improvement and maintaining the service as the customers' expectation. On the other hand, technology development relates to all techniques for converting inputs into outputs, which helps reduce the amount of time while running the main activity. As can be seen in Fig. 3, the monitoring and cotrolling architecture should allow the process alignment towards business initiative and plan especially related to validating report and risk management. In addition, financing consists of obtaining the inputs or resources for supplying raw materials and negotiating until the appropriate price is obtained [53].

D. Service Transition

Data Dissemination Diagram is a diagram that describes the relationship between application components, data entities, and business services. The purpose of this artifact is to show the relationship between business services and the applications and data used. This diagram also shows how logical data is physically represented in application components. Unlike the processes and technologies, organizations are made up of different characteristics of people, organizational structures, communication models, and cultures, business processes are viable and available technologies are critical to the success of ITIL's implementation but the primary ingredient related to the competence and determination of the employee [54], [55]. The main issue with implementing ITIL is the fact that ITIL tells organizations "what they have to do" but "how they should do" is not clear [56], [57]. Today, there is no complete framework that can be used as a comprehensive framework for IT governance to ensure consistency between service management and organizational concepts and manufacturing. In fact, different frameworks are often used as complements, most often at the same time. In addition to the difficulties associated with managing both initiatives, parallel EA and ITIL projects involve overlapping investment and costs. In fact, using a shared infrastructure does not allow you to avoid repeating data, procedures, and personnel or to comply with different initiatives (ITIL and EA) [58].

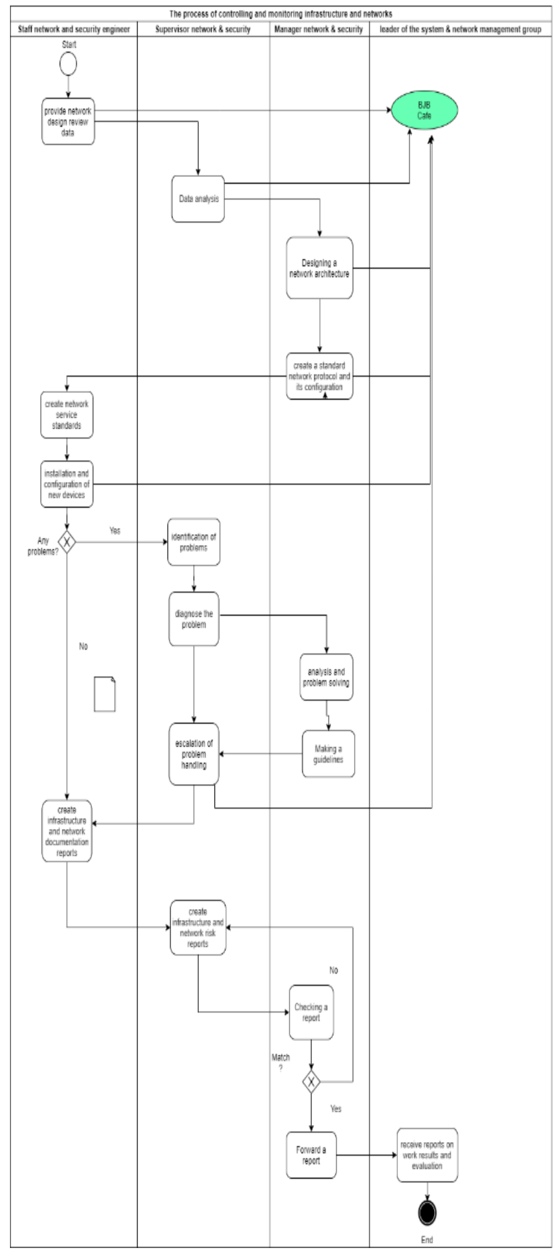


Fig. 3. Controlling and Monitoring Infrastructure.

Brown and Winter [59] proposed an EA extension to incorporate ITIL and SOA. In their proposal, EA is an important concept, but ITIL only considers IT operations. The EA provides an overview of the IT architecture, and ITIL is designated as an important part of the management process for providing services. Along with IT services, the concept of SOA is also integrated into the EA at the application architecture level. ITIL and SOA are integrated into the EA as a framework for providing IT services, and this integrated research focuses only on the IT services provided. On the other hand, Thorne is in close agreement with his previous research, working on the relationship between ITIL and TOGAF, but with a different focus, which mentioned that EA is a spiritually fundamental concept of organizational engineering, with a focus on EA development, and includes ITIL as a framework for the operational model of the IT services provided. In his research, he argued that both frameworks can be used together by mapping the two approaches. TOGAF covers EA development, participates in the conceptual life cycle of products, and ITIL guarantees the delivery and management of IT services to users and consumers [60].

As can be seen in Fig. 4, API or microservice become primary choice in order to allow hybrid infrastructure to be taken in the network management system that accommodate legacy system to be integrated with high end technology. Despite the perceived need for different teams and tools, TOGAF requires an EA repository and ITIL requires a configuration management database (CMDB), so the two frameworks complement each other. It fits. A recent study [61] promoted by the Prophet of God provides a service-based framework for the EA to meet the ITSM requirements of ITIL V3 and extends the EA to include the service architecture layer of ITIL service design. This indicates that it is necessary. An IT service architecture model has been proposed and is the service layer of EA. However, it does not explain how to do this or the relationships between architectures. Therefore, to be able to implement an effective and efficient system requires planning, implementation, organization, and evaluation according to the needs and values of each organization [62]. Interestingly, previous studies have shown that ITIL adoption has increased, the number of operational benefits achieved has increased, and the level of maturity of business alignment with IT has increased [63]. To maintain results, IT service upgrades identify critical success factors (CSF) as benchmarks for improving IT service implementations. Or achievement, goal conditions need fairness process, recommendations to eliminate gaps that occur in each process. Establish IT service management policies and guidelines that consist of policy documents, standard operating procedures (SOPs), and other related document attachments from operational equivalence. The results of assessing the fairness of the process have led to the achievement of the goal, and there are some records to help the goal be achieved continuously [64].

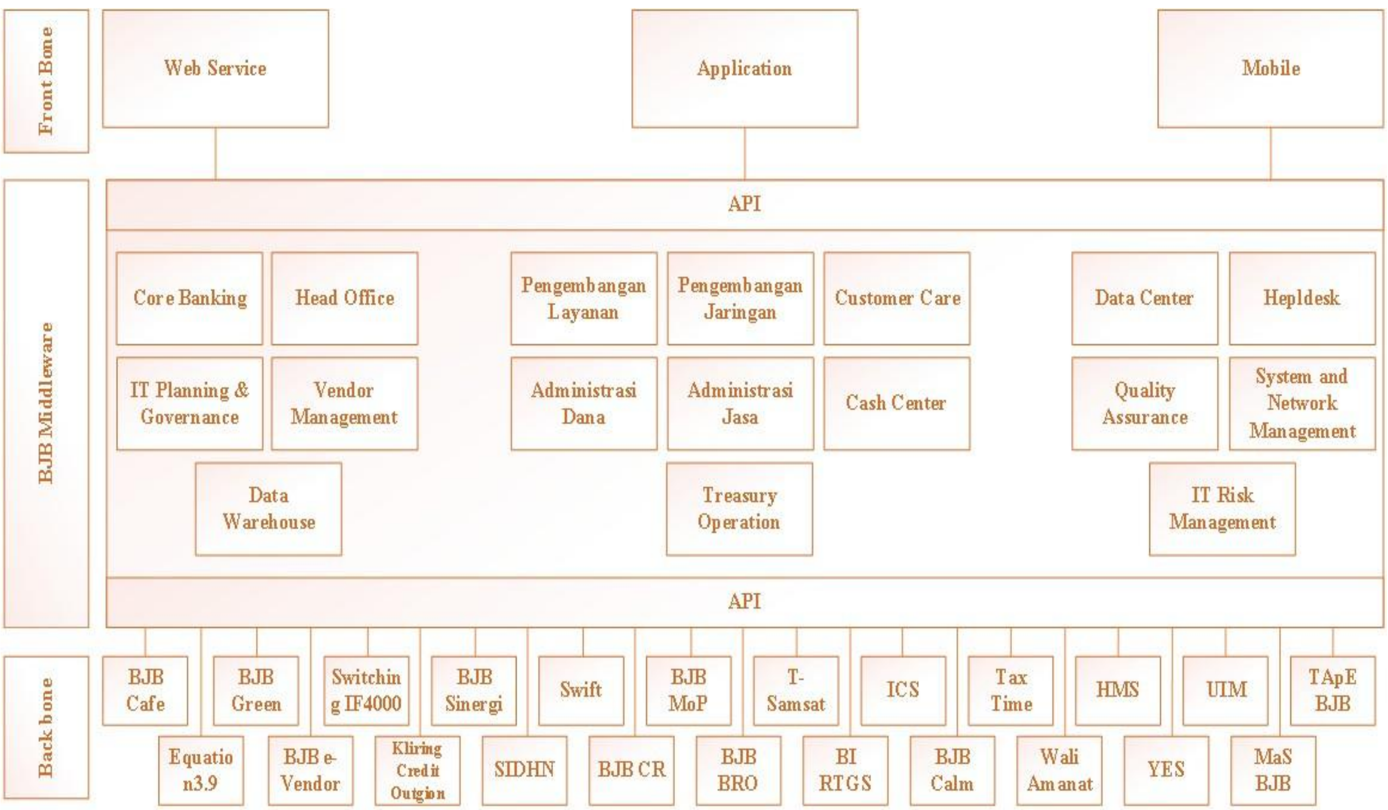


Fig. 4. Targeting Infrastructure Network Management.

Apply information technology in a timely and appropriate manner according to your current business strategy, objectives, and needs to align your business requirements with the relevant IT services. To the extent that IT applications, infrastructure, and organizations enable and support business strategies and operations, including processes, to achieve this by aligning information system (IS) capabilities with business goals [65]. However, many IT development teams do not pay attention to program management frameworks and standards, so planning, reporting and monitoring work results, especially if not all employees of the team have sufficient experience. It becomes difficult to control. It would be desirable for each Software Company and IT consultant to implement common frameworks and standardizations so that the software development process is more oriented and high quality services are produced to increase the business relation [66], [67].

E. Service Operation

As can be seen in Table I, the service operation manages the services the company is currently using, focuses on service management practices, and ensures that the service provider's service to its customers is effectively achieved. Degree 1 (Unstructured Data Exchange), is an unstructured data exchange. Degree 2 (Structured Data Exchange), is a structured data exchange and has certain method. Degree 3 (Seamless Sharing of Data), is an exchange of data using a system so that the data has been automated. Degree 4 (Seamless Sharing of Information), is a continuation of Degree 3 where information exchange occurs in real time. Code for data types includes A: Formal message exchange, B: Common data exchange, C: Complete data exchange and D: Real-Time data exchange.

TABLE I. IT OPERATION

		IT Operation			
		IT handling Problems	IT Asset Governance	System Testing	Secured Application and Users
IT Operation	IT handling Problems	4D	4D		4D
	IT Asset Governance	3C	4D		3C
	System Testing	2B		4D	3C
	Secured Application and Users	3C		2B	4D

Creating business value is generally discussed as an established, thoughtful, and anticipated process that revolves around the maturity of the process and the capabilities of the target organization [68]. Consistency with business needs, achievable level of service through negotiation, predictable and consistent operation, efficiency of service delivery, services and processes that can be measured and improved, IT services and potential current and future business needs. A common language that defines alignment, improved quality of IT services, and long-term costs for delivering short-term services, better communication between suppliers and customers, and terms [69]. Lack of project management contributes significantly to project failure without communication or tracking of employee comments throughout the ITIL implementation process. Administrators sought to implement ITIL as part of their business rather than as a project, without considering a stable organizational culture.

According to end users, most employees weren't interested in implementing ITIL because top management didn't notify them of the need for ITIL, and most employees weren't willing to commit to the project. It should be noted that after management began asking about their lack of commitment to new processes, employees began to change their behavior towards ITIL certification in their relevant duties [70]. The research results shown that the most mentioned factors for implementing ITIL are training (21%), senior management commitment (18%), awareness (11%), and change management (9%). %), operational implementation and maintenance (7%) and project management (7%) [71]. Changes in organizational culture can be laborious and time consuming and can be difficult to achieve within the scope and time frame of a process project such as ITIL. Open and honest relationships with vendors are important and are usually discussed in the context of strategic and long-term outsourcing projects. Effective coordination between multiple vendors is critical to a successful implementation [72].

The facilitative condition is required in the implementation, which is defined as the degree to which an individual believes that there is an organizational and technical infrastructure to support the use of the system. If there is a top management commitment, this will provide the project with the resources and funding needed to train consultants, staff, and acquire the necessary skills [73]. Understanding the reasons behind the success or failure of an ITIL implementation requires an analysis of the types of problems by corporate actors regarding the relevance of the ITIL critical success factors implementation and their reflexive self-judgment. Employees make their own decisions and act in ways that enhance or weaken the rationalization of IT department resources [74]. Therefore, to support organizations in the IT sector and to make them more service-oriented, current best practices need to include a modern service-oriented perspective [75]. Competitiveness is the sum of all the factors that contribute to business continuity in a competitive environment, and understanding the importance of adopting good business practices by a company is a continuous activity that provides a competitive advantage that is useful for expansion.

This can be defined as gaining a position in a favorable, profitable and sustainable market, making it difficult to enter new competitors and efficiently searching for new markets. The data show that companies are more interested in results related to the operational part of information technology, and do not show focus, management, or oversight related to management's strategic vision. There may be other factors related to the corporate culture. It prioritizes operational activities due to the business impact on innovation. This includes different types of changes, depending on your organization's resources, capabilities, strategies, and requirements [76], [77]. It is imperative that each organization analyze the context listed for market demand, as there is room to implement innovation in the industry without investigating or analyzing its impact on competitiveness.

F. Continual Service Improvement

Continuous Service Improvement (CSI) has room for improvement at all stages of the service lifecycle by measuring and improving the efficiency, effectiveness, service levels, technologies, and processes used to manage the entire service, which in the end support on how to identify the area. The most important goal of an IT plan is always to align IT features and activities with business objectives and requirements, such as making decisions about the scope, size, and pace of IT projects. However, aligning IT with business goals requires a stable, reliable, and relatively homogeneous business foundation in the form of a widely accepted, committed transportation plan. Unfortunately, the availability and reliability of this business plan is arguably the best for most organizations. Even if available, the business plan often expires after the IT planner accepts the business plan. In many cases, the time lag between the business planning process and the IT planning process is too long to allow [78]. In conducting business operations, organizations face specific challenges, issues, and issues that arise from the expectations of their respective sectors and clients. In general, satisfaction comes from comparing emotions that resemble a passenger's image with real or physical situations. Dissatisfaction does occur when the actual situation does not meet the customer's terms and standards [79]. However, simplifying business processes by eliminating unnecessary areas and activities is the first step for an organization to better control and manage the flow of infrastructure development.

In fact, attention should be paid to the impact of organizational changes that are consistent with growing interest in strategies for creating an appropriate organizational culture with project managers and leadership action capabilities [80]. Unfortunately, many large companies are inefficient in implementing their needs applications due to the difficulty of implementation and the complexity of the implementation process [81]. When you start using the product, the executive can add value in the form of personalization or personalization to increase user participation, and when starting to use the required product, the social or environmental awareness of distribution by the general approach of representing the finished product through certain changes cause many problems. Knowing the right things about the company's business processes can create a specific model for more effective and efficient user engagement lead to continual improvement taking into account the many factors that a company can consider implementing software, which is expected to be possible in the long term execution [82]. By linking IT and business with ITIL, you can strengthen communication between IT and business, share knowledge, increase efficiency, and link activities within your organization [83]. With ITIL, you help deliver value by increasing effectiveness, competition and market space while increasing customer satisfaction. The resource department is responsible for protecting the confidentiality, availability and integrity of service assets by providing accurate configuration information and ensuring that only allowed components and allowed changes are used [84], [85].

The role of technology in banking has already been mentioned several times, but given its widespread impact at the enterprise level, the use of technology and automation is also worth paying attention to individually as part of an overall efficiency improvement effort with comprehensive goals are in three aspects. Firstly, providing an application that allows customers to execute transactions and obtain information on a self-service basis without the effort of employees. Secondly, using the technology to reduce the time employees spend searching for information. Lastly, using the automated business rules and decision making models to move work faster and more efficiently throughout your business. Improving vendor management does not just mean putting pressure on sellers to lower prices. Rather, it is a focused effort designed to maximize the value possible from the seller's relationship. Choosing a vendor that closely matches your bank's business goals is important. Maintaining strong vendor performance is supported by SLAs and vendor scorecards for monitoring performance issues such as system availability, response time, and direct costs. These tools help you get a more complete picture of your vendor relationships. Ultimately, improving the organizational success and profitability of a bank requires more than just efficiency. Successful banks must be able to provide value and service to their customers at competitive prices and still generate acceptable profits.

There is no one-size-fits-all approach. Some banks are actively promoting electronic account opening, remote capture of deposits via smart devices, and accounts designed to be virtually paperless. Other banks with large corporate customers often take a completely different approach, focusing on personalized services by relationship managers and support teams dedicated to each eligible account. The high-value businesses created by this approach can offset the additional costs. Other useful tools include visual metrics and performance charts, as well as "in-line" incentives such as rewards that are directly related to individual or group performance and practices, as well as organizational performance and resilience.

Many organizations have also succeeded in redefining work roles, using more flexible work arrangements, providing mobility for offsite work, and outsourcing more professional activities. Electronic documents can be step-by-step with minimal delay and virtually no additional cost. Most importantly, electronic imaging enables parallel processing of documents, allowing multiple steps in progress of a transaction to be completed at the same time. Of course, digital signatures, signature panels, and online processes can often take you one step further by eliminating paper altogether. Technology not only helps automate core processes, but also plays a clear role in efforts to improve banking channels. This affects not only how customers interact with banks, but also how banks communicate important information internally and manage sales and customer relationship activities. Other basic ways to reduce costs include integrating vendors and measuring costs for similar services on the market. Also, keep in mind that vendor relationships can affect a regulator's view of a company's risk profile as can be seen in table II that showed the roadmap for improment and maintaining the demand.

TABLE II. ROADMAP

No.	Project	Estimation Duration (Month)	Period																			
			Year-1				Year-2				Year-3				Year-4				Year-5			
			TW1	TW2	TW3	TW4	TW1	TW2	TW3	TW4	TW1	TW2	TW3	TW4	TW1	TW2	TW3	TW4	TW1	TW2	TW3	TW4
1	BJB Middleware	10																				
2	bjb CR	6																				
3	IT Infrastruktur Planning	10																				
4	MaS BJB	7																				
5	BJB Sinergi	6																				
6	Tax Time	6																				
7	bjb MoP	7																				
8	BJB e-Vendor	5																				
9	TApE BJB	6																				
10	bjb BRO	6																				

The alignment will give better understanding of business management about the importance of information systems, while at the same time improving your understanding of business goals and objectives. Through technological innovation, information systems managers are also increasing the return on IS investment [86]. Indeed, the strategic integrity of business and information systems has been consistently reported as a major concern for business and IT managers in various industries. IT business integration gives organizations greater flexibility and increases the return on their IT investment. This will increase profitability and maintain competitive advantage. Therefore, failure to take advantage of information technology can have a significant negative impact on a company's performance and survival [87]. Business and Information Technology (IT) safety refers to the optimal synchronization of business objectives, dynamic processes and related technical services provided by information technology innovation. This has been taken into account by researchers, IT executives, and consultants since the early 1970s [88]. The stronger impact on financial performance was driven by operational adjustments that suggested that achieving the link between the organization and the IT infrastructure was of prime importance to the model organization. Indeed, it has been suggested that infrastructure and coordination of operations are more important than information technology and business strategy coordination [89].

V. Conclusion

The complexity of the existing problems with the development of Information Technology is directly proportional and has penetrated various industrial sectors. In the IT Operation function, namely: IT management and management that has not been good, limited automation, control of information security that is less effective, and technological infrastructure that does not support business needs. Some of the problems mentioned certainly cause many parties to be disadvantaged, especially the company's business operations and allow for over and even low budget in the company. Another impact obtained is that it can pose a fatal risk in the development of IT that allows fraud to occur so that the Management Information System (MIS) becomes unreliable. It also causes delays when processing reports between units and from branch offices to the center. In its design, EA uses a framework that will provide a blueprint that is used systematically to identify the current state of the company and the desired IT architecture design. Besides, to achieve alignment between IT strategies and business strategies that support optimization of IT governance, a best practice is needed that is used as a foundation for company operational services. The integration between the framework and best practice is expected to complement each other and produce an EA design to support the achievement of the company's strategic goals. Development bank as government owned company engaged in banking that uses the key performance index as an indicator used to determine how far the strategy has been carried out by the company by the company's vision and mission. Therefore, it is expected for verification and validation model can be utilized to evaluate project that has been delivered in respected period as well to analyze the benefit delivered that align with IT business perspectives.

References

[1] N.A. Hamidi, M. Rahimi G.K., A. Nafarieh, A. Hamidi and B. Robertshon, “Personalized Security Approaches in E-Banking Employing Flask Architecture over Cloud Environment”. Procedia Computer Science 21, 18-24, 2013.

[2] A. Kousaridas, G. Parissis, T. Apostolopoulos. “An Open Financial Services Architecture based on the Use of Intelligent Mobile Devices”. Electronic Commerce Research and Applications 7, 232-246.

[3] D. Kellezi, C. Boegelund and W. Meng. "Towards Secure Open Banking Architecture: An Evaluation with OWASP". NSS, LNCS 11928, 185-198, 2019.

[4] M. Lubis, M. Saputra and W.A. Nurtrisha. "Financial Technology Development Framework for Prosperity of the Nation and Potential Direction". *ACM ICCCM 2021*.

[5] S. Heydari, M.H. Shakib and A. Khamseh, "IT-Business Strategic Alignment and Organizational Agility: The Moderating Role of Environmental Uncertainty," Journal of System Management, vol 1, 35-52, 2020.

[6] A. Al-Surmi, C. Guangming and D. Yanqing, "The impact of aligning business, IT, and marketing strategies on firm performance," Industrial Marketing Management, 84, pp. 39-49, 2020.

[7] K. Harsch and M. Festing, "Dynamic talent management capabilities and organizational agility—A qualitative exploration," Human Resource Management, 59(1), pp. 43-61, 2020.

[8] S. Panda and S. Rath, "Strategic IT-business alignment and organizational agility: from a developing country perspective," Journal of Asia Business Studies, Vol. 12 No. 4, pp. 422-440. 2018. https://doi. org/10. 1108/JABS-102016- 0132 .

[9] A.S. Eldin, A. Elnour and R. Hassan, "Process Level Social Media Business Value Configuration of SMEs in Saudi Arabia," Int. Journal of Advanced Comp. Science and Application, vol. 11(11) 2020.

[10] F. Almehmadi, "Developing an Information Management Strategy e-Government in Saudi Arabia," ," Int. Journal of Advanced Comp. Science and Application, vol. 11(11) 2020.

[11] P. Appiahene, N. Ussiph and Y.M. Missah, "Information Technology Impact on Productivity: A Systematic Review and Meta-Analysis of the Literature," Int. Journal of Information Communication Technologies and Human Development vol. 10(3), pp. 39-61, 2018.

[12] J.D. Twizeyimana and A. Andersson, "The Public Value of E-Government – A Literature Reiew," Government Information Quaterly vol. 36(2), pp. 167-178, 2019.

[13] M. Scott, W. DeLone and W. Golden, "Measuring eGovernment Success: A Public Value Approach," European J. of Information Systems vol. 25(3), 2016.

[14] J. Benitez, A. Castilo, J. Llorens and J. Braojos, "IT-enabled Knowledge Ambidexterity and Innovation Performance in Small U.S. Firms: The Moderator Role of Social Media Capability," Information & Management vol. 55(1), pp. 131-143, 2018.

[15] I. Kawtar, B. Salah and D. Karim, "Impact of Change in Business IT Alignment: Evaluation with CBITA Tool," Int. Journal of Adv. Computer Science and Applications vol. 11(10), 2020.

[16] O. Avila and K. Garces, "Change Management Support to Preserve Business-Information Technology Alignment," J. of Computer Information Systems vol. 57(3) 2017.

[17] A. Asil and N. F-H. Farahmand, "Design and Implementation of Strategic Agility Evaluation Model with Structural Equation Modelling Approach," Research Article 18(1) 2019.

[18] D.A. Almajali and Z.M. Dahalin, "Factors Influencing IT-Business Strategic Alignment and Sustainable Competitive Advantage: A Structural Equation Modelling Approach," Communications of the IBIMA 261315, 2011.

[19] M. Henriques, J.B. de Vasconcelos, G. Pestana and A. Rocha, "Strategic Alignment IT-Business: Towards a Proactive e-Public Sector," J. of Information Systems Engineering & Management 4(2), 2019.

[20] M. Henriques, J.B. de Vasconcelos, G. Pestana and A. Rocha,"IT-Business Strategic Alignment in Social era," 14th Iberian Conf. on Information Systems and Technologies, IEEE 2019.

[21] A.R Ahlan, M. Lubis and A.R. Lubis, "Information Security Awareness at the Knowledge-based Institution: Its Antecedents and Measures," Procedia Computer Science 72, pp. 361-373, 2015.

[22] E. Rosmaini, T.F. Kusumasari, M. Lubis and A.R. Lubis, Insight to Develop Privacy Policy in Indonesia," Journal of Physics: Conference Series 978, 2018.

[23] M. Lubis, A.R. Lubis, B. Lubis and A. Lubis, "Incremental Innovation towards Business Performance: Data Management Challenges in Healthcare Industry in Indonesia," MATEC Int. Conf. on Industrial Electrical and Electronics (ICIEE) 2018.

[24] M. Lubis, A.R. Lubis and E. Ernovianti, "Disruptive Innovation Service Oriented Framework: A Case Study of Transportation in Indonesia," Scitepress, Int. Conf. on Multidisciplinary Research (ICMR) 2018.

[25] A. Widjajarto, M. Lubis and U. Yunan, "Architecture Model of Information Technology Infrastructure based on Service Quality at Government Institution," Procedia Comp. Sci. 161, pp. 841-850, 2019.

[26] S.T.N. Andi, A.A.N. Fajrillah and M. Lubis, "IT Roadmap to Improve Business Strategy using TOGAF ADM: A Case Study of Government-Owned Electricity Company," J. of Physics: Conf. Series 1361, 2019.

[27] M. Lubis, H.H.R. Ananza and F.D. Suryaputro, "Analysis and Design of Policy and Standard Operating Procedure for Information Technology in the Communication and Information Services Department," IEEE Int. Conf. on Interactive Digital Media 2020.

[28] M. Lubis, M. Fathoni and A.R. Lubis, "New Product Development Architectural Framework for Sustainability and Innovation within Telecommunication Industry," ACM Int. Conf. on Computer and Communications Management 2020.

[29] A.M. Jannah, M. Lubis, R.R. Saedudin and F.D. Suryaputro, "Designing the Smart Health Function towards Puskesmas (Citizen Health Centre) based on Smart City Concept," IOP Conf. Series: Materials Sc. and Engineering 847, 2020.

[30] M. Lubis, R.C. Annisyah and L.L. Winiyanti, "ITSM Analysus using ITIL v3 Service Operation in PT. Inovasi Tjaraka Buana," IOP Conf. Series: Materials Sc. and Engineering 847, 2020.

[31] A.E. Wardani, A.A.N. Fajrillah, S.F.S. Gumilang and M. Lubis, "Project Evaluation for Business and IT Alignment with Enterprise Architecture for Water Distribution Company," J. of Physics: Conf. Series 1361, 2019.

[32] M.K. Haris, A.A.N. Fajrillah and M. Lubis, "Business Value Assessment for Global Service Provider Industry: Opportunities and Solution," J. of Physics: Conf. Series 1361, 2019.

[33] S. Charoensuk, W. Wongsurawat and K.D. Ba, "Business-IT Alignment: A Practical Research Approach," The Journal of High Technology Management Research 25(2), 2014.

[34] S. Baina, P-Y. Ansias, M. Petit and A. Castiaux, "Strategic Business/IT Alignment using Goal Models," Proc. Of BUSITAL 2008.

[35] A.J.G. Silvius, B. de Waal and J. Smit, "Business and IT Alighment; Answers and Remaining Questions," Pacific Asia Conference on Information Systems, 2009.

[36] A. Ghildyal and E. Chang, "IT Governance, IT/Business Alignment and Organization Performance for Public Sectors," Journal of Economics, Business and Management, vol. 5(6), 2017.

[37] M.A. Ardakan, H. Hatefian and Y. Gorji, "Aligning IT and Business Strategies Adopting an HOQ, a Case Study in a Large Scale Iranian Bank," Int. Conf. on Information and Financial Engineering, IPEDR vol. 12, 2011.

[38] K. Doumi, S. Baina and K. Baina, "Strategic Business and IT Alignment: Representation and Evaluation," J. of Theoritical and Applied Information Technology vol. 47(1), pp. 41-52, 2013.

[39] J. Luftman, R. Papp and T. Brier, "Enablers and Inhibitors of Business-IT Alignment," Communications of the Association for Information Systems vol. 1(11), 1999.

[40] J. Kaidalova, "Enterprise Architecture Modeling for Business and IT Alignment," PoEM, Short and Doctoral Consortium Papers, pp. 108-116, 2015.

[41] E.D. Madyatmadja, J.F. Andry and A. Chandra, "Blueprint Enterprise Architecture in Distribution Company using TOGAF," J. of Theoritical and Applied Information Technology vol. 98(12), 2020.

[42] A. Fergina and I.D. Sumitra, "Designing Enterprise Architecture Planning in Mobile News Applications using TOGAF ADM," IOP Conf. Series Materials Scc. And Engineering 662:022098 2019.

[43] B. Imbing and J. Andry, "Enterprise Architecture Planning for Contata Music School Institute using Zachman," Journal of Systems Integration 3, 2019.

[44] M.M. Mei and J. Andry, "The Alignment of Business Process in Event Organizer and Enterprise Architecture using TOGAF," Jurnal Ilmiah Teknologi Informasi vol 17(1), pp. 21-29, 2019.

[45] F.A. Pasaribu, J.H. Sipahutar, B.P. Situmorang, S. Sfenrianto and E.R. Kaburuan, "Designing Enterprise Architecture in Hospitals Group," Int. Conf. on Inf. and Communications Technology (ICOIACT) 2019.

[46] P. Ranting and J. Andry, "Re-design Business Process at Forwarding Company Based on Enterprise Architecture Planning," Int. J. of New Media Technology (IJNMT) vol. VI(2), pp. 68-73, 2019.

[47] H. Tannady, J. Andry and F.E. Gunawan and J. Mayseleste, "Enterprise Architecture Artifacts Enablers for IT Strategy and Business Alignment in Forwarding Services," Int. J. of Adv. Trends in Comp. Science and Engineering vol. 9(2), pp. 1465-1472, 2020.

[48] B.G. Sudarsono, J. Andry and N. Az, "Design Information System Order Fulfillment using Archimate Modelling," Int. J. of Advanced Trends in Comp. Science Engineering vol. 9(2), pp. 1303-1310, 2020.

[49] B.G. Sudarsono, J.S. Antouw, J. Andry and A. Sani, "Enterprise Architecture Landscape using TOGAF Framework for Offshore Company," Test Engineering and Mgmt vol. 83, pp. 16880-16889, 2020.

[50] N. Zulfarian and I.D. Rosiyadi, "Designing Enterprise Architecture for Academics Information System Platform using the Open Group Architecture Framework Architecture Development Method," IOP Conf. Series Materials Science and Engineering 879:012066, 2020.

[51] B.G. Sudarsono, J. Andry, P. Ranting and A. Abd. Rahman, "Redesign the Forwarding Company's Business Process using the Zachman Framework," J. of Theoretical and Applied Infor. Technology v. 98(16), pp. 3222-3232, 2020.

[52] F.S. Lee, C. Aziza, R. Nathanael and J. Andry, "Architecture Information System in Electrical Distribution Company using TOGAF," Int. J. of Advanced Trends in Computer Sc. and Engineering vol. 9(5), pp. 7149-7156, 2020.

[53] R.E. Riwanto and J. Andry, "Designing Enterprise Architecture Enable of Business Strategy and IS/IT Alignment in Manufacturing using TOGAF ADM Framework," Int. J. of Information Technology and Business vol. 1(2), 2019.

[54] N. Gama, R.N. da Silva and M.M. da Silva, "Using people-CMM for diminishing resistance to ITIL," Int. J. of Human Capital and Inf. Tech. Professionals vol. 2, pp. 29-43, 2011.

[55] R.F.D. Pereira and M.M. da Silva, "A maturity model for implementing ITIL v3," 6th World Congress on Services 2010.

[56] M. Marrone and L. Kolbe, "ITIL and the Creation of Benefits: An Empirical Study on Benefits, Challenges and Processes," 18th European Conf. on Inf. Systems 2010.

[57] A. Lahtela and M. Jantti, "Improving IT service management processes: a case study on IT service support," 17th European Conf. Systems, Software and Services Process Improvement 2010.

[58] N. Gama, P. Sousa and M.M. da Silva, "Integrating Enterprise Architecture and IT Service Management," 21st Int. Conf. on Information Systems Development 2012.

[59] C. Braun and R. Winter, " Integration of IT Service Management Into Enterprise Architecture," ACM Symposium on Applied Computing, pp. 1215-1219, New York, 2007.

[60] S. Thorn, "TOGAF and ITIL". The Open Group (ed.), vol. Catalog number W071, pp. 26, San Francisco 2007.

[61] A. Nabiollahi, R.A. Alias, S. Sahibuddin, "A Service Based Framework for Integration of ITIL V3 and Enterprise Architecture," International Symposium in Information Technology (ITSim), vol. 1, pp. 1-5, Kuala Lumpur 2010.

[62] H. Tannady, J. Andry, Y.T. Suyoto and A. Herlian, "Business Architecture of Public Guest Service for University using TOGAF ADM Framework," Technology Reports of Kansai University vol. 62(5), pp. 2421-2428, 2020.

[63] M. Marrone and L.M. Kolbe, "Uncovering ITIL claims: IT executives' perception on benefits and Business-IT alignment," Inf. Systems and e-Business Management vol. 9(3), pp. 363-380, 2011.

[64] A. Hermanto and G. Kusnanto, "Evaluation of the information technology system services for medium higher education based on ITIL (a case study of polytechnic XYZ)," 4th Int. Conf. on Computer App. and Inf. Processing Technology 2017.

[65] N. Zeinolabedin, M. Khademi and N. Rahbar, "Assessing efficiency of ITIL framework to align business and IT," Research Inventy: Int. J. of Engineering and Science, vol. 2(5), pp. 13-26, 2013.

[66] H.F. Rahmani and E. Hikmawati, "Combining SDLC Method and ITIL Framework by Involving Auditors," Jurnal AKSI (Akuntantsi dan Sistem Informasi) vol. 5(1), pp. 6-12, 2020.

[67] H.N. Prasetyo, "Maturity Level of Business Relation Services based on ITIL Framework in Digital Start up X Indonesia," Test Engineering and Management 83, pp. 8043-8049, 2020.

[68] J. Cusick, "Business value of ITSM. Requirement or mirage," Advancing IT and Software Engineering, January 2020.

[69] A.S. Alqahtani, "Critical Success Factors in Implementing ITIL in the Ministry of Education in Saudi Arabia: An Exploratory Study," Int. J. of Advanced Computer Science and Applications 8(4), 2017.

[70] N. Ahmad and Z.M. Shamsudin, "Systematic Approach to Successful Implementation of ITIL," Procedia Computer Science vol. 17, pp. 237-244, 2013.

[71] S. Sebaaoui, M. Lamrini, A. Bouayad and L. el Abbadi, "Design of an ITIL implementation model in a company," Information systems 2019.

[72] N. Hajiheydari and S. Mehravani, "ITIL Adoption Critical Success Factors: Applying Meta-Synthesis Approach," Int. Conf. on Future Information Technology 2010.

[73] N. Ahmad, N.T. Amer, F. Qutaifan and A. Alhilali, "Technology adoption model and a road map to successful implementation of ITIL," J. of Enterprise Information Management vol. 26(5), 2013.

[74] A. el Yamami, S. Ahriz, K. Mansouri, M. Qbadou and E. Illoussamen, "Developing an Assessment Tool of ITIL Implementation in Small Scale Environments," Int. J. of Advanced Com. Science and App. 8(9), 2017.

[75] S. Cronholm, H. Gobel, M. Akesson, "ITIL Compliance with Service-Dominant Logic," e-Service Journal 11(2), 2020.

[76] J. Schaefer, I.C. Baierle, E. Nara, G.B. Benitez and C. Haetinger, "Corporate Governance and Information Technology: A Hybrid Vision based on COBIT and ITIL for Small Companies," XL encontro Nacional de Engenharia de Producao 2020.

[77] I.C. Baierle, G.B. Benitex, E. Nara, J. Schaefer and M.A. Sellitto, "Influence of open innovation variables on the competitive edge of small and medium enterprises," J. of Open Innovation Technology Market and Complexity vol. 6(4), 2020.

[78] R. Alit, S. Sugiarto and M. Idhom, "Evaluation of Information Technology Infrastructure Management using IT Balanced Scorecard and COBIT Framework 4.1 on Domain Deliver and Support," Proc. of the International Conference on Science and Technology 2018.

[79] M. Lubis, R. Fauzi, A.R. Lubis and R. Fauzi, "A Case Study of Universities Dormitory Residence Management System," 16 th Cyber and IT Service Management (CITSM) 2018.

[80] M. Lubis, R. Fauzi, A.R. Lubis and R. Fauzi, "Analysis of Project Integration on Smart Parking System in Telkom University," 16 th Cyber and IT Service Management (CITSM) 2018.

[81] L. Syafiraliany, M. Lubis and R.W. Witjaksono, "Analysis of Critical Success Factors from ERP System Implementation in Pharmaceutical Fields by Information System Success Model," 4th International Conference on Informatics and Computing (ICIC) 2019.

[82] M. Lubis, I.R. Wardhani and R.W. Witjaksono, "Examining the Means of Engagement (MOE) for Enterprise Resource Planning (ERP) Adoption in Indonesia: Factors and Measures," 3rd Int. Conference on Electrical, Telecommunication and Com. Engineering (ELTICOM) 2019.

[83] M. Gervalla, N. Preniqi and P. Kopacek, "IT Infrastructure Library (ITIL) Framework Approach to IT Governance," IFAC PapersOnLine 51(30), pp. 181-185, 2018.

[84] G. Shaykhian, K. Said and G. Alqarna, "ITIL Its Effectiveness: Decision Makers' Perspectives," Annual Conference & Exposition 2019.

[85] H. Metin, N. Preniqi and M. Gervalla, "The Effect of the Organizational Culture, Organizational Structure and Usage of Social Media on

Corporate Governance," International Conference University for Business and Technology 2017.

[86] R.M. Kekwaletswe and P.C. Mathebula, "Aligning Information Systems Strategy with the Business Strategy in a South African Banking Environment," Proc. of the Conf. for Information Systems Applied Research 2014.

[87] P.V. Malyzhenkov and M.I. Ivanova, "An Architectural Approach to IT-Business Alignment," Information Systems and Technologies in Business vol. 3(41), pp. 56-64, 2017.

[88] A. Ullah and R. Lai, "A Systematic Review of Business and Information Technology Alignment," ACM Transaction on Management Inf. Systems vol. 4(1/4), 2013.

[89] W. Afandi, "The Impact of Strategic IT-Business Alignment: Evidence from Saudi Private and Midsize Enterprises," Int. J. of Business and Social Science vol. 8(10) 2017.

A Lasso-based Collaborative Filtering Recommendation Model

Hiep Xuan Huynh[1]
College of Information and Communication Technology
Can Tho University (CTU)
Can Tho City, Vietnam

Vien Quang Dam[2]
Faculty of Information Technology
Can Tho Vocational College (CTVC)
Can Tho City, Vietnam

Long Van Nguyen[3]
Information & Communication Technology Department
Ministry of Public Security (MPS)
Hanoi, Vietnam

Nghia Quoc Phan[4]*
Assessment Office
Tra Vinh University (TVU)
Tra Vinh Province, Vietnam

***Abstract*—This paper proposes a new approach to solve the problem of lack of information in rating data due to new users or new items, or there is too little rating data of the user for items of the collaborative filtering recommendation models (CFR models). In this approach, we consider the similarity between users or items based on the lasso regression to build the CFR models. In the commonly used CFR models, the recommendation results are built only based on the feedback matrix of users. The results of our model are predicted based on two similarity calculated values: (1) the similarity calculated value based on the rating matrix; (2) the similarity calculated value based on the prediction results of the Lasso regression. The experimental results of the proposed models on two popular datasets have been processed and integrated into the recommenderlab package showed that the suggested models have higher accuracy than the commonly used CFR models. This result confirms that Lasso regression helps to deal with the lack of information in the rating data problem of the CFR models.**

Keywords—UBCF-LASSO; IBCF-LASSO; Lasso regression

I. Introduction

The recommendation models [1][2][7] are widely used in the fields of commerce, education, and entertainment. It saves users time when searching for needed information easily and quickly based on transaction data or rating data of users in the past. For example, Facebook displays ads related to keywords that users search for; YouTube automatically jumps to clips like the clips the user just watched. The Amazon sales site has been very successful in using the recommendation model [8]. Products are rated by users on a scale of 1 to 5 when they shop online. Transactional information and customer evaluations for products are collected to assist in increasing the accuracy of the recommender model. Therefore, every time a customer visits the website, they are always suggested products that are predicted from previously collected data.

The above example shows that the recommender system plays a very important role in e-commerce [4] and many areas of daily life around us. Therefore, building a good recommendation system to find products according to user requirements is the desire of not only researchers but also an important investment in the development strategy of each company.

Many collaborative filtering recommendation models have been successfully applied in the field of online e-commerce [4][7][10]. It is also considered as one of the effective solutions to solve the problem of information explosion for online systems with a rapidly increasing number of users. These recommender models recommend users what products they need to buy based on the results of their ratings on the products. To predict products according to user preferences based on a rating matrix, collaborative filtering recommendation models perform calculations based on the efficient exploitation of statistical methods and data mining techniques [13]. However, collaborative filtering recommendation models still face objective problems that need further research and improvement. It is a matter of lack of information in rating data of users due to new users or new items, or too little rating data of users for items.

In this study, we propose a new method to improve the accuracy of collaborative filtering recommendation models due to the lack of user rating information (new users, new items, and sparse data) by looking at the correlation relationship between users or products based on Lasso regression. In particular, the proposed models are built based on two similarity methods: the similarity method is built based on Lasso regression [15] and the similarity method is built based on rating data [13].

This study is structured into six sections. Section one includes an overview of the recommendation models and research questions. Section two briefly reviews the linear Lasso regression. Section three presents the steps to calculate the similarity based on the linear Lasso regression. Section four proposes the collaborative filtering recommendation models based on similarity Lasso regression. Section five shows the experimental results of the proposed models on two popular datasets (MovieLenss and Jester5k) integrated into the recommenderlab package [13]. The concluding section presents a summary of the results achieved.

*Corresponding Author.

II. Collaborative Filtering

A recommendation system is based primarily on a set of users, a set of items, and a set of user ratings based on those data items; represented in a matrix (see Fig. 1).

Collaborative filtering [7][1][13] is the process of determining the missing values / ratings from a rating matrix. Let $U = \{u_1, u_2, \dots, u_m\}$ be a set of m users and $I = \{i_1, i_2, \dots, i_n\}$ be a set of n items obtained from a certain supermarket.

Ratings [13][7][1] are stored in a $m \times n$ user-item rating matrix $R = (r_{jk})$ where each row represents a user u_j with $1 \le j \le m$ and columns represent items i_k with $1 \le k \le n$. r_{jk} represents the rating of user u_j for item i_k. The values of r_{jk} is rated from 1 to q (for example $q = 5$) or missing. From this point of view recommender systems solve a regression problem to predict missing values [13] (see Fig. 1).

The first approach is to predict the rating value for a $user - item$ pair. Suppose we have a dataset about the interest of m users (users) in n products (items), corresponding to an $m \times n$ matrix, where the element in row i, column j represents the price value the rating of the ith user on the jth item. Our work is to fill the empty values in the matrix, in other words, we will predict the user's rating value on the items that the user has not rated.

In fact, we don't really need to predict all user ratings on items to make recommendations to users. Instead, we just need to suggest k most suitable products (items) to the user or determine k users that best match the item. To find the most suitable k _items for the user, we need to calculate the "distance/similarity/compatibility" to find the k neighbors ($k - neighborhood$) that best match the user u_a (see Fig. 2).

Collaborative filtering algorithms are also divided into memory-based and model-based [18].

A. User-based Collaborative Filtering

User-based recommendation uses the similarity of a group of users' purchasing or purchasing patterns to predict what items that user will buy or choose.

B. Item-based Collaborative Filtering

Item-based recommendation will use similarity in the purchase relationship of items to predict which items the user will buy or choose.

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8
u_1	?	4.0	4.0	2.0	1.0	2.0	?	?
u_2	3.0	?	?	?	5.0	1.0	?	?
u_3	3.0	?	?	3.0	2.0	2.0	?	3.0
u_4	4.0	?	?	2.0	1.0	1.0	2.0	4.0
u_5	1.0	1.0	?	?	?	?	?	1.0
u_6	?	1.0	?	?	1.0	1.0	?	1.0
	:	:	:	:	:	:	:	:
u_a	?	?	4.0	3.0	?	1.0	?	5.0
$\hat{r_a}$	3.5	4.0			1.3		2.0	

Fig. 1. User Rating Matrix (URM).

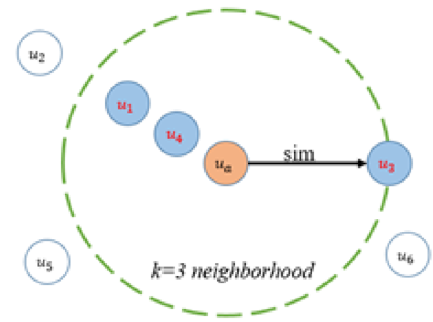


Fig. 2. k- $neighborhood$ (k=3).

III. Lasso Regression

Lasso Regression (Least Absolute Shrinkage Selection Operation) [15] [19] was first proposed by Robert Tibshirani in 1996.

$$\hat{\beta}^{lasso} = \mathrm{argmin}_{\beta} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 \tag{1}$$

$$subject\ to\ \sum_{j=1}^{p} |\beta_j| \le t \tag{2}$$

In *Lagrangian form* [19]:

$$\hat{\beta}^{lasso} = \mathrm{argmin}_{\beta} \left\{\frac{1}{2}\sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \lambda \sum_{j=1}^{p} |\beta_j|\right\} \tag{3}$$

This is a good method to narrow predictors by removing unimportant attributes based on the absolute value of the weights of the regression model by the following formula:

$$\hat{\beta}(\lambda) = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij}\right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \tag{4}$$

The purpose of the Lasso regression model is to minimize prediction errors [6][15]. In practice, the tuning parameter λ controls the strength of the penalty, when λ is large enough some of the coefficients β are exactly zero, in this way can reduce the dimensionality of the model. The larger the parameter λ, the more coefficients β shrink to zero.

There are many advantages to using Lasso regression, first, it can provide a good prediction accuracy because shrinking and removing the coefficient can reduce the variance without significantly increasing the standard deviation. This is especially useful when we have a small number of observations and many attributes.

In addition, Lasso regression helps to increase the interpretability of the model by removing variables that are not related to the explanatory variable, which also avoids the overfitting of the model. So, Lasso regression is a good choice to build a recommendation model that avoids underfit or overfit when we choose too few or too many variables in the model.

IV. Lasso-based Similarity

As described in the previous section, Lasso regression could build a good predictive model for large datasets by selecting important attributes and removing unimportant attributes in the dataset. This method can give good prediction results on a sparse matrix, and it is suitable for overcoming the weakness of the lack of rating data of the collaborative filtering recommender model.

Therefore, the function to calculate the user similarity matrix based on Lasso regression is built as follows:

LUS (rating matrix; newdata)
Input: rating matrix, newdata
Output: URM (user result matrix)
Begin
Step 1: Building Lasso regression based on rating matrix
User_Lasso = Lasso (rating matrix);
Step 2: Using Lasso regression to build user similarity matrix
For each row of the rating matrix
Begin
Value = predict (User_Lasso, newdata)
URM = cbind (URM, Value)
End
Step 3: return (URM)
End

Like building a user similarity matrix, the function to calculate the item similarity matrix based on Lasso regression is built as follows:

LIS (rating matrix; newdata)
Input: rating matrix, newdata
Output: IRM (item result matrix)
Begin
Step 1: Building Lasso regression based on rating matrix
Item_Lasso = Lasso (rating matrix);
Step 2: Using Lasso regression to build item similarity matrix
For each column of the rating matrix
Begin
Value = predict (Item_Lasso, newdata)
IRM = cbind (IRM, Value)
End;
Step 3: return (IRM)
End;

V. Recommendation Framework

This section presents the content of two proposed models based on Lasso regression: the UBCF-LASSO model and the IBCF-LASSO model. The UBCF-LASSO model is designed based on the user similarity matrix integrated between the Lasso-User-Similarity (LUS) matrix calculated by the linear regression Lasso and the user similarity matrix calculated from the rating data in the way of the traditional UBCF models [12][13]. Similar to the above approach, the IBCF-LASSO Model is designed based on the integrated item similarity matrix between the LUS matrix calculated by the linear regression Lasso and the item similarity matrix calculated from the rating data in the way of the traditional IBCF models [12][13].

A. UBCF-LASSO

The UBCF-LASSO model is designed with two input parameters: the user's rating matrix for the items: $R_{UxI} = [r_{i,j}]$ with $U = \{u_1, u_2, \dots, u_n\}$ is n users; $I = \{i_1, i_2, \dots, i_m\}$ is m items, and u_a is an user who needs recommendation.

The UBCF-LASSO model has an overall block diagram design structure as follows.

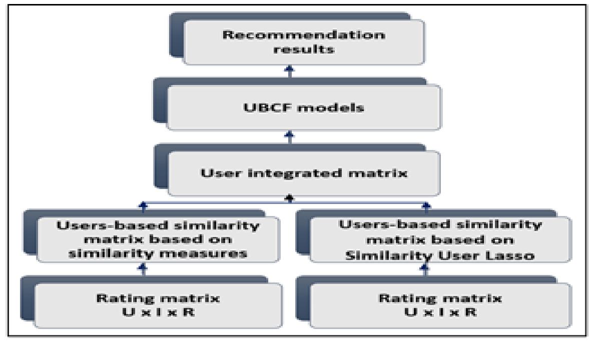


Fig. 3. UBCF-LASSO Model.

Fig. 3 shows the implementation steps of the UBCF-LASSO model. In the first step, the model builds a user similarity matrix from a rating matrix based on the similarity measures. In the second step, the model continues to build the user similarity matrix from the rating matrix based on the LUS. In the third step, the model builds the integration matrix by adding two user similarity matrices from the two steps above. In the last step, the model uses the integration matrix to predict the items to recommend to the user u_a (who needs recommendation).

B. IBCF-LASSO

The IBCF-LASSO model is designed similar to the UBCF-LASSO model with two input parameters: the user's rating matrix for the items: $R_{UxI} = [r_{i,j}]$ with $U = \{u_1, u_2, \dots, u_n\}$ is n users; $I = \{i_1, i_2, \dots, i_m\}$ is m items, and u_a is an user who needs recommendation. However, when building similarity matrices, the model calculates similarity values based on item similarity.

The IBCF-LASSO model has an overall block diagram design structure as follows.

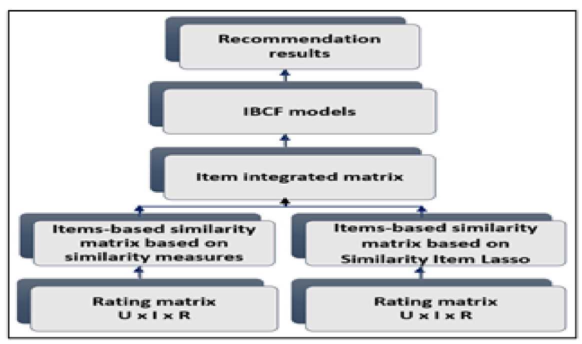


Fig. 4. IBCF-LASSO Model.

Fig. 4 shows the implementation steps of the IBCF-LASSO model. In the first step, the model builds an item similarity matrix from a rating matrix based on the similarity measures. In the second step, the model continues to build the item similarity matrix from the rating matrix based on the Lasso-Item-Similarity (LIS). In the third step, the model builds the integration matrix by adding two item similarity matrices from the two steps above. In the last step, the model uses the integration matrix to predict the items to recommend to the item i_a(item needs recommendation).

VI. EXPERIMENT

A. Datasets

The experimental part is deployed on two popular datasets for research on collaborative filtering models, which are the MovieLense dataset (100k) [3][13] and the Jetter5K dataset (5k sample) [11][13]. These two datasets have been processed and integrated into the recommenderlab package [13].

The MovieLense dataset was collected through the MovieLens website during the seven-month period from September 19th, 1997 through April 22nd, 1998. The dataset contains about 100.000 ratings (1-5) from 943 users on 1664 movies. This dataset is stored in the sparse matrix format of the "realRatingMatrix" class. This matrix is similar in structure to the size of the dataset with rows equal to the number of users, columns equal to the number of movies, and nearly 7 percent of the cells of the matrix have rating values between 1 and 5 (the "null" value is "0").

The Jetter5k is a dataset of the Jester Online Joke Recommendation System collected from April 1999 to May 2003. It contains a sample of 5,000 anonymous users who rated 100 jokes. This dataset is also stored in the sparse matrix format of the "realRatingMatrix" class. However, this dataset has two major differences from the MovieLense dataset. The first is that each user must have a rating for more than 30% of the total jokes. The second is that the rating value for jokes is a real number value between -10.00 and 10.00 (the "null" value is "99").

Both above datasets are randomly selected using the k-fold cross-validation technique (with k=5) [13][17]. This technique requires performing k times to evaluate the proposed models. In each evaluation, the models use one-fold as the testing set and the other $k-1$ folds as the training set. This technique always makes sure that each tuple (row) has at least one occurrence in the testing set. The overall evaluation result of the proposed models is the average result of k times evaluations.

B. Tools

The experiments in this study were performed on the ARQAT tool developed in the R language by our research group [14][16]. In this ARQAT tool, we integrate the recommenderlab [13][16] and the glmnet [6][16] packagses. The recommenderlab is a framework for developing and testing recommendation algorithms while the glmnet fits generalized linear and similar models via penalized maximum likelihood. This ARQAT tool also includes functions for experimental deployment such as: preprocessing of experimental data, calculating similarity matrices, installing recommendation models and methods of evaluating recommendation models.

C. MovieLense

1) Accuracy based on model's predicted values: The method of calculating model accuracy based on predicted values is a method of measuring the deviation between the predicted model's values and the user's actual rating values. In this study, we use three common measures: MSE (Mean Squared Error), RMSE (Root Mean Square Error), MAE (Mean Absolute Error) [5][9][13] to compare the error value of the proposed models with the error value of traditional CFR models.

To ensure the accuracy of the results of comparing the models, we experiment on 4 models: UBCF-LASSO, IBCF-LASSO, UBCF [13], and IBCF [13] with the same training set and testing set.

The results of the comparison of error indexes (RMSE, MSE, MAE) of the four models are presented in TABLE IThis result shows that the model UBCF-LASSO (RMSE: 1.140918; MSE: 1.301695; MAE: 0.892930) has an value of errors is lower than that of the UBCF model (RMSE: 1.146269; MSE: 1.313934; MAE: 0.902714) and the IBCF_LASSO model (RMSE: 1.320553; MSE: 1.743861; MAE: 0.944156) with a lower value of errors than the IBCF model (RMSE: 1.388423; MSE: 1.927719; MAE: 1.050939) on the MovieLense dataset.

2) Accuracy based on model's recommendation results: The method of evaluating the accuracy of the model based on the recommendation results is a method of determining the accuracy of the model by comparing the model's prediction results with the user choices on the testing set. This method uses the confusion matrix to calculate the values of the indicators: Precision, Recall, and F-measure [5][9][13]. Precision and Recall are two commonly used metrics used to evaluate recommender models.

To get the comparison results for this evaluation method, we run all four models: UBCF-LASSO, IBCF-LASSO, UBCF, IBCF on the same training set and testing set.

The results of comparing 3 indexes (Precision, Recall, and F-measure) of the 4 models are detailed in Fig. 5. This result shows that the UBCF-LASSO model has a higher Precision index than the UBCF model (Precision: 0.6 vs. Precision: 0.5) and the IBCF-LASSO model has a higher Precision index than the IBCF model (Precision: 0.5 vs. Precision: 0.4).

3) Comparing Prec/Rec ratio of models: As mentioned in section 2 above, Precision and Recall are two commonly used metrics to evaluate recommender models. However, in some cases when Precision and Recall are inversely proportional to each other, we can use a harmonious combination of Precision and Recall evaluating the overall efficiency of the model. Specifically.

In this study, we build a comparison chart based on the Precision/Recall ratios [13] of four models to better see the performance of the proposed models compared to other published models.

TABLE I. MODEL'S ERROR INDICATORS ON MOVIELENSE

CRF models	**Error indicators**		
	RMSE	***MSE***	***MAE***
UBCF	1.146269	1.313934	0.902714
UBCF-LASSO	1.140918	1.301695	0.892930
IBCF	1.388423	1.927719	1.050939
IBCF-LASSO	1.320553	1.743861	0.944156

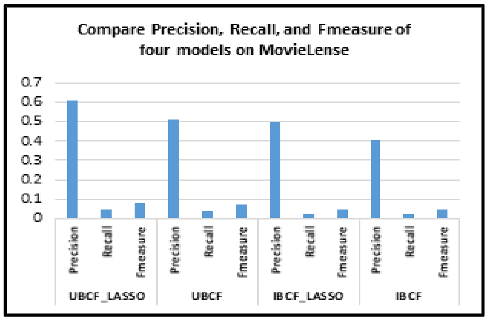


Fig. 5. Comparison Chart of Indicators: Precision, Recall, and F-measure of Four Models on MovieLense.

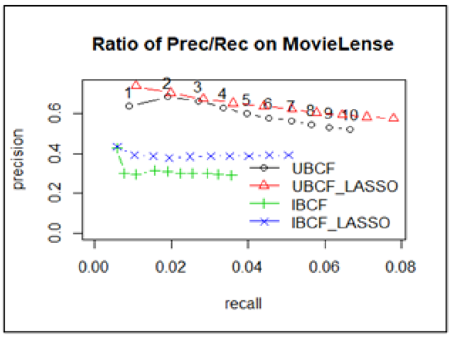


Fig. 6. Comparing the Prec / Rec Ratio of Four Models on MovieLense.

Fig. 6 shows that the UBCF-LASSO model has a better Precision/Recall ratio curve than the UBCF model and the IBCF-LASSO model also has a better Precision/Recall ratio curve than the IBCF model on the MovieLense. The result shows that Lasso regression has increased the accuracy of two models: UBCF-LASSO and IBCF-LASSO. This shows that Lasso regression is suitable to overcome the weakness of the lack of rating data of the CFR models.

D. Jester5K

1) Accuracy based on model's predicted values: In this section, the two proposed models continue to be evaluated for accuracy on the real number ranking dataset. The experimental content deployed on four models is like that deployed with the MovieLense dataset.

From the experimental results, we continue to calculate the error indexes (MSE, RMSE, MAE) of the two proposed models to compare with these indexes of the traditional CFR models.

The results of the comparison of error indexes of the four models are presented in TABLE IIThis result shows that the model UBCF (RMSE: 5.380568; MSE: 28.950510; MAE: 4.243284) has an index of error that is higher than that of the UBCF-LASSO model (RMSE: 5.256894; MSE: 27.634931; MAE: 4.096698). Meanwhile, these indicators on both IBCF-LASSO and IBCF models are almost equal.

2) Accuracy based on model's recommendation results: In this evaluation, we continue to experimentally run four models on the Jester5K dataset and compare the accuracy indexes (Precision, Recall, and F-measure) of two proposed models with the accuracy indexes of two traditional CFR models.

The results of comparing the three above indicators of two proposed models and two traditional CFR models are presented in Fig. 7. This result shows that the accuracy indexes of the two proposed models are higher than those in the traditional CFR models. Especially, the precision value of the UBCF-LASSO model is higher than the precision value of the UBCF model (Precision: 0.898 vs. Precision: 0.698) and the precision value of the IBCF-LASSO model is higher than the precision value of the IBCF model (Precision: 0.647 vs. Precision: 0.447).

3) Comparing Prec/Rec ratio of models: Like the experimental part on the MovieLense dataset, we continue to build a comparison chart based on the Precision/Recall ratios [13] of four models to better see the performance of the proposed models compared to other published models. Fig. 8 shows that the two proposed models both have a higher Precision/Recall ratio curve than the two traditional CFR models. This result again shows that Lasso regression has increased the accuracy of the CFR models on the Jester5K dataset. This once again confirms that Lasso regression helps to deal with the lack of information in the rating data problem of the CFR models.

TABLE II. MODEL'S ERROR INDICATORS ON JESTER5K

CRF models	**Error indicators**		
	RMSE	***MSE***	***MAE***
UBCF	5.380568	28.950510	4.243284
UBCF-LASSO	5.256894	27.634931	4.096698
IBCF	5.844125	34.153799	4.410457
IBCF-LASSO	5.849720	34.219222	4.413152

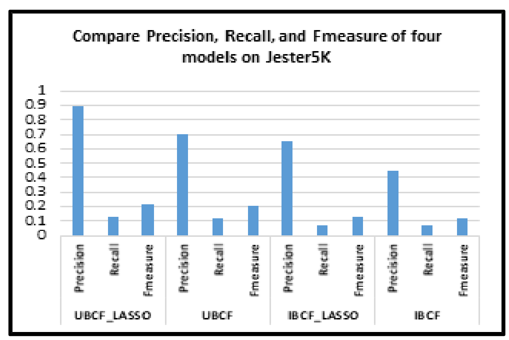


Fig. 7. Comparison Chart of Indicators: Precision, Recall, and Fmeasure of Four Models on Jester5K.

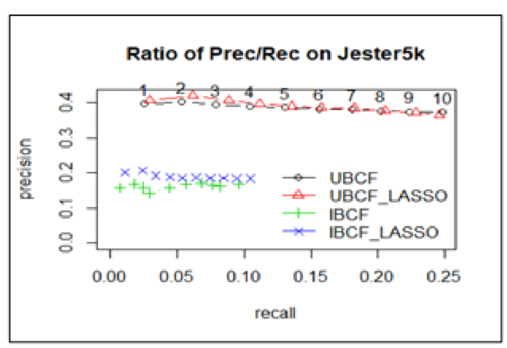


Fig. 8. Comparing the Prec / Rec Ratio of Four Models on Jester5K.

VII. Conclusion

Collaborative filtering model is one of the effective technical solutions to provide customer support on e-Commerce sites. Recommended collaborative filtering models are mainly based on user or product similarity to make recommendations to online customers from rating data. However, these models always face the problem of sparse data on e-commerce sites such as new customers, new products, or too little information about customer reviews of products.

In this approach, an integration matrix between Lasso regression similarity and rating data similarity is constructed in a way that is appropriate to make predictions for new users. Experiments on two popular datasets (MovieLense and Jester5K) suggest that the proposed models provide the recommendation result comparable to significantly better accuracy.

Furthermore, while the accuracy of the traditional CFR models is very dependent on rating data, our model is more accurate than the traditional CFR models even when the number of ratings of users is very small.

References

[1] C. C. Aggarwal, Recommender systems: The textbook, 1st ed., Springer International Publishing Switzerland, 2016.

[2] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, “Recommender systems survey,” Knowledge-Based Systems, vol. 46, 2013, pp. 109–132.

[3] F. M. Harper and J. A. Konstan, “The MovieLens datasets: History and context,” ACM Transactions on Interactive Intelligent Systems, vol. 5, issue 4, 2015, pp. 1-19.

[4] F. T. A. Hussien, A. M. S. Rahma, and H. B. A. Wahab, “Recommendation systems for e-commerce systems an overview,” Journal of Physics: Conference Series, vol. 1897, IOP Publishing, 2021, pp. 1-14.

[5] F. Zhang, T. Gong, V. E. Lee, G. Zhao, C. Rong, and G. Qu, “Fast algorithms to evaluate collaborative filtering recommender systems,” Knowledge-Based Systems, vol. 96, 2016, pp. 96–103.

[6] J. Friedman, T. Hastie, R. Tibshirani, “Regularization paths for generalized linear models via coordinate descent,” Journal of Statistical Software, vol. 33, issue 1, 2010, pp. 1–22.

[7] G. Adomavicius and A. Tuzhilin, “Towards the next generation of recommender systems: A survey of the state-of-the-art and possible extensions,” IEEE Transactions on Knowledge and Data Engineering, vol. 17, issue 6, 2005, pp. 734-749.

[8] G. Linden, B. Smith, and J. York, “Amazon.com recommendations item-to-item collaborative filtering,” IEEE Internet Computing, vol. 7, issue 1, 2003, pp. 76-80.

[9] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, “Evaluating collaborative filtering recommender systems,” ACM Transactions on Information Systems, vol. 22, no. 1, 2004, pp. 5–53.

[10] J. Tang, S. Wang, X. Hu, D. Yin, Y. Bi, Y. Chang, and H. Liu, “Recommendation with social dimensions,” Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16), 2016, pp. 251-257.

[11] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins, Eigentaste, “A constant time collaborative filtering algorithm,” Information Retrieval, vol. 4, 2001, pp. 133-151.

[12] M. P. Robillard, W. Maalej, R. J. Walker, and T. Zimmermann, Recommendation Systems in Software Engineering, Springer, 2014.

[13] M. Hahsler, recommenderlab: A Framework for developing and testing recommendation algorithms, Intelligent Data Analysis Lab, Southern Methodist University (IDA@SMU), 2022.

[14] L. P. Phan, N. Q. Phan, K. M. Nguyen, Huynh, H. X. Huynh, and F. Guillet, “Interestingnesslab: a framework for developing and using objective interestingness measures,” Advances in Intelligent Systems and Computing, vol. 538, Springer, 2017, pp. 302–311.

[15] R. Tibshirani, “Regression Shrinkage and Selection via the Lasso,” Journal of the Royal Statistical Society, Series B (Methodological), vol. 58, no. 1, 1996, pp. 267-288.

[16] G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to statistical learning – with applications in R, 2nd ed., Springer Texts in Statistics, 2021.

[17] R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI), vol. 2, pp. 1137–1143, 1995.

[18] J.S. Breese, D. Heckerman, and C. Kadie, “Empirical analysis of predictive algorithms for collaborative filtering,” Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence, 1998, pp. 43–52.

[19] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning – Data Mining, Inference and Prediction, 2nd ed., Springer Series in Statistics, 2008.

Face Age Estimation using Shortcut Identity Connection of Convolutional Neural Network

Shohel Pramanik, Dr. Hadi Affendy Bin Dahlan
Information Science and Technology
University Kebangsaan Malaysia
Selangor, Malaysia

***Abstract*—Depletion of skin and muscle tone has a considerable impact on the appearance of the face, which is constantly evolving. Algorithms necessitate a large number of aging faces for this purpose. Another popular deep learning technique is convolutional neural networks. In a recent study, many computer vision and pattern recognition problems have been successfully tackled using it. But these methods have architectural issues (e.g., the training process) that have a negative impact on their age estimation performance. As a result, a whole new approach is proposed in this research to address the issue. Using a convolutional neural network framework and resnet50 architecture, researchers were able to detect the age of a human face. This proposed shortcut identity connection strategy, which enables age estimation from the face image, has improved the success of the resnet50 architecture. To be able to tell a person's age from a picture of their face, it was important to know the characteristics of aging. As a result, the rhetorical classification method, which employs the resnet50 structure, is used to shift the face aging levels to a probability level. All of the 50 layers in the proposed residual network have a residual block that connects them. As far as face-aging databases go, ImageNet and FG-NET are both good choices for the proposed age estimation process. In the training session, the experiment results are 2.27 and 2.38, based mostly on the mean absolute error. The test accuracy results for the ImageNet dataset are 81.75% with the FG-NET dataset and 57% with the ImageNet dataset.**

Keywords*—*Shortcut identity connection; ImageNet; residual connection; face aging

I. Introduction

There are numerous real-world applications for age estimation from facial photographs, including security monitoring, missing-person inquiries, entertainment, crime investigation, forensic art, cosmetology, biometrics, and facial recognition. Because of the rising interest in this topic among scientists, the automatic age estimate has attracted a large number of them. It is possible to define an automatic age estimate as either a predefined age range or an actual age.

Multiple ways have been investigated to depict age using facial photos to estimate it. Unfortunately, most of the techniques require a significant number of facial aging datasets and an excessive level of architectural fitness to be implemented. Due to their inaccuracies, these approaches take a long time to train and also lack robustness [12]. To train deep neural networks, researchers have turned to the ResNet framework, which was released last year. The method can employ shortcut identity connections to avoid overfitting in this network's numerous layers. When used for image classification, ResNet is capable of achieving good performance in prediction [1].

The Shortcut Identity Connection of ResNet50, as shown in Fig. 1, is a convolutional neural network is used to estimate an individual's age. The age estimation problem may be solved using a minimal face aging dataset with the proposed approach. Overfitting architectural designs have been solved through shortcut identity connections. In addition, the input data size can be matched to the design using this technique. Different images are trained and tested on various datasets to establish a shortcut identity connection in the ResNet50 architecture. As a final step, the algorithm used the testing dataset and a live image taken by the webcam to estimate age using an estimator. Detailed explanations of the algorithm can be found in the third section.

When deep networks were seen to be converging, a degradation issue arises [1]. As a result of this, the gradients between neurons were completely wiped out [2]. No additional layers or overfitting will solve this gradient problem [1]. Active functions like the sigmoid active function and the tanh active function [3] are to blame for this.

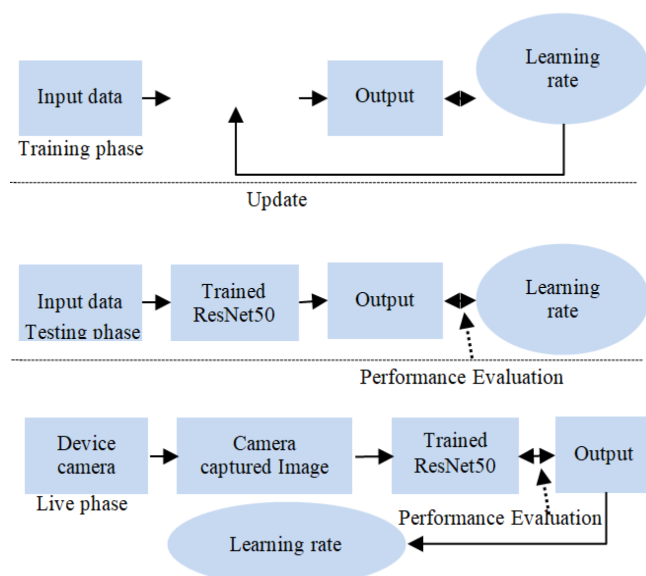


Fig. 1. Proposed Algorithm.

The following are the main points of this manuscript: The proposed shortcut identity connection method demonstrates the high capability of resnet50 by freezing layers from the structure; the presented method can accurately estimate the ages of global and regional people. The proposed age estimation process is immediate and uses an image taken directly by the device webcam.

II. Related Work

A. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is a feed-forward artificial neural network approach [34]. Individual neurons are structured such that they can respond to the overlapping areas of the visual field. In 1968, Hubel and Wiesel [4] studied the cat's visual cortex and discovered that the cortex had a complex arrangement of neurons. In order to cover as much of the subregional visual field as possible, these neurons are placed in a grid-like pattern. It is these neurons that perform the filtering of the input image and work with respect to the natural face photo. Each layer of the CNN design uses a differential function to turn the input volume into the output volume. The architecture of CNN is divided into several separate tiers. Fig. 2 displays the CNN layers.

The work of CNN layers [13] has given below:

The convolution layer serves as the foundation upon which CNN is built, and it is the first step in the learning process. CNN parameters can calculate by multiplying each filter's width and height (w and h), the preceding layer's filter (F), and the current layer's filters (n). $P = (w \times h \times f) \times n$.

During the training session, filters and parameters will learn. While processing the image classification, each filter generates an activation map. In the pooling layer, there are no parameters. The CNN pooling layer does not require any kind of learning. To avoid overfitting during training sessions, it is also beneficial to use the pooling layer. So, the pooling layer p parameter is zero.

A Fully Connected Layer is made up of feed-forward neural networks. The "Fully Connected Layer" is the network's final layer and is referred to as such because of its connectivity. An input is flattened from the final Pooling or Convolutional Layer. A fully connected layer has the most factors to consider when comparing it to other layers. The ideal parameters for a fully linked layer may be found easily because each neuron is connected to every other neuron in the system.

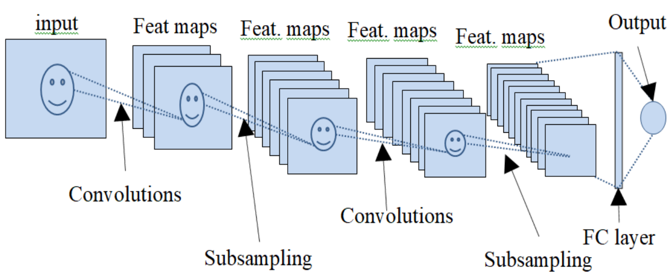


Fig. 2. Neural Networks are constructed using a Common Method known as Convolution. There are Several Separate Layers in CNN Topologies, Including an Input Layer, Convolution Layers, and a Fully Connected Layer.

The generated image is presented in the figure above in the top right corner (Fig. 2), is known as the "Feature Map" [31]. It's the number of confusing features. As a result, the input image information is reduced. When the number of confusing features is 2, the algorithm reduces the image. If the number of confusing features is 4, the algorithm further reduces the image, which makes it easier to process. The question is whether the filter function detector causes information loss. The higher the number in the feature map, the better the filtering process, which implies that the algorithm is not missing many features. Essentially, researchers create all of the feature maps that are required, as well as all of the filters that are required (e.g., edge detection, blur detection, bump detection). Using the Softmax function, the method can assure that all of the estimation will add up to 1. The cross-entropy function and the Softmax function are frequently used interchangeably [32]. After applying the Softmax function in CNN, the next step is to test the reliability of the model using the cross-entropy function to maximize the performance of the neural network [35]. Using the cross-entropy function has several advantages. One of the best features is that if the output value is much smaller than the actual value at the start of the back-propagation, the gradient descent will be very slow. Since Cross-Entropy uses the logarithm, it helps the network evaluate even the largest errors.

B. Residual Network (ResNet)

VGG networks [5] encourage the use of simple residual networks. [5] In order to reduce layer complexity, the number of filters is doubled when the feature map's size is only half-complete [1]. Residual connections were introduced in [1], where they proposed theoretical and practical proof for the use of signal merging for image recognition. This is the original residual connection, and this is an improved version that decreases the computational cost of 11 convolutions (Fig. 3).

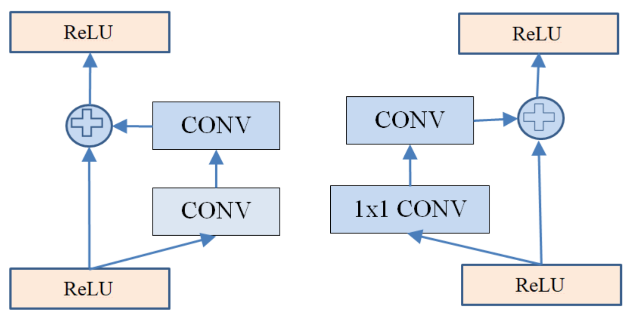


Fig. 3. Residual Connection.

The 50-layer of residual network architecture [1] detail has given below:

It is possible to get a stride size of 2 with a kernel size of 7 * 7 and 64 unique kernels in the first convolution layer. With a stride length of two, which is the maximum pooling layer. As a result of the three levels being repeated three times, the network now contains nine layers after 1*1, 64 kernels, and 3*3, 64 kernels. By the time the procedure reaches the 12th layer, the model has a 3 * 3-layer kernel, a 1 * 1, 128-layer kernel, and a 1 * 1-layer kernel that has been replicated four times in total. There is one more layer in this level, making the total number of layers in the kernel 1*1*1, 1024. There are now a total of 18 tiers to pick from after the procedure's sixth iteration. It now has a total of nine levels of kernels: 3 * 3, 512, and 1 * 1, 2048. In the end, a Softmax function and a 1000-node connected layer reduce the algorithm to a single layer. So, this is a typical pool, no? The activation function and the maximum and the average pooling layer are not calculated. Adding all of the numbers together, they get a convolutional network with 50 layers: 1 + 9 + 12 + 18 + 9 + 1.

Age estimation can be seen as both non-linear and non-category regression problems. In the current age estimation frameworks, the use of images of aging and age estimation modules is common. Shape and texture data from facial photos are commonly used in aging trait techniques. They fall under the categories of anthropometric models, AAM, AGES, age distribution, and appearance models. Regression methods or age-related institutional categories can be used to estimate an individual's age from this point onward. In the most recent research, categorical and regression methodologies are experienced jointly in the most recent research, resulting in a hybrid structure [29]. To create age estimates based solely on facial features, the robust multi-sample regression mastering rule set is also used [30]. Table I presents existing ways for estimating a person's age.

TABLE I. EXISTING AGE ESTIMATION METHODS

Network Architecture	***Publishing Year***	***References***
Multi-purpose CNN	2017	[14]
Diagnosing DLM	2017	[15]
Deep cumulatively and comparatively LM	2017	[16]
AAFs of CNN	2017	[17]
DMTL	2017	[18]
VGG-NET-GPR	2018	[19]
ELM	2018	[20]
DAG-CNN	2018	[21]
The CNN triplet ranking	2019	[22]
DeepAge	2019	[23]
Multitasks-AlexNet	2019	[24]
SADAL & VDAL	2020	[25]
LRN	2020	[26]
CR-MT	2020	[27]
MA-SFV2	2020	[28]

C. Shortcut Identity Connection

Adding more layers in the architecture of a deep neural network allows it to better recognize multiple layers of features in an image. Due to over-exercise, additional layers may not always yield precise results. The problem can be alleviated by establishing a shortcut identity connection between residual blocks. The layer of the neural network where most accuracy will be achieved must be identified [36]. Excessive fitness reduces accuracy, so don't add any more layers after you've found the one with the highest accuracy. The "shortcut identity connection" refers to the identity link between the beginning residual block and the correct stage of a residual block. When the shortcut identity connection is used, all remaining blocks in the architecture will be eliminated [33]. The residual building block [1] is defined as:

$$b = F(a, \{W_i\}) + a, \quad (1)$$

There are two levels here, a is input, and b is output level [37]. The function represents the remaining knowledge. According to Eqn. (1), the size of input a and F must be equal. Use a linear projection if the dimensions of the input image and F don't match up. This issue can solve with a shortcut identity connection if they aren't.

$$b = F(a, \{W_i\}) + W_c a, \quad (2)$$

The experiments in this paper involve varying dimensions of images to estimate age. The shortcut identity connection is effective for the degradation problem, and there is a need to match input dimensions. Age estimation from different dimensional input images can be robust in real-life experience.

III. AGE ESTIMATION METHODOLOGY

The ResNet50 residual network is used in this paper to present a new age estimate approach called "age estimation using the shortcut identity connection of CNN" [33]. Data preprocessing, model creation or choosing a pre-made model, learning rate estimation, and model training are all part of the algorithm's process of determining age estimation. These parts work together to figure out how old you are.

A. Data Preprocessing

This module consists of input data from GitHub (ImageNet) and a loaded data set from the device (FG-NET dataset by Yanwei Fu). The model's expectations for data preprocessing are using ResNet50-selected auto normalization pixel values. Several steps are involved in the data preprocessing modules: acquiring the ImageNet data set, importing all libraries required for the ResNet50 model, identifying missing values, encoding the face aging data, splitting the dataset for the training phase, the testing phase, and the feature scaling.

There are several steps involved in data acquisition before the model may be built and developed. The algorithm needs to create a data set accurately, and then the algorithm uses a shortcut identity connection to match dimensions. A number of graphs and figures were created using the Matplotlib Python program for the experiment. In addition to importing datasets from multiple sources, and data preparation also requires the import of datasets. Setting the data set's current directory was required before import. For the experiment's training and testing phases, the algorithm needed data set splitting to calculate the mean absolute error (MAE), validation error, loss, and validation loss, as well as delete a particular row or column when more than 80% of the values were missing. When dividing a dataset in half, the ratio is 80:20. In the end, the data preparation module applies feature scaling. Within a particular range of values, the algorithm had to summarise several sorts of variables in a dataset. These things are very important in age estimation experiments of any significance: the epochs, loss functions, error functions, and validation functions for each epoch are very important as well.

B. Model Creation

The residual network has 50 layers in this experiment. In brackets, you'll see the residual building blocks (Table II). It takes 50 layers of the residual network to achieve 3.8 billion FLOPs per second [1].

The algorithm used shortcut identity connections for identity mapping to use different dimensional images and set up a freezing process to ignore layers from the model to avoid overfitting. To wrap up, the algorithm presents an alternative CNN design based on shortcut identity connections (Fig. 4).

C. Estimate Learning Rate

During these trials, the algorithm used the range of learning rates to estimate the optimal learning rate for the algorithm and data set under consideration. A maximum learning rate of 0.0001 has been employed in the one-cycle policy [6] (1 e-4]. Algorithm training sessions make use of the term "epoch term." There is a count of how many times the algorithm has been trained. A batch size of 512, ResNet50, and momentum equal to 0.9 was used, and the weight decrease was equivalent to 1e-4 [6]. Since it has a rate of learning close to 1, it can be put to good use. CIFAR10 and ResNet56's training loss, validation accuracy, and validation loss are illustrated in Fig. 5's characteristic plot (a). Fig. 5 shows the age estimation method's log scale of learning rate (b).

D. Training Model

In this experiment, the auto-fit method employed a learning rate schedule with an automatic initial stop on the plateau and a reduction of the maximum learning rate [7]. The auto-fit strategy, when supplied with the fit_onecycle parameter, reduces the learning rate of each cycle using cosine. To estimate age, the algorithm has used the Softmax image classifier with the ResNet50 model pre-trained on ImageNet.

TABLE II. 50 LAYERS OF THE RESIDUAL NETWORK

CNN layer name	*Output size of layer*	*Layer*
Convolution1	112×112	7×7, 64 stride 2
Convolution2_n	56×56	3×3 max pool, stride 2 $\begin{bmatrix} 1\times1,64 \\ 3\times3,64 \\ 1\times1,256 \end{bmatrix}\times3$
Convolution3_n	28×28	$\begin{bmatrix} 1\times1,128 \\ 3\times3,128 \\ 1\times1,512 \end{bmatrix}\times4$
Convolution4_n	14×14	$\begin{bmatrix} 1\times1,256 \\ 3\times3,256 \\ 1\times1,1024 \end{bmatrix}\times6$
Convolution5_n	7×7	$\begin{bmatrix} 1\times1,512 \\ 3\times3,512 \\ 1\times1,2048 \end{bmatrix}\times3$
	1×1	Avg. Pool, 1000-d fc, Softmaxt
Floating point operations per second (FLOPs) [1]		3.8×10^9

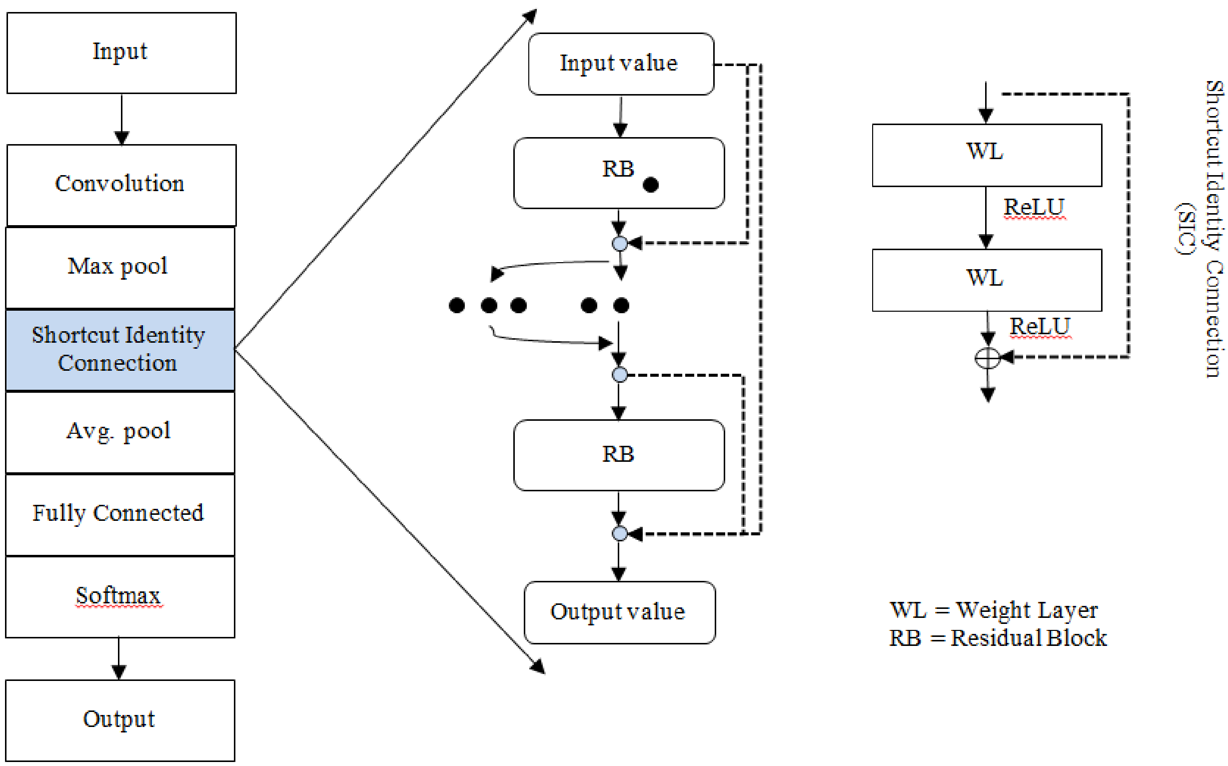
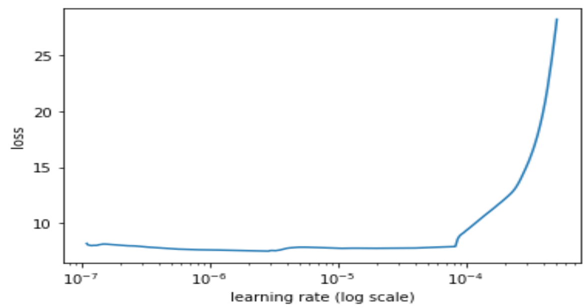


Fig. 4. Proposed an Alternative CNN Architecture using Shortcut Identity Connection.

(a) CIFAR10, ResNet50; Learning Rate Loss.



(b) Age Estimation Approach Learning Rate.

Fig. 5. Learning Rate Log Scale.

IV. Experiments and Results

To obtain these results, the algorithm used the ImageNet and FG-NET datasets together. The experiments use a variety of epochs to evaluate the outcomes. Epoch specifies how many times the algorithm runs during the training process. Each sample of the training data set is only transmitted through the neural network once during a single epoch in a single neural network [33]. One epoch is made up of one or even more batches. The facial photos of Malaysian and Bangladeshi citizens are used to test the suggested approach. Images from the ImageNet and FG-NET databases are shown in Fig. 6. Another experiment was performed on the FG-NET database utilizing the ResNet50 design. To see how the shortcut identity connection worked, the algorithm used images of different sizes and resolutions.

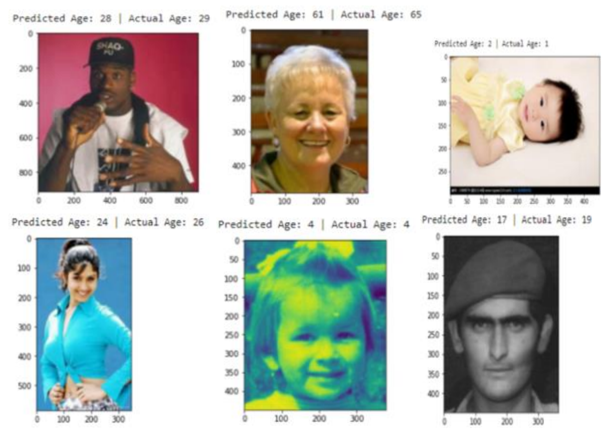
A total of 21690 photos from the ImageNet collection were utilized to train the algorithm, and a further 2411 images that have been verified were used to conduct the testing. To obtain the most accurate results while using the fewest facial photos possible, we have sought to reduce the number of images used. The ImageNet data set can be used to estimate an age with an average accuracy of 81.75 percent and MAE of 2.27 in the training session (Fig. 7). The evaluation method of Mean Absolute Error (MAE) [8] is shown in Eqn. (3).

$$MAE = \sum_{i=1}^{n} |\widehat{l_i} - l_i\,|/N_t, \quad (3)$$

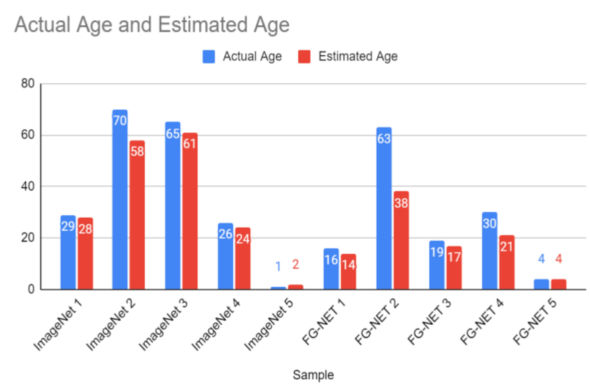
Where: $\widehat{l_i}$ age is recognized, i is a testing sample, and the total number of testing samples is N_t.

Only 352 facial photos from the FG-NET database were used throughout the training period to test how this method worked with a smaller number of facial aging cases. Fig. 8 shows the FG-NET dataset age estimation accuracy [23]. In the training session, MAE for the FG-NET dataset was 2.38. The accuracy rate is 57% across the experiment.

A parameter includes added to estimate age automatically from the camera-captured facial image. The system will automatically turn on the device's camera at the end of the training and testing session of the algorithm. It helps to capture the live facial image to estimate age. After capturing the photo, the estimator will estimate the age. The accuracy of the camera-captured image is 93.75%. A sample has been shown in Table III. Only images from the ImageNet database have been used in training to figure out how old people are from their live webcam captured images, so far.



(a) Sample Facial Images.



(b) Age Estimation from Facial Images.

Fig. 6. (a) Shows Sample Facial Images (b) with an Estimated Age.

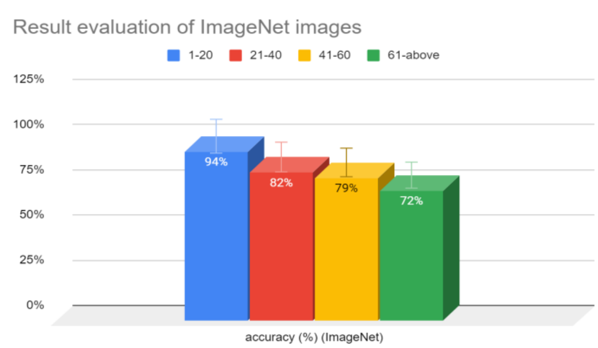


Fig. 7. The Testing Result Evaluation of ImageNet Images.

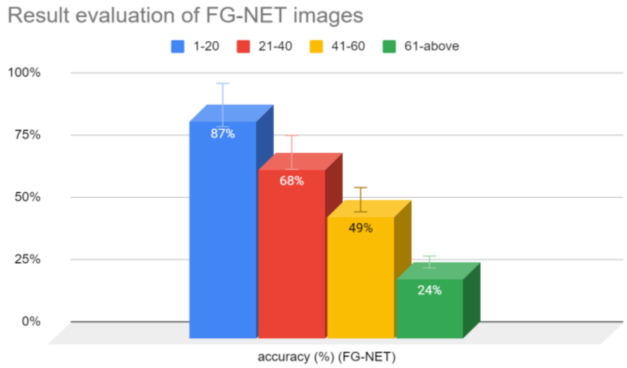


Fig. 8. Testing Results Evaluation of FG-NET Images.

TABLE III. RESULTS OF CAMERA CAPTURED IMAGES

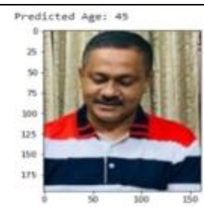
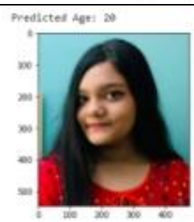
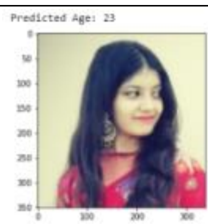
Actual Age	Estimated Age	Facial Images
48	45	
19	20	
26	25	
19	23	
29	29	

Table IV compares the proposed approach's performance to other methods for estimating an age from facial photos and finds that the suggested method outperforms them all. ImageNet testing results have not been found to compare with the proposed algorithm's results.

TABLE IV. THE COMPARISON OF AGE ESTIMATION RESULTS

Methods	Database	MAEs
WAS [9]	FG-NET	8.06
AAS [9]	FG-NET	14.83
KNN [10]	FG-NET	8.24
BP [10]	FG-NET	11.85
SVM [10]	FG-NET	7.25
AGES [9]	FG-NET	6.77
AAM + 2D-DCT [8]	FG-NET	5.39
CMT learning [11]	MORPH-II, FGNET	3.04
AEUSICR (proposed)	ImageNet, FG-NET	2.27, 2.38

V. CONCLUSION AND FUTURE WORK

The proposed ResNet50-based CNN model for estimating age from a facial image has the best performance of any model that's already out there. Use the shortcut identity connection method to come up with a new CNN architectural design. Using different dimensional images and freezing many layers of CNN architecture, this method has been effectively implemented. The image is classified using the softmax classifier, and the positive value is normalized. Improved calculation speed and accuracy are two benefits of shortcut identity connection use in CNN systems. The proposed method achieves MAE 2.27 and MAE 2.38 using the ImageNet and FG-NET datasets. In the future, the algorithm will consider the proposed technique to determine gender and ethnicity and improve the accuracy of the age estimate method by adjusting the ResNet method. Age estimation thus relies on the capacity to classify face images. The proposed method can apply to small and big databases for facial image classification.

ACKNOWLEDGMENT

I have my deepest gratitude to Dr. Hadi Affendy Bin Dahlan, the doctor who supervised me. Patience, time, and encouragement he has given me have been greatly appreciated. He taught me a lot about image processing and computer vision in particular. Hadi Affendy bin Dahlan, in particular, taught me a great deal about doing and writing a research report. I'd want to express my gratitude to the individuals who provided feedback on my review article, which allowed me to make the necessary corrections.

REFERENCES

[1] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778. 2016.

[2] Bengio, Yoshua, Patrice Simard, and Paolo Frasconi. "Learning long-term dependencies with gradient descent is difficult." IEEE transactions on neural networks 5, no. 2 (1994): 157-166.

[3] Szandała, Tomasz. "Review and comparison of commonly used activation functions for deep neural networks." In Bio-inspired neurocomputing, pp. 203-224. Springer, Singapore, 2021.

[4] D. H. Hubel and T. N. Wiesel, "Receptive fifields and functional architecture of monkey striate cortex," The Journal of physiology, vol. 195, no. 1, pp. 215–243, 1968.

[5] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

[6] Smith, Leslie N. "A disciplined approach to neural network hyper-parameters: Part 1--learning rate, batch size, momentum, and weight decay." arXiv preprint arXiv:1803.09820 (2018).

[7] Maiya, Arun S. "ktrain: A low-code library for augmented machine learning." arXiv preprint arXiv:2004.10703 (2020).

[8] Asuman Günay and Vasif V. Nabiyev, "Age Estimation Based on AAM and 2D-DCT Features of Facial Images" International Journal of Advanced Computer Science and Applications (IJACSA), 6(2), 2015. http://dx.doi.org/10.14569/IJACSA.2015.060217.

[9] T. Cootes, G. Edwards and C. Taylor, "Active appearance models", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 6, pp. 681-685, Jun 2001.

[10] A. Lanitis, C. Taylor and T. Cootes, "Toward automatic simulation of aging effects on face images", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 4, pp. 442-455, April 2002.

[11] Yoo, ByungIn, et al. "Deep facial age estimation using conditional multitask learning with weak label expansion." IEEE Signal Processing Letters 25.6 (2018): 808-812.

[12] Maria Luiza C. Passini, Katiusca B. Estébanez, Grazziela P. Figueredo and Nelson F. F. Ebecken, "A Strategy for Training Set Selection in Text Classification Problems" International Journal of Advanced Computer Science and Applications (IJACSA), 4(6), 2013. http://dx.doi.org/10.14569/IJACSA.2013.040608.

[13] Hjelm, R. Devon, Alex Fedorov, Samuel Lavoie-Marchildon, Karan Grewal, Phil Bachman, Adam Trischler, and Yoshua Bengio. "Learning deep representations by mutual information estimation and maximization." arXiv preprint arXiv:1808.06670 (2018).

[14] Ranjan, S. Sankaranarayanan, C. D. Castillo, & R. Chellappa, "An all-in-one convolutional neural network for face analysis," in 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), pp. 17-24, May, 2017.

[15] J. Xing, K. Li, W. Hu, C. Yuan, & H. Ling, "Diagnosing deep learning models for high accuracy age estimation from a single image," in Pattern Recognition, vol. 66, pp. 106-116, 2017.

[16] K. Li, J. Xing, W. Hu, & S. J. Maybank, "D2C: Deep cumulatively and comparatively learning for human age estimation," in Pattern Recognition, vol. 66, pp. 95-105, 2017.

[17] L. Hou, D. Samaras, T. M. Kurc, Y. Gao, & J. H. Saltz, "Convnets with smooth adaptive activation functions for regression," in Proceedings of machine learning research, vol. 54, pp. 430, 2017.

[18] F. Wang, H. Han, S. Shan, & X. Chen, "Deep multitask learning for joint prediction of heterogeneous face attributes," in 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), pp. 173-179, May, 2017.

[19] J. Wan, Z. Tan, Z. Lei, G. Guo, & S. Z. Li, "Auxiliary demographic information assisted age estimation with cascaded structure," IEEE transactions on cybernetics, vol. 48, no. 9, pp. 2531-2541, 2018.

[20] M. Duan, K. Li, & K. Li, "An Ensemble CNN2ELM for Age Estimation," IEEE Transactions On Information Forensics And Security, vol. 13, no. 3, 2018.

[21] S. Taheri, & Ö. Toygar, "On the use of DAG-CNN architecture for age estimation with multi-stage features fusion," Neurocomputing, vol. 329, pp. 300-310, 2019.

[22] W. Im, S. Hong, S. E. Yoon, & H. S. Yang, "Scale-Varying Triplet Ranking with Classification Loss for Facial Age Estimation," in Computer Vision – ACCV 2018. Lecture Notes in Computer Science, vol. 11365. Springer, Cham, 2018.

[23] O. Sendik, and Y. Keller. "DeepAge: Deep Learning of face-based age estimation," in Signal Processing: Image Communication, vol. 78, pp. 368-375, 2019.

[24] Savov, M. Ngo, S. Karaoglu, H. Dibeklioglu, & T. Gevers, "Pose and Expression Robust Age Estimation via 3D Face Reconstruction from a Single Image," in Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 0-0, 2019.

[25] H. Liu, P. Sun, J. Zhang, S. Wu, Z. Yu, & X. Sun. "Similarity-Aware and Variational Deep Adversarial Learning for Robust Facial Age Estimation," IEEE Transactions on Multimedia, 2020.

[26] P. Li, Y. Hu, X. Wu, R. He, & Z. Sun. "Deep label refinement for age estimation," Pattern Recognition, vol. 100, p. 107178, 2020.

[27] N. Liu, F. Zhang, & F. Duan. "Facial Age Estimation Using a Multitask Network Combining Classification and Regression," IEEE Access, vol. 8, pp. 92441-92451, 2020.

[28] X. Liu, Y. Zou, H. Kuang, & X. Ma. "Face Image Age Estimation Based on Data Augmentation and Lightweight Convolutional Neural Network," Symmetry, vol. 12, no. 1, pp. 146, 2020.

[29] Choi, S.E., Lee, Y.J., Lee, S.J., Park, K.R. and Kim, J., 2011. Age estimation using a hierarchical classifier based on global and local facial features. Pattern recognition, 44(6), pp.1262-1281.

[30] Ni, B., Song, Z. and Yan, S., 2011. Web image and video mining towards universal and robust age estimator. IEEE Transactions on Multimedia, 13(6), pp.1217-1229.

[31] Gong, S., Liu, C., Ji, Y., Zhong, B., Li, Y. and Dong, H., 2018. Advanced image and video processing using MATLAB (Vol. 12). Springer.

[32] Myers Jr, L.R., Keller, J.G., Rogers, S.K. and Kabrisky, M., 1998, March. Evolution programs for Bayesian training of neural networks. In Applications and Science of Computational Intelligence (Vol. 3390, pp. 90-98). SPIE.

[33] S. Pramanik and H. Affendy Bin Dahlan, "Age Estimation Using Shortcut Identity Connection of ResNet50 Based on Convolutional Neural Network," 2021 International Conference on Electrical Engineering and Informatics (ICEEI), 2021, pp. 1-7, doi: 10.1109/ICEEI52609.2021.9611146.

[34] Pedrycz, Witold, and Shyi-Ming Chen, eds. Deep learning: Concepts and architectures. Springer International Publishing, 2020.

[35] Al-Mushayt, Omar Saeed. "Automating E-government services with artificial intelligence." IEEE Access 7 (2019): 146821-146829.

[36] Chen, Chun-Fu, Quanfu Fan, Neil Mallinar, Tom Sercu, and Rogerio Feris. "Big-little net: An efficient multi-scale feature representation for visual and speech recognition." arXiv preprint arXiv:1807.03848 (2018).

[37] Jain, Anil K., Arun Ross, and Salil Prabhakar. "An introduction to biometric recognition." IEEE Transactions on circuits and systems for video technology 14, no. 1 (2004): 4-20.

Framework to Deploy Containers using Kubernetes and CI/CD Pipeline

Manish Kumar Abhishek, D. Rajeswara Rao, K. Subrahmanyam
Department of CSE
Koneru Lakshmaiah Education Foundation
Vaddeswaram, India

***Abstract*—Containers are continuously replacing the usage of virtual machines and gaining popularity in terms of scalability and agility in IT Industry. The key concept behind containers is Operating system based virtualization. In cloud, computing containers are getting deployed in terms of computing instances whereas in premises they are getting deployed using Docker as a part of CI/CD pipelines using Jenkin Server. When containers are going to be increased in number, its deployment and resource management is always a concern which is managed using the Kubernetes. Kubernetes is used to deploy and manage the containers in an autonomous manner and Rancher is used to manage the Kubernetes Cluster in an efficient manner. First Analysis is done for the scheduler, resource management which is used by Kubernetes to deploy the containers and proposed a framework which will automate the whole process using the helm-charts, ansible scripts from container deployment to the management of Kubernetes Cluster in a scalable manner. It is fully automated framework and can be used to deploy the scalable applications in form of containers as Docker images. CI/CD pipeline is also considered using Jenkin Server.**

***Keywords*—Containers; Docker; Jenkin; Kubernetes; rancher; virtualization**

I. Introduction

In an agile environment, the application needs to be scalable, performant and highly available based on customer requirements and to achieve the same, containers are widely used in IT industry. The releases are so frequent in terms of delivery in cloud based SaaS (Software as a Service) environment. This is one of the main reasons containers are gaining popularity and used in cloud computing environments and High Performance Computing [1]. Containers are very lightweight in terms of application deployment as one whole package including its required libs, binaries and other dependencies, if any. When an application is split in terms of micro services, every feature is implemented as a micro service and going to be deployed using containers. As a whole Product, the entire application is split in terms of multiple micro services and to have scalability and high availability, every application has a fail back mechanism and to avail the same, each application is deployed with at least two containers. In case one is down another will be available to serve the request or both are used to distribute the workload. In result of this if one product has n numbers of application then need n multiply by two containers. The huge numbers of containers are going to be deployed and then needs to be managed. This requirement of managing high number of containers deployment is always a concern. These containers are provisioned using the Docker images. Docker is an open source platform to bundle the services in form of containers. Using various components, the applications are bundled as Docker images. The required libraries, binaries and other dependencies are defined as a part of configuration in terms of Docker File. Docker file is converted in form of image and then deployed by running commands. Using commands, the image is deployed and the application starts running within container within few seconds. The complete lifecycle of the application revolves around the container lifecycle. Using Docker the state of application is also maintained via commit the container state as an image and tagging it with multiple versions accordingly. If at one point, application crashes, the committed state of container in form of image can be easily deployed again and application will start running with same state when it is committed. This is one of the reason containers provides fault-tolerance and high availability for applications. They are highly trending for application deployment and wherever micro services are getting designed. In comparison of virtual machines, more weightage has been given to the containers in cloud computing as well as in high performance computing. The overhead of application deployment is reduced as it runs on an operating system isolated layer which is portable without use of a hypervisor.

Kubernetes is used for cluster systems to support the container based application deployment. Containers where they are going to be deployed in known as “pod” and it manages thee multiple pod deployment across the physical servers, scaling out the application at run time with multiple workloads. It provides multiple services and tools which are widely available. It is used to avoid downtime of an application. If one container gets stopped or crashed, the another one needs to be up and running in next second. This is the behaviour which is handled easily by Kubernetes. It also offers service registry and load balancing. Multiple containers can reside within a pod to use its file systems and other services belong to a particular pod. The functional or dynamic programming where resource provisioning is so frequent in terms of milliseconds and containers are used, the deployment and its performance need to be monitored. For example: AWS Lambda where multiple user streams are generating the events which are processed by a lambda function. The whole process is executed by deploying a container and billed at 100ms interval of time. The container will be stopped as the function completes its execution. This container deployment, management, monitoring where lambda function is hosted and its performance impacts the provider ability to facilitate the more efficient charging alternatives to the users to process the stream based applications.

Framework is proposed for deploying the containers using Kubernetes based on high performance and fully automated to process the requests which need multiple deployments of containers within few milliseconds. It will identify the required states associated with pods and containers. It can be further used as a configuration to monitor the resources and other details acquired from a Kubernetes Cluster deployment. Using this framework, individuals can plan the capacity support for applications scalability and can do the evaluation of containers and pods which can impact the application performance.

The structure of this paper is as follows: the second section is the analysis and related work; the third section describes the proposed design of framework using Jenkin server and CI/CD pipeline; the fourth section is the evaluation and results; and the fifth section is the conclusion and acknowledgement.

II. Analysis and Related Work

Hardware virtualization and operating system virtualization in terms of virtual machines and containers are always being a research topic from a performance perspective in terms of computing resources such as CPU, memory and storage workloads [2]. In spite of being so much analysis, it is found that many are not familiar or reluctant to use the formal methods. Cloud computing is using a minimal amount of work done on using the formal methods from a performance perspective [3]. Under [4], it is provided as a cyclic design based on particular functional algorithm specifications. Later, it went with the computing resource availability specifications holding the data, control and resources workflow. It was based on Petri Net model [5] capturing the details of functionalities and computing resources requirements involved in the running environment. It starts first with the analyses of the application deployment lifecycle and then understanding of the execution behaviour at run time. It combines the analyses with simulation and predicts the non-functional and functional requirements. Over a time of period, this model is enhanced with the inclusion of performance minimal and maximum boundaries. It allows the competition of resources consumption via formulating the model which is not considered in nude queued networks. The requirement of these models to work is the historical data which need to be feed in form of temporary data. The virtual machine performance has been evaluated in cloud computing environments [6]. Using [7], [8] the containers and virtual machines performance have been evaluated with multiple performance metrics. The few designed have been evaluated in past to manage the containers using Docker and Kubernetes but there is a limitation exists in research area around containers deployment and its management using the Kubernetes architecture. In [9], Containers using Docker performance results into a degradation of network and CPU based negligible performance impact in specified configurations. Kubernetes is not using fully nested-container strategy. It uses the partial one having pod concept where the same IP is used across the containers deployed within that pod. It uses multiple performance metrics for Pod start up and REST API request-response time. Kubemark is used for the Kubernetes Cluster performance evaluation.

Kubernetes is not a traditional platform based system. It operates at container and offers flexibility, monitoring, scaling, load balancing and deployment of containers [10]. There is no limitation for application type with any amount of workload. It is used for containers not for source code deployment. Using its API, required specifications can be declared for the containers which eliminate the requirement of orchestration where steps are executed one by one in sequential order. It is holding a complete independent set of controlled processes which drives continuously the present state to the targeted one. How to reach from one point to another does not matter which make it easy to use extensible and resilient. It is formed using a set of worker machines known as nodes which are used to execute the application in containers and mainly hosts the pods. There is a control panel which is responsible to manager the worker nodes and pods in the Kubernetes Cluster including scheduling, start-up of the new pod, detecting and responding to the triggered events.

Kube-API server is used for the API which is the frontend of control panel and offers horizontal scalability. Fig. 1 shows the high level flow of container's deployment using the Kubernetes. Kube-Scheduler is used to select a node to the newly created pod to run on. For scheduling the node to the pod several factors need to be considered which includes resources requirements, specifications, deadline and infrastructure based policy constraints. There is also kube-controller manager which manages the different type of controllers. For example: Job Controller, Node Controller, Service Account & Token Controller and Endpoint Controller. etcd is used to store the info about the cluster in terms of key-value pair. For cloud based environment, it also offers cloud-controller manager which links the Kubernetes cluster to the cloud based API. Multiple components are also running on the nodes which manage the running pods. Kubelet and kube-proxy are among those components. Kubelet is an agent which runs on the node to make sure that containers are running fine in pod where kube-proxy manages the network rules to make the communication inside and outside of cluster. Containerd is one of the container runtime used by Kubernetes which is mainly holding the responsibility of running the containers [11]-[13]. Kubernetes monitors the container resources via saving the time-series based metrics in a centralised data base. It offers an UI in form Dashboard using which users can monitor the resources and can search on logging i.e. view the logged activity perform against the running application in container. The pod will live till the containers are running which are deployed inside it. Its lifecycle depends on the container lifecycle. The pod required to be waiting till the containers have been created. Using Object Nets [14]-[15] abstraction, pods and containers can be represented as System and Token Nets. To improvise the legibility, the creation part is hidden and apart from this, it is assumed that as long as resources are present, the scheduler is going to allocate a single node to a single pod. If the resource exhausted, the pods will reside in the waiting queue. The node represents the management of resources. For every node, there will be a token which identify the node and its available computing resources.

The allocated resources to a pod will be released based on the policy value i.e. "release" or "failure". Containers creation will get started only when pod is going to be assigned to node. The pod will wait in waiting queue till containers creation will get over. Once all containers get created, pod will be moved to the running state. It will remain in same state till the time containers will not terminate. If any container gets terminated, pod will go in runningFailed state. If container gets restarted without any failure, pod will come back to running state and come to the success state once all containers finished without any failure. Fig. 2 shows the same behaviour. The different transition states are represented from TS1 to TS7. Table I elaborates all the transition states.

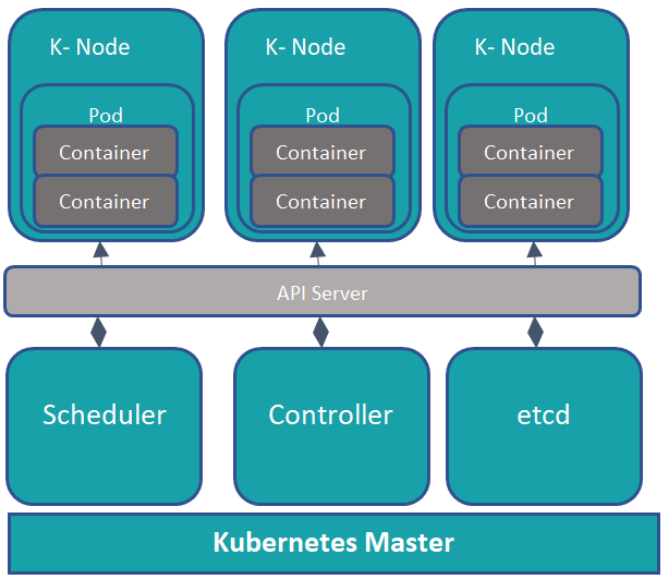


Fig. 1. Container Deployment using Kubernetes.

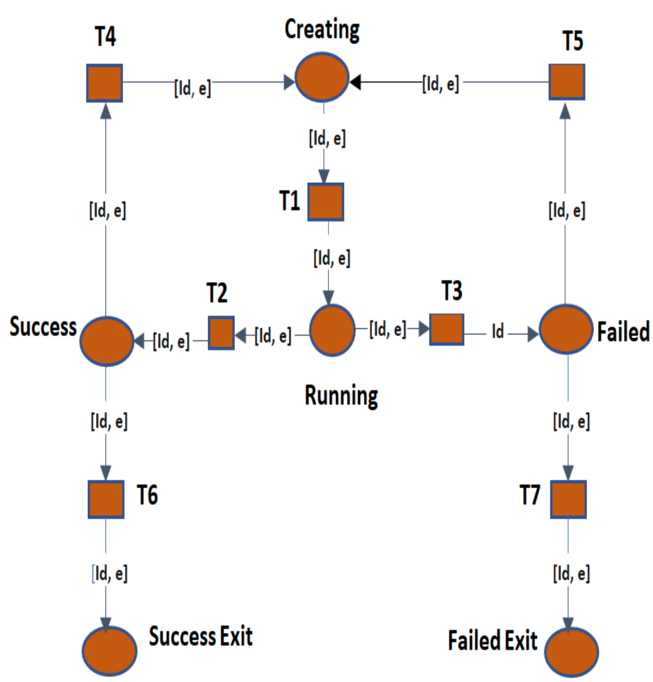


Fig. 2. Transition States Model.

TABLE I. TIME BASED TRANSITION STATES IN THE MODEL

Transition States	**Description**
TS1	Creation time of a conatiner.
TS2	Execution time of a container
TS3	Time until next failure
TS4, TS5	Time taken to restart a container
TS6, TS7	Successful termination of a conatiner

III. PROPOSED DESIGN

Here, the proposed framework is explained to deploy the containers using the Kubernetes cluster which is going to be managed by Rancher. The applications are bundled in form of a single entity as container which will get deployed on a pod. Multiple applications are considered based on different workloads and resource requirements. It consist multiple stages from building the application to its deployment. At first the application will be built and bundled in terms of a jar/war or based on application type. Secondly, Docker file will be created that will hold the multiple instructions, configurations required to execute the application. Once Docker file will be defined, Docker image will be created using the same; it will be tagged based on release version. Using the Jenkin server the application will be built as a part of CI/CD pipeline which will be responsible to take care of whole process from building the application till its execution. Helm Charts have been used in form of YAML files to deploy these container images. In these YAML files, multiple steps have been defined. For example: version, stages from checkout to deployment , repository from where it is needed to checkout the application source code, building the Docker image with name, Docker file, its tag, working directory and deployment details. Ansible scripts have been used to install the helm charts in an automated manner. Whenever the new version of application is available, using CI/CD pipeline; it will be checkout, built, converted into a Docker image and gets installed using the helm-charts via running the Kubernetes commands [16]. Fig. 3 shows the proposed architecture to deploy the containers using Kubernetes followed by CI/CD pipeline.

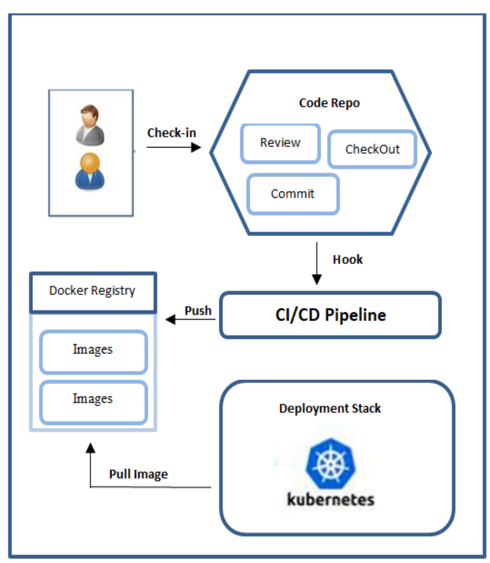


Fig. 3. Proposed Model using Kubernetes followed by CI/CD Pipeline.

A. Jenkin Server

Jenkin Server is used to automate the whole process of check out the application source code, getting compiled, and bundled in appropriate package/build and its conversion to the Docker image and persist it to Docker registry. The main advantage of using the Jenkin server is facilitating the CI/CD pipeline [17] to deploy the containers in form of helm charts using Kubernetes Cluster. It is easily configurable and extendible. For every application, one pipeline is defined with respective set of instructions in Jenkin Server. It also helps us to check the error at run time by using its flexible UI. For Docker images and helm charts, leveraged its plugin architecture and found it really helpful at every stage of pipeline.

B. Continuous Integration and Deployment (CI/CD)Pipeline

Jenkin Server is used for delivering and deploying the application as container in form Continuous Integration and deployment pipeline which is split into multiple stages. Main branch of source code is targeted to pull the source code. In the first stage of CI/CD pipeline [18], whenever the code will be pushed to the respective release branch, it will start building the application in terms of Docker image. The second pipeline will be created for the helm install. It will first setup the infrastructure in terms of database, ElasticSearch, Kafka, Zookeeper if any. After the infrastructure build, it will start pushing the helm packages and charts to respective node. The helm package will be extracted and start executing the YAML files using the ansible scripts. It will run the helm install commands to deploy the containers on pod. As a next stage of pipeline, it will check the health of container. If it is up and running, it will end the pipeline successfully else in case of failure, it will wait for some time as retry else will terminate the container and exists. The Kubernetes Cluster is responsible for the containers deployment but to manage the multiple Kubernetes Cluster health, Rancher Server is used [19]. It provides a Dashboard using which multiple Kubernetes Cluster are monitored and managed.

IV. Evaluation and Result

Evaluation is carried out and benchmarked the overhead during the deployment of containers using Kubernetes with the consideration of following scenarios. (i) Multiple containers i.e. 17 in count are deployed within single pod. (ii) Multiple pods i.e. 4 in count and single container deployment per pod. These 4 pods are deployed on single physical host of 16 cores on Kubernetes node. Mean is represented by symbol (μ_i) and standard deviation is represented by symbol (σ_i). For comparing the results it is equated like N_0 which is the difference both means as $\mu_1 - \mu_2 = 0$ and N_1 is going to be the difference as: $\mu_1 - \mu_2 \neq 0$. TC represents the total number of containers.

For the CPU intensive benchmarking used the pov-ray 3.7 for the measurement of overhead over the pods. Kubernetes allows the containers CPU reservation based on Docker. Sharing of multiple CPUs is directly proportional to the Docker-based reservations. IN scenario (i), as there are multiple containers on a single pod, so it is going to be distributed but in scenario (ii) where single container is deployed per pod, one container can consume the all CPU cores. Table II shows the results for same where execution time is linear. It is found that for the CPU usage, Kubernetes has introduced about 13% overhead. It is concluded that for CPU intensive applications instead of deploying single container per pod, multiple containers deployment on a single pod is recommended. Multiple containers are not resulting into the addition of any type of overhead. If application has the extensive tasks to do at the same time then deployment of application in terms of one replica is not recommendable.

For the I/O intensive, BZip is used for the measurement of overhead. During this experiment, N_0: $\mu_1 - \mu_2 = 0$ is targeted. Table III shows the outcome of measured results of using the BZip for all 17 containers execution time during the compression of the UNIX kernel. It is found that I/O intensive applications are not impacting the deployment. The overhead is almost negligible even the file system is shared across all the deployed containers.

For the network benchmarking, used the iperf server [20] and client deployment on the pods. Server and client are on same physical machine. The containers are deployed in a pod is going to share the IP address in terms of network connection. The TCP based traffic has been monitored by running the tests for about 1 minute. Both the scenarios (i) and (ii) have been considered to find out the impact of network connection. Tables IV and V show the results for both scenarios where single container per pod and multiple container in a pod. It is found that for the running application more than 5 containers. It is concluded that group of few containers on a single pod is better than having higher number of containers deployed within a pod. This number can be fine-tuned based on workloads and application type.

TABLE II. POV-Ray for CPU Intensive based Application

TC	Scenario 1		Scenario 2		N_0?
	μ_1	σ_1	μ_2	σ_2	
1	122..34	0.42	122.23	0.38	Yes
5	467.56	0.94	469.14	0.59	No
9	936.80	0.71	936.58	0.68	Yes
13	1411.66	1.57	1414.30	1.25	No
17	2360.22	1.14	2364.37	3.87	Yes

TABLE III. BZip for Input / Output Intensive based Application

TC	Scenario 1		Scenario 2		N_0?
	μ_1	σ_1	μ_2	σ_2	
1	15.04	0.15	14.97	0.23	Yes
5	15.93	0.14	15.916	0.15	Yes
9	18.19	1.40	18.88	0.53	Yes
13	21.73	1.42	20.33	1.09	Yes
17	35.34	2.67	34.58	0.98	Yes

TABLE IV. N/W BENCHMARKING USING C-PERF CLIENT FOR SCENARIO 1

TC	μ_1(GB)	σ_1	$\sum BW_i/TC$(GB)
1	1.86	0.07	1.86
5	9.62	0.24	2.14
9	16.65	0.11	1.90
13	15.98	0.25	1.26
17	17.87	1.23	2.13

TABLE V. N/W BENCHMARKING USING C-PERF CLIENT FOR SCENARIO 2

TC	μ_1(GB)	σ_1	$\sum BW_i/TC$(GB)	N_0?
1	1.88	0.05	1.88	Yes
5	9.82	0.08	2.19	Yes
9	16.95	0.12	2.04	Yes
13	17.18	0.20	1.16	No
17	18.17	1.35	2.19	No

V. CONCLUSION

For containers deployment, Kubernetes is highly recommendable. Wherever there is a need to provision the high number of computing instances frequently within seconds, the overhead attached to the containers and resource allocation is a limitation. In this paper proposed a flexible, automated and performance based framework that can be used by developers or students in their labs to deploy the containers using Kubernetes. It can be used for any application release, capacity planning and for resource management. It is highly flexible in nature using the helm-charts. In this framework not only the deployment of containers are outlined but also focused on managing the Kubernetes cluster using the Rancher. The life cycle of container is also elaborated and pods internally. The CI/CD pipeline based on Jenkin Server is making this framework fully automated. It is not only offering the fault-tolerance but also support the horizontal scalability of an application in terms of containers. The fully automated framework is elastic in nature without any single manual interruptions.

ACKNOWLEDGMENT

I am really thankful to the Koneru Lakshmaiah Education Foundation for allowing me to choose my research area as per my interest and my guide optimistic nature as well as other staff members who encouraged me to complete this research work.

REFERENCES

[1] Abhishek, Manish. (2020). Containerization for shipping Scientific Workloads in Cloud. International Journal of Advanced Trends in Computer Science and Engineering. 9. 5327. 10.30534/ijatcse/2020/166942020.

[2] Abhishek, Manish. (2020). High Performance Computing using Containers in Cloud. International Journal of Advanced Trends in Computer Science and Engineering. 9. 5686. 10.30534/ijatcse/2020/220942020.

[3] V. Rastogi, C. Niddodi, S. Mohan, and S. Jha, “New directions for container debloating,” in Proceedings of the 2017 Workshop on Forming an Ecosystem Around Software Transformation, ser. FEAST ’17. New York, NY, USA: ACM, November 2017.

[4] Merino, Alberto & Tolosana-Calasanz, Rafael & Bañares, José & Colom, José. (2015). A Specification Language for Performance and Economical Analysis of Short Term Data Intensive Energy Management Services. xxx-yyy. 10.1007/978-3-319-43177-2_10.

[5] T. Murata, "Petri nets: Properties, analysis and applications," in Proceedings of the IEEE, vol. 77, no. 4, pp. 541-580, April 1989, doi: 10.1109/5.24143.

[6] J. Hwang, S. Zeng, F. y. Wu and T. Wood, "A component-based performance comparison of four hypervisors," 2013 IFIP/IEEE International Symposium on Integrated Network Management (IM 2013), 2013, pp. 269-276.

[7] W. Felter, A. Ferreira, R. Rajamony and J. Rubio, "An updated performance comparison of virtual machines and Linux containers," 2015 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS), 2015, pp. 171-172, doi: 10.1109/ISPASS.2015.7095802.

[8] M. Raho, A. Spyridakis, M. Paolino and D. Raho, "KVM, Xen and Docker: A performance analysis for ARM based NFV and cloud computing," 2015 IEEE 3rd Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE), 2015, pp. 1-8, doi: 10.1109/AIEEE.2015.7367280.

[9] M. Amaral, J. Polo, D. Carrera, I. Mohomed, M. Unuvar and M. Steinder, "Performance Evaluation of Microservices Architectures Using Containers," 2015 IEEE 14th International Symposium on Network Computing and Applications, 2015, pp. 27-34, doi: 10.1109/NCA.2015.49.

[10] Abhishek M.K., Rajeswara Rao D. (2022) A Scalable Framework for High-Performance Computing with Cloud. In: Tuba M., Akashe S., Joshi A. (eds) ICT Systems and Sustainability. Lecture Notes in Networks and Systems, vol 321. Springer, Singapore. https://doi.org/10.1007/978-981-16-5987-4_24.

[11] M ondal, S.K., Pan, R., Kabir, H.M.D. et al. Kubernetes in IT administration and serverless computing: An empirical study and research challenges. J Supercomput 78, 2937–2987 (2022). https://doi.org/10.1007/s11227-021-03982-3.

[12] A. Tesliuk, S. Bobkov, V. Ilyin, A. Novikov, A. Poyda and V. Velikhov, "Kubernetes Container Orchestration as a Framework for Flexible and Effective Scientific Data Analysis," 2019 Ivannikov Ispras Open Conference (ISPRAS), 2019, pp. 67-71, doi: 10.1109/ISPRAS47671.2019.00016.

[13] Wei-guo, Zhang & Xi-lin, Ma & Jin-zhong, Zhang. (2018). Research on Kubernetes' Resource Scheduling Scheme. ICCNS 2018: Proceedings of the 8th International Conference on Communication and Network Security. 144-148. 10.1145/3290480.3290507.

[14] Valk, Rüdiger. (2003). Object Petri Nets -- Using the Nets-within-Nets Paradigm. 819-848. 10.1007/b98282.

[15] Kummer, Olaf & Wienberg, Frank & Duvigneau, Michael & Schumacher, Jörn & Köhler-Bußmeier, Michael & Moldt, Daniel & Rölke, Heiko & Valk, Rüdiger. (2004). An Extensible Editor and Simulation Engine for Petri Nets: Renew. Applications and Theory of Petri Nets 2004. 3099. 484-493. 10.1007/978-3-540-27793-4_29.

[16] Spillner, Josef. (2019). Quality Assessment and Improvement of Helm Charts for Kubernetes-Based Cloud Applications.

[17] Moutsatsos, Ioannis & Hossain, Imtiaz & Agarinis, Claudia & Harbinski, Fred & Abraham, Yann & Dobler, Luc & Zhang, Xian & Wilson, Christopher & Jenkins, Jeremy & Holway, Nicholas & Tallarico, John & Parker, Christian. (2016). Jenkins-CI, an Open-Source Continuous Integration System, as a Scientific Data and Image-Processing Platform. Journal of Biomolecular Screening. 22. 1087057116679993. 10.1177/1087057116679993.

[18] S. A. I. B. S. Arachchi and I. Perera, "Continuous Integration and Continuous Delivery Pipeline Automation for Agile Software Project Management," 2018 Moratuwa Engineering Research Conference (MERCon), 2018, pp. 156-161, doi: 10.1109/MERCon.2018.8421965.

[19] Vergara Vargas, Jeisson & Umaña, Henry. (2017). A Model-Driven Deployment Approach for Scaling Distributed Software Architectures on a Cloud Computing Platform.

[20] Tirumala, Ajay & Cottrell, Les & Dunigan, Tom. (2003). Measuring end-to-end bandwidth with Iperf using Web100. 10.2172/813039.

Adaptive Generation-based Approaches of Oversampling using Different Sets of Base and Nearest Neighbor's Instances

Hatem S Y Nabus[1], Aida Ali[2], Shafaatunnur Hassan[3]
Siti Mariyam Shamsuddin[4], Ismail B Mustapha[5]
Department of Computer Science, School of Computing
Faculty of Engineering, University Technology Malaysia
(UTM), Johor, Malaysia

Faisal Saeed[6]
Department of Computing and Data Science
School of Computing and Digital Technology
Birmingham City University
Birmingham, UK

***Abstract*—Standard classification algorithms often face a challenge of learning from imbalanced datasets. While several approaches have been employed in addressing this problem, methods that involve oversampling of minority samples remain more widely used in comparison to algorithmic modifications. Most variants of oversampling are derived from Synthetic Minority Oversampling Technique (SMOTE), which involves generation of synthetic minority samples along a point in the feature space between two minority class instances. The main reasons these variants produce different results lies in (1) the samples they use as initial selection / base samples and the nearest neighbors. (2) Variation in how they handle minority noises. Therefore, this paper presented different combinations of base and nearest neighbor's samples which never used before to monitor their effect in comparison to the standard oversampling techniques. Six methods; three combinations of Only Danger Oversampling (ODO) techniques, and three combinations of Danger Noise Oversampling (DNO) techniques are proposed. The ODO's and DNO's methods use different groups of samples as base and nearest neighbors. While the three ODO's methods do not consider the minority noises, the three DNO's include the minority noises in both the base and neighbor samples. The performances of the proposed methods are compared to that of several standard oversampling algorithms. We present experimental results demonstrating a significant improvement in the recall metric.**

***Keywords*—Class imbalance; nearest neighbors; base samples; initial selection; SMOTE**

I. Introduction

One of the most challenging machine learning problems to both the academia and industry in the last couple of decades is one associated with learning from data that is unbalanced [1]. This problem is known to arise in both binary and multiclass classification tasks when data instances from one class, known as the majority class occur more frequently than instances of other classes, known as the minority classes [2]. This obvious disproportion in the distribution of data instances across classes leans the classifier towards significant bias to the majority class which in turn results in the misclassification of instances of other classes [3]. What makes the class imbalance problem more interesting is the fact that the minority class is often the class of interest in most real-life application domain, thus, the cost of misclassifying the minority class is often higher than that of the majority class [4, 5]. For instance, given a machine learning fraud detection system, legitimate transactions occur more often than fraudulent ones, but the cost of misclassifying a fraudulent transaction as legitimate is greater than the opposite. Therefore, approaches to addressing class imbalance problem are aimed at increasing the accuracy and sensitivity of the classifier to the minority class.

The approaches to dealing with class imbalance problem can broadly be grouped into two categories [6]. The first category entails algorithmic creation/modification to improve learning of the minority class samples. The second category of approaches is the most popularly used category, data level methods, which resamples the data distribution to ensure balanced data distribution across the respective classes via oversampling, under-sampling or their hybrid combination.

This paper focuses on oversampling methods that involve the generation of synthetic data samples to augment the minority class. A leading oversampling method that serves as the basis for most of the recent oversampling methods is the Synthetic Minority Over-sampling Technique (SMOTE) algorithm [2]. SMOTE basically generates artificial samples along the length of the line joining neighboring minority class samples.

SMOTE has also inspired several approaches to counter the issue of class imbalance. It is standard benchmark for learning from imbalanced data [7]. Based on SMOTE, several techniques have been proposed in the literature, and these techniques have been categorized according to some properties include: (1) initial selection of instances to be oversampled (technically called base samples), (2) integration with Under-sampling as step in the technique, (3) type of interpolation, (4) operation with dimensionality changes, (5) adaptive generation of synthetic examples, (6) possibility of relabeling and (7) filtering of noisy generated instances.

Each SMOTE-based extension might have different properties from the aforementioned aspects. However, a large number of them use the three common aspects include: initial selection, type of interpolation (the common type is 'range restricted'), and the adaptive generation of the new samples. This study, therefore, focuses on those three properties.

The most common standard technique that utilizes initial selection and the 'range restricted' interpolation aspects is SMOTE_BORDERLINE [8]. This research, thus, started with adopting the same initial selection of instances to be oversampled in SMOTE-BORDERLINE. The common standard technique that uses adaptive generation of synthetic examples is ADASYN [9], and this is also adopted in this study to be used in our proposed techniques. The minority classes have been classified into three different groups namely safe, danger, and noise; according to its level of difficulty [8, 10-12].

Consequently, six new oversampling techniques, namely, ODO1, ODO2, ODO3, DNO1, DNO2, and DNO3 are proposed. For the ODO techniques, only the borderline examples (Danger group) of the minority class are over-sampled, while in case of the DNO techniques, both the minority danger and minority noise examples are oversampled. The main difference between the three ODO methods lies in the criteria for choosing the nearest neighbors (NN) group. In ODO1, the NN is the minority class except the minority noises, while in ODO2; the NN is the same as the base example which are the borderline examples (minority danger). In ODO3, the NN group includes the whole classes except the minority noises.

Similarly, the main difference between the three DNO methods is the criteria for choosing the nearest neighbors. In DNO1, the NN is the minority class, while in DNO2, the NN are the same as the base examples which consist of the Danger and Noise examples. Lastly, in DNO3, the NN group consists of the whole classes (minority and majority).

Table I shows how each of the proposed methods differs from the standard techniques (SMOTE, Borderline1, Borderline2, and ADASYN). Moreover, in this study, three aspects are added for more clarification about the methods and they are: (1) Nearest Neighbor group, (2) 'how to choose from NN group' and (3) 'noise considered?'

Hence, the major contribution of this study includes the implementation of the proposed methods as well as a tabular overview showing the differences between the methods in details and more clarifications, and this includes the initial selection / base samples used, the NN groups, the method of NN selection, type of interpolation, adaptive generation, and the representation of the minority noises (noises considered?) as shown in Table I. The proposed oversampling techniques were experimentally analyzed using four classification algorithms and evaluation metrics across 15 publicly available datasets from Machine Learning Repositories. The performances of the proposed methods are compared to SMOTE, Borderline SMOTE and ADASYN oversampling methods. In addition, statistical analysis was also carried out using Friedman aligned and Holm's tests.

The organization of this article is as follows. An overview of pertinent studies and oversampling methods is provided in Section II while the procedure of the proposed methods is listed in Section III followed by the experimental design in Section IV. The experimental results and conclusion are respectively presented in Sections V and VI.

II. Related Work

Given that this study focuses on oversampling through synthetic data generation which is a data level approach, a short review of related studies is presented here in this regard. References [7, 13, 14] are important articles for an in-depth review of imbalance resolution approaches. The most basic form of oversampling is known as Random Oversampling which involves random sampling of minority class samples with replacement till it matches the size of the majority class samples. A major drawback of this approach is high likelihood of overfitting that results from the exposure of the classifier to the same information.

An oversampling approach that sidesteps the challenges associated with basic random oversampling is SMOTE which involves synthetic data generation along the length of the line joining neighboring minority class samples. SMOTE generates synthetic samples for any minority class including minority noises which also participate as nearest neighbors. However, when the separation between majority and minority class clusters is not clear, noisy samples may be generated [2]. On the other hand, borderline-SMOTE methods [8] intend to prevent producing noisy samples by detecting the boundary instances between the majority and minority classes, which are then utilized to identify useful informative minority class samples. Although both SMOTE-Borderline1 and SMOTE-Borderline 2 do not generate any sample for minority noises, dealing with those noises as nearest neighbors may generate new samples located near the noises or overlap with them. The study in [9] aims to distribute the new synthetic samples according to the level of difficulties by making the most difficult samples have more new samples. However, this approach results in that minority noises will have the big portion of the new synthetic samples.

From the afore-highlighted, it is obvious that the methods vary in how they deal with the base and nearest neighbor's samples. Similarly, some of them give the minority noises the advantage of being more represented in the new samples while others ignore them completely. However, the use of other different groups is still lacking, therefore, using different sample groups of the base and nearest neighbors are needed.

III. Proposed Methods' Procedure

Suppose that the whole training set is X, the minority class is P and the majority class is N, and $P=\{p_1,p_2,\ldots,p_{num}\}$, $N = (n_1,n_2,\ldots,n_{num})$ Where p_{num} and n_{num} are the number of minority and majority examples. The detailed procedure of ODO1 explained in Fig. 1.

The difference between ODO1, ODO2, and ODO3 is the NN groups as we mentioned above. Additionally, the difference between ODO's techniques and DNO's techniques is that, in DNO's methods, minority noises are added to both base samples and NN samples as declared in Table I. Further, In situations where the NN is from the majority class, a random value between 0 and 0.5 will be multiplied by the difference between the base example and its nearest negative example as in SMOTE_Borderline2 [8].

TABLE I. DIFFERENCES BETWEEN THE OVERSAMPLING METHODS

Method	**Initial Selection/ Base samples**	**NN group**	**How to choose from neighbors**	**Type of interpolation**	**Adaptive generation**	**Noise considered?**
SMOTE	Any from minority	The 5 NN all minority	Randomly	On the line between base and NN New sample= base + (rand(0,1)*diff)	-	Base ➔yes NN ➔yes
Borderline1	Minority_Danger (3,4)	the 5 NN (minority)	Randomly	New sample= base + (rand(0,1)*diff)	-	Base ➔NO NN ➔yes
Borderline2	Minority_Danger (3,4)	the 5 NN (minority + majority)	Randomly	*(range restricted)* *If NN is minority* New sample= base + (rand(0,1)*diff) *If NN is majority* New sample= base + (rand(0,0.5)*diff)	-	Base ➔No NN ➔yes
ADASYN	Minority (1,2,3,4,5)	the 5 NN all minority	Randomly	Weighted distribution New sample= base + (rand(0,1)*diff)	Weighted distribution	Base ➔yes NN ➔yes
ODO1	Minority_Danger (3,4)	the 5 NN (minority-noise)	Randomly	New sample= base + (rand(0,1)*diff)	Weighted distribution	Base ➔No NN ➔No
ODO2	Minority_Danger (3,4)	the 5 NN (minority Danger)	Randomly	New sample= base + (rand(0,1)*diff)	Weighted distribution	Base ➔No NN ➔No
ODO3	Minority_Danger (3,4)	the 5 NN (minority-noise) + majority	Randomly	*(range restricted)* *If NN is minority* New sample= base + (rand(0,1)*diff) *If NN is majority* New sample= base + (rand(0,0.5)*diff)	Weighted distribution	Base ➔No NN ➔No
DNO1	Minority_Danger and Noise (3,4,5)	the 5 NN (minority)	Randomly	New sample= base + (rand(0,1)*diff)	Weighted distribution	Base ➔Yes NN ➔Yes
DNO2	Minority_Danger and Noise (3,4,5)	the 5 NN (minority Danger and Noise)	Randomly	New sample= base + (rand(0,1)*diff)	Weighted distribution	Base ➔Yes NN ➔Yes
DNO3	Minority_Danger and Noise (3,4,5)	the 5 NN (minority + majority)	Randomly	*(range restricted)* *If NN is minority* New sample= base + (rand(0,1)*diff) *If NN is majority* New sample= base + (rand(0,0.5)*diff)	Weighted distribution	Base ➔Yes NN ➔Yes

IV. EXPERIMENTAL DESIGN

The performance of the proposed methods is evaluated using 15 benchmark imbalanced datasets of varying imbalance rations (IR) from the Machine Learning Repositories (UCI, Kaggle, Keel, Datahub) and this is a common practice in class imbalance learning. Table II shows a summary of the 15 datasets. The performances of the proposed oversampling techniques were evaluated and compared with SMOTE, SMOTE_Borderline1, SMOTE_Borderline2, and ADASYN. Since accuracy has been shown in representative works as an insufficient evaluation metric for imbalanced datasets, Recall, and F1-measure are employed in this study. Additionally, the four classifiers considered for evaluation in this study are Decision Trees (DT) [15], Logistic Regression (LR) [16], RandomForest (RF) [17] and Support Vector Machine (SVM) [18].

For each combination of dataset, classifier and evaluation metric, an aligned ranking score is used to rank each oversampling method including the baseline. In addition to the 10 oversampling algorithms considered in this study, the performance of the classifiers on the original dataset without oversampling is also used as the baseline.

Thus, the best performing method has the biggest ranking score while the smallest ranking score indicates the worst performing method. Additionally, two statistical tests, Friedman aligned ranks and Holm, were also used to further establish the significance of our findings. While the Friedman aligned rank's test recognizes the difference in outcomes obtained from many attempts when the normality assumption may not hold, the Holm's test is a nonparametric t-test used to establish whether a control method outperforms comparative methods.

Only Danger Oversampling (ODO1) algorithm:
Step 1. Extract the X_min as the minority samples.
Step 2. Define m_min and m_maj as the number of minority class examples and the number of majority class examples, respectively. Therefore, $m\_min \leq m\_maj$ and $m\_min+m\_maj = X$.
Step 3. Calculate the degree of class imbalance: $d = m\_min/m\_maj$, where $d \in (0, 1]$.
Step 4. Calculate the number of synthetic data examples that need to be generated for the minority class: $G = (m\_maj - m\_min) \times \beta$ Where $\beta \in [0, 1]$ is a parameter used to specify the desired balance level after generation of the synthetic data. $\beta = 1$ means a fully balanced data set is created after the generalization process.
Step 5. Determine the three Minority groups (Noise, Danger, Safe)
Step 6. Now, we find the KNN (K=5) for each example x_i in the danger group in the whole training dataset X.
Step 7. calculate the ratio r_i defined as: $r_i = \Delta_i/K$, $i = 1, ...,X\_d$. X_d is the number of examples in Danger group. where Δ_i is the number of examples in the K nearest neighbors of x_i that belong to the majority class, therefore $r_i \in [0, 1]$.
Step 8. Normalize r_i according to $r_i^{\wedge} = r_i / \sum_{i=1}^{X\_d} ri$, so that $r_i^{\wedge}$ is a density distribution ($\Sigma\, r_i^{\wedge} = 1$)
Step 9. Calculate the number of synthetic data examples that need to be generated for each minority_danger example x_i: $gi = \hat{}r_i \times G$ where G is the total number of synthetic data examples that need to be generated for the minority_danger class.
Step 10. Determine the minority group without noises X_min_no_noise = (X_min)-(Noise)
Step 11. find the KNN (K=5) for each example x_i in the danger group in the X_min_no_noise. In this step, we guarantee that we don't use any minority noise as a NN.
Step 12. For each minority_danger class data example x_i, generate g_i synthetic data examples according to the following steps: Do the Loop from 1 to g_i: (i) Choose a minority data example (x_{zi}) randomly from the nearest neighbors for data x_i. (ii) Generate the synthetic data example: $s_i = x_i + (x_{zi} - x_i) \times \lambda$ where $(x_{zi} - x_i)$ is the difference vector in n dimensional spaces, and λ is a random number: $\lambda \in [0, 1]$. End Loop

Fig. 1. ODO1 Algorithm.

To evaluate the performance of the classifiers on each dataset and method, a stratified k-fold cross validation experimental setup was applied with k = 5. Each oversampling method is performed on only the training portion dataset during k-fold CV and tested on their respective test folds [19]. The presented results represent the means validation performance. When the data you are using to train a machine learning algorithm happens to have the information you are trying to predict that is called Data leakage [20]. Therefore, to prevent leaking the data, the data preparation was performed within cross validation folds.

The hyperparameter tuning of the classifiers was done on the original datasets with no oversampling (baseline) and then the obtained optimal parameters are used when applying the oversampling methods to have fairness with all techniques, while the various oversampling algorithms' hyperparameters were tuned using the default values, except an important parameter in this study that is k nearest neighbor which must be equal to 5 in all oversampling techniques since the proposed methods are built on this number of nearest neighbors. The classifiers and standard oversampling algorithms were implemented using Python modules Scikit-Learn [21] and Imbalanced-Learn [22].

V. Experimental Result and Discussion

At first, in the favor of explaining more about the nature of work of the oversampling standard techniques and the proposed methods, this research visualized their generating of the new samples using a synthetic dataset as you can see in Fig. 2, in addition to the detail description in Table I.

On your imbalanced classification problem, you can choose to use precision or recall. The number of false positive errors will be reduced if precision is maximized, while the number of false negative errors will be reduced if recall is maximized. As a result, precision may be a better fit for classification problems where false positives are a concern. Alternately, recall may be more appropriate on classification problems when false negatives are more important [23]. With dataset such as Breast Cancer, the concern is the recall, therefore, try to reduce the False Negative (FN) as possible as can, while with dataset such as Spam mails dataset, the task will be more focus on precision since it is needed to reduce the False Positive (FP) the most. This study tries to improve the recall without hurting the precision too much.

For each combination of classifier and evaluation metric, the mean rankings of the oversampling approaches over data sets are shown in Table III. The Friedman aligned test is used to statistically confirm the conclusion and the results are shown in Table IV. As a result, the null hypothesis is rejected at a significance level of 0.05., i.e., the oversampling methods do not perform equally in mean rankings for all evaluation metrics.

Table V shows that our proposed method DNO3 is always the first or the second winner with all classifiers when the metric measure is the recall, therefore, DNO3 oversampler is used as a control method in the Holm's test to see if DNO3 result is a significant or not. The adjusted p-values are shown in Table VI.

DNO3 ranked as the best method among all techniques regarding the recall results, and then DNO1 coming as the second. By looking at the differences between the DNO1 and DNO3, the only difference is the NN samples. DNO3 will deal with all classes in the NN whether they are minority or majority class, while DNO1 will only consider the minority class in the NN. This shows the importance of considering both minority and majority classes in the nearest neighbors.

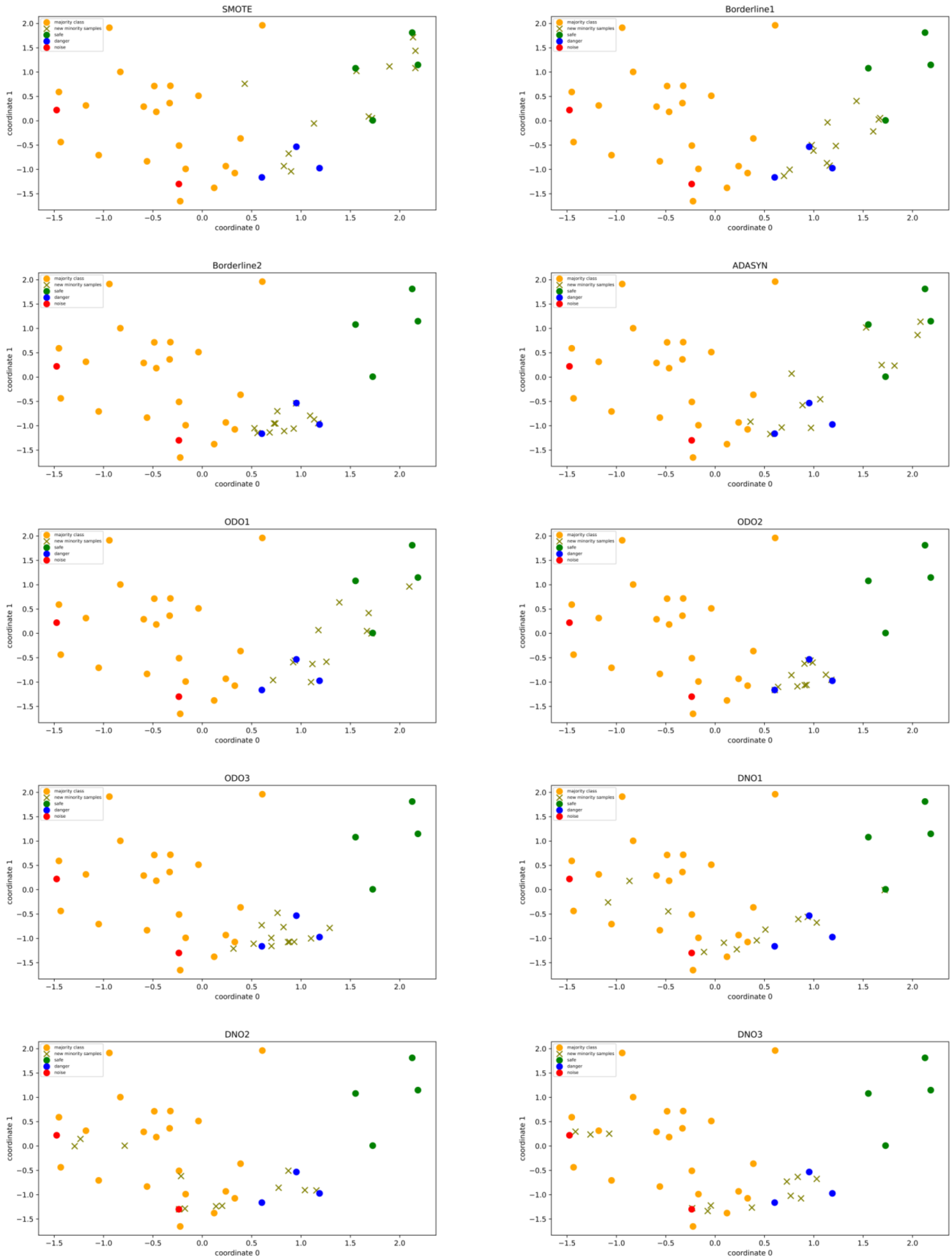


Fig. 2. The Distribution of the New Synthetic Samples using different Oversampling Methods.

TABLE II. DATASETS' DESCRIPTION

Name	**Instances**	**Attributes**	**IR**
abalone-20_vs_8-9-10	1916	8	72.69
Adult	48842	7	3.18
Covertype	38501	55	13.02
pima-indians-diabetes	768	9	1.87
glass4	214	9	15.47
Ionosphere	351	35	1.79
Mammography	11183	7	42.01
oil-spill	937	50	21.85
page-blocks0	5472	10	8.79
Phoneme	5404	6	2.41
poker-8_vs_6	1477	10	85.88
poker-8-9_vs_6	1485	10	58.4
Satimage	6435	37	9.28
Vehicle Silhouttes_0	846	19	3.25
yeast5	1484	8	32.73

TABLE III. RESULTS FOR MEAN RANKING OF THE OVERSAMPLING METHODS ACROSS THE DATASETS. THE BOLD HIGHLIGHTS THE BEST PERFORMING METHOD

Metric	**Baseline**	**SMOTE**	**BL1**	**BL2**	**ADASYN**	**ODO1**	**ODO2**	**ODO3**	**DNO1**	**DNO2**	**DNO3**
Algorithm: DT											
Recall	35.97	89.33	73.37	103.83	87.30	67.73	53.63	65.07	**121.30**	99.67	**115.80**
F1	**124.70**	101.80	96.00	63.43	74.97	**109.00**	92.10	66.03	69.10	61.80	54.07
Algorithm: LR											
Recall	9.80	80.53	80.97	**106.50**	90.40	80.00	58.53	95.83	101.50	96.50	**112.43**
F1	77.73	**114.80**	86.00	71.93	76.17	**111.93**	99.80	81.77	73.63	61.07	58.17
Algorithm: RF											
Recall	18.27	74.97	77.97	96.70	107.47	62.47	66.53	68.83	**119.07**	105.50	**115.23**
F1	105.53	101.47	**108.43**	76.43	71.20	**108.20**	94.97	55.53	75.17	70.67	45.40
Algorithm: SVM											
Recall	19.50	74.70	73.00	82.27	96.73	57.57	77.37	95.87	**106.93**	106.70	**122.37**
F1	86.50	100.67	85.43	54.43	76.93	**101.60**	**105.63**	88.77	81.87	76.47	54.70

TABLE IV. RESULTS FOR FRIEDMAN'S TEST

Metric	P value
Algorithm: DT	
Recall	0.00000
F1	0.00042
Algorithm: LR	
Recall	0.00000
F1	0.00799
Algorithm: RF	
Recall	0.00000
F1	0.00078
Algorithm: SVM	
Recall	0.00000
F1	0.04599

TABLE V. THE WINNING METHODS AMONG ALL METRICS AND CLASSIFIERS

	Recall	F1
DT	**DNO1**	None
LR	**DNO3**	SMOTE
RF	**DNO1**	BL1
SVM	**DNO3**	**ODO2**

TABLE VI. RESULTS FOR HOLMS' TEST. THE BOLD HIGHLIGHTS STATISTICAL SIGNIFICANCE (RECALL – CONTROL METHOD = DNO3)

RECALL							
DT	**adjusted p_values**	**LR**	**adjusted p_values**	**RF**	**adjusted p_values**	**SVM**	**adjusted p_values**
Baseline	**0.00005**	Baseline	**0.00000**	Baseline	**0.00000**	Baseline	**0.00000**
ODO2	**0.00329**	ODO2	**0.01803**	ODO1	**0.02240**	ODO1	**0.00183**
ODO3	**0.02908**	ODO1	0.50404	ODO2	**0.04196**	BL1	**0.03726**
ODO1	**0.04105**	SMOTE	0.50404	ODO3	0.05474	SMOTE	**0.04402**
BL1	0.09000	BL1	0.50404	SMOTE	0.12593	ODO2	0.05936
ADASYN	0.51162	ADASYN	1.00000	BL1	0.16331	BL2	0.10763
SMOTE	0.51693	ODO3	1.00000	BL2	1.00000	ODO3	0.51500
DNO2	1.00000	DNO2	1.00000	DNO2	1.00000	ADASYN	0.51500
BL2	1.00000	DNO1	1.00000	ADASYN	1.00000	DNO2	0.73831
DNO1	1.00000	BL2	1.00000	DNO1	1.00000	DNO1	0.73831

Among the common standard techniques (SMOTE, SMOTE_BORDERLINE1 (BL1), SMOTE_BORDERLINE2 (BL2), and ADASYN), the BL2 is the best in Recall results. Comparing SMOTE_BORDERLINE2's structure with DNO3 shows the importance of considering the minority noise in the base samples since SMOTE_BORDERLINE2 is not considering that, as well as the weighted distribution of the new samples used by DNO3 that creates more new samples for the most difficult samples which is not the way used in SMOTE_BORDERLINE2.

From the above analysis this study depicts that there are three factors can affect the detection of the minority class; the first is that the minority's noises and danger samples which should be considered in the initial selection / base samples, and the second factor is that the minority noises, danger, and also the majority samples should be considered in the nearest neighbors samples, and last but not least is that the distribution of the new synthetic samples should be also weighted distributed so that the more difficult samples will be given more new synthetic samples. These factors can help reducing the false negative (FN) examples and this, in turn, increases the recall.

VI. CONCLUSION

DNO'S techniques performances were the best in Recall, and specifically DNO3 that outperformed all standard techniques in recall metric. This study shows the importance of considering minority noises and danger samples whether as base samples or nearest neighbors' group. Furthermore, the majority class samples should be under concern in the nearest neighbors' group. Finally, the weighted distribution (adaptive generation) of the new samples can help to get better Recall result. Taking everything into account, next work should consider not only the minority danger and minority noise groups, but also different groups of difficult minority samples including the minority safe samples.

REFERENCES

[1] N. Chawla, N. Japkowicz, and A. Kolcz, "Workshop learning from imbalanced data sets II," in Proc. Int'l Conf. Machine Learning, 2003.

[2] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," Journal of artificial intelligence research, vol. 16, pp. 321-357, 2002.

[3] D. Devi and B. Purkayastha, "Redundancy-driven modified Tomek-link based undersampling: A solution to class imbalance," Pattern Recognition Letters, vol. 93, pp. 3-12, 2017.

[4] P. Domingos, "Metacost: A general method for making classifiers cost-sensitive," in Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, 1999, pp. 155-164.

[5] K. M. Ting, "An instance-weighting method to induce cost-sensitive trees," IEEE Transactions on Knowledge and Data Engineering, vol. 14, no. 3, pp. 659-665, 2002.

[6] A. Ali, S. M. Shamsuddin, and A. L. Ralescu, "Classification with class imbalance problem: A Review," Int. J. Advance Soft Compu. Appl, vol. 7, no. 3, 2015.

[7] A. Fernández, S. Garcia, F. Herrera, and N. V. Chawla, "SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary," Journal of artificial intelligence research, vol. 61, pp. 863-905, 2018.

[8] H. Han, W.-Y. Wang, and B.-H. Mao, "Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning," in International conference on intelligent computing, 2005: Springer, pp. 878-887.

[9] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence), 2008: IEEE, pp. 1322-1328.

[10] K. Borowska and J. Stepaniuk, "Imbalanced data classification: A novel re-sampling approach combining versatile improved SMOTE and rough

sets," in IFIP International Conference on Computer Information Systems and Industrial Management, 2016: Springer, pp. 31-42.

[11] K. Borowska and M. Topczewska, "Data preprocessing in the classification of the imbalanced data," Advances in Computer Science Research, 2014.

[12] K. Borowska and M. Topczewska, "New data level approach for imbalanced data classification improvement," in Proceedings of the 9th International Conference on Computer Recognition Systems CORES 2015, 2016: Springer, pp. 283-294.

[13] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches, IEEE T. Syst. Man Cy. C, 42, 463–484," ed, 2012.

[14] N. Chawla, "Data mining for imbalanced datasets: an overview (Periodical style)," Dept. of Computer Science and Engineering, Notre Dame Univ., US, 2005.

[15] J. R. Quinlan, Programs for machine learning. 1993.

[16] P. McCullagh, "Generalized linear models," European Journal of Operational Research, vol. 16, no. 3, pp. 285-292, 1984.

[17] L. Breiman, "Random forests," Machine learning, vol. 45, no. 1, pp. 5-32, 2001.

[18] C.-C. Chang, "" LIBSVM: a library for support vector machines," ACM Transactions on Intelligent Systems and Technology, 2: 27: 1--27: 27, 2011," http://www. csie. ntu. edu. tw/~ cjlin/libsvm, vol. 2, 2011.

[19] M. S. Santos, J. P. Soares, P. H. Abreu, H. Araujo, and J. Santos, "Cross-validation for imbalanced datasets: avoiding overoptimistic and overfitting approaches [research frontier]," ieee ComputatioNal iNtelligeNCe magaziNe, vol. 13, no. 4, pp. 59-76, 2018.

[20] R. Nisbet, J. Elder, and G. D. Miner, Handbook of statistical analysis and data mining applications. Academic press, 2009.

[21] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," the Journal of machine Learning research, vol. 12, pp. 2825-2830, 2011.

[22] G. Lemaıtre, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning," arXiv preprint arXiv:1609.06570, 2016.

[23] J. Brownlee, Imbalanced classification with Python: better metrics, balance skewed classes, cost-sensitive learning. Machine Learning Mastery, 2020.

Social Group Optimization-based New Routing Approach for WMN's

Bhanu Sharma, Amar Singh
School of Computer Applications
Lovely Professional University
Jalandhar, India

***Abstract*—Wireless Mesh Networks (WMNs) are hop-to-hop communication networks that are quickly deployable, dynamically self-organizing, self-configuring, self-healing, self-balancing, and self-aware. In WMNs, a node can leave or join the network at any time. Due to the mobile nature of nodes, the routes between source and destination can change frequently. Computing the shortest path under dynamic conditions for a given time constraint imposed due to node mobility can also be placed in the class of highly complex problems. However, as the network size grows, the performance of the nodes decreases. As a result, we require Soft Computing approaches to handle this problem. This article proposes a Social Group Optimization (SGO) based routing approach to wireless mesh networks. The proposed approach was implemented in MATLAB and tested on different dynamic nodes network scenarios. We compare the performance of the proposed approach with Ant Colony Optimization (ACO), Ad-hoc On-demand Distance Vector (AODV), Dynamic Source Routing (DSR), BAT, Biogeography-based optimization (BBO), and Firefly Algorithm based routing approaches. We observe that the proposed approach outperformed all other approaches on the more than 1000 node network scenarios.**

***Keywords*—ACO; AODV; DSR; BAT; BBO; wireless mesh network; social group optimization**

I. Introduction

A mesh network is a rising era that can provide broadband Internet access, wi-fi nearby area coverage, and network connectivity to operators and customers at a low cost. Wireless Mesh Network is a network that consists of n number of nodes that help in communication within a network structure. Because of its characteristics like self-healing, fault-tolerant and self-organizing, WMNs make the communication process more reliable. There are three types of components in Wireless Mesh Network: Mesh Clients, Wireless Mesh Routers, and Wireless Mesh gateways. Mesh clients, also known as Wireless Mesh clients, are the end-user devices that access the network to run applications like email, games, tracking location, etc. Examples of these clients are mobile, laptops, smartphones, etc. They have limited resources, and sometimes they might not be connected to the network. Client WMNs consists of routers with no cabling connecting the nodes and can work using inbuilt radios in network nodes. Unlike traditional wireless access points, every node consists of inbuilt hardware gateways to connect it with neighbor nodes.

Routing in WMNs is a challenging issue due to the dynamic nature of networks. In WMNs, any node can move anywhere without informing other nodes. Due to the mobile nature of nodes, the routes are frequently changing. Computing the shortest path under dynamic conditions for a given time constraint imposed due to node mobility can also be placed in the class of highly complex problems. So, for efficient communication, there is a need for routing approaches that can discover the new routes quickly before the network structure changes again. It is a well-known fact that the probability of finding the best solution is significantly less for the highly nonlinear and complex problems falling under the class of NP-hard or NP-complete problems. The computational cost for such cases may be so high that the best solutions may not be affordable. Under such circumstances, wherever best can be replaced with reasonable solutions, soft computing approaches perform better than their hard computing counterparts. Thus, suggesting that soft computing-based approaches to dynamic shortest path route evaluation shall provide better results than the hard computing-based approaches such as AODV, DSR, etc. The hard computing-based algorithms work well on WMNs with less than 1000 nodes. This article proposes a new soft computing-based routing approach for WMNs. The proposed approach works upon the social learning behavior of humans. The proposed algorithm worked very well on large nodes network scenarios.

We divide our paper into 5 sections. Section 1 of the paper presents the motivation for the paper. Section 2 discusses the overview of the SGO (Social Group Optimization) Algorithm. Section 3 presents the SGO based approach for shortest route evaluation in WMNs. Section 4 discusses the implementation and performance analysis of the proposed approach for WMNs. Section 5 concludes the paper.

II. Social Group Optimization Algorithm

Social Group optimization group is the nature-inspired computing approach that is based on the human decision-making process. This algorithm is based on peer learning and increasing an individual's knowledge. The SGO algorithm works in two phases: The Improving Phase and Acquiring Phase. Here each person in the group is considered as a candidate solution. Each candidate tends to acquire knowledge and possess to increase fitness value. On the other hand, the one with the highest expertise or the best person tries to distribute the knowledge among all the candidate solutions, resulting in upgrading the knowledge level of the entire group members.

A. Improving Phase

In the improving Phase, each candidate solution communicates/interacts with the best person present in the group. As a result, an individual's knowledge shall upgrade with the help of the one with the highest fitness value. The pictorial representation of this Phase can be seen in Fig. 1.

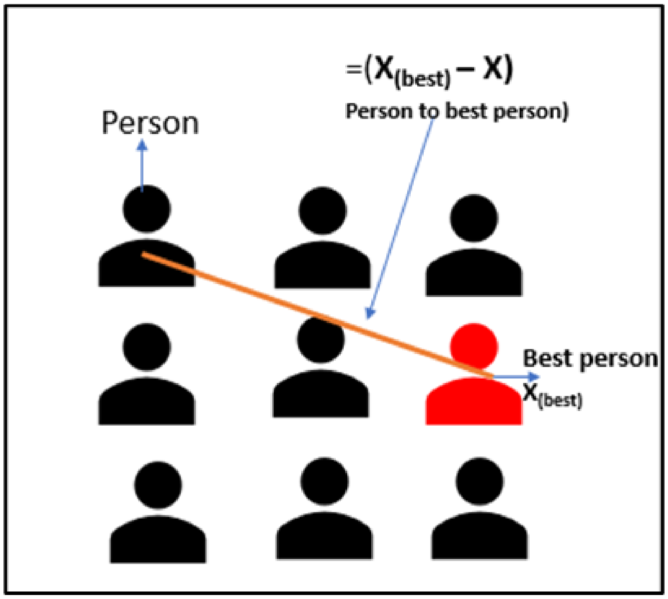


Fig. 1. Shows the Improving Phase Working.

$$X_{new} = cX + r(X_{best} - X) \quad (1)$$

The new solution computed with c, a self-introspection parameter, is here X $_{new}$, r is the random number between (0,1), and X is the previous solution. X $_{best}$ represents the one with the highest fitness value. With (X $_{bes}$t, X), the individual interacts with the best person in the group, which helps them enhance their expertise. The focus will be on the acquisition phase once the new solution has been produced. This entire process works in two steps: Improving and acquiring Phases. Previously the improving Phase has been discussed. Once the communication between the partner and the best person occurs, the algorithm moves to the next phase, i.e., acquiring.

B. Acquiring Phase

Now this Phase works in two steps:

Step 1: Mutual interaction with the other partners in the group.

Step 2: Mutual Interaction with the other people in the group + considering the best person from Phase 1 (Fig. 2).

In the case of Step 1, each person interacts with other people in the group by using the following Eq 2:

$$Person\ to\ partner\ or\ peer\ = X - X_p \quad (2)$$

Where X is the person and X_p are their colleagues. In Step 2, communication will be done with the best person in the group. Here each member interacts with the best person only, which is calculated using the following Eq calculates. 3.

$$Person\ to\ Best\ person = (X_{best} - X) \quad (3)$$

Here X_{best} represents the person with the highest fitness value or with the best knowledge. Once the acquiring completes, each person's ability in the group is updated using the following Eq. 4.

$$X_{new} = X + r1(X - X_p) + r2\ (X_{best} - X) \quad (4)$$

With the help of these two steps, a new solution is generated by X_p(partner solution). The obtained solution could be used to solve different types of optimized problems. Here if the knowledge of the person X is greater than the partner X_p. The objective function for Maximization is given by equation 5. Similarly, if the knowledge of the person X is less than the partner X_p , then the objective function is given by following (Eq. 6)

$$X_{new} = X + r1(X - X_p) + r2\ (X_{best} - X)$$
$$if\ X > X_p \quad (5)$$
$$X_{new} = X - r1(X - X_p) + r2\ (X_{best} - X)$$
$$if\ \ X < X_p \quad (6)$$

and for the same condition, the objective function for Minimization is shown below in (Eq. 7 and Eq. 8)

$$X_{new} = X + r1(X - X_p) + r2\ (X_{best} - X)$$
$$if\ \ X < X_p \quad (7)$$
$$X_{new} = X - r1(X - X_p) + r2\ (X_{best} - X)$$
$$if\ \ X > X_p \quad (8)$$

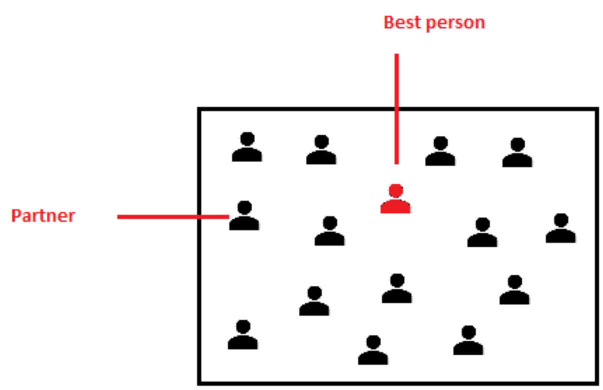


Fig. 2. SGO with the Best Person and the Partner.

III. Social Group Optimization-Based Dynamic Shortest Path Route Evaluation in Wireless Mesh Network

This section proposes an SGO based approach for least-cost route evaluation in WMNs. Routing is the process of establishing a communication link between a source and a destination node. The cheapest path in WMNs is challenging to determine due to the dynamic nature of network nodes. Routing techniques for WMNs are often divided into two phases. The first Phase is the route discovery phase. The source nodes should discover the best route within the specific time frame. Once the route is discovered, data communication is performed in the second Phase. In the second phase, source and destination nodes can communicate. Data transfer continues until a specified time interval or network structure changes.

When estimating a minimal or near-cost route, these dynamic scenarios and time constraints complicate the routing process. In most cases, the exact shortest path cannot be calculated, so the shortest path must be replaced with the least cost path so that the path can be served as an input to the existing dynamic environment [1]. Due to the complexity of this in Wireless Mesh Network, A soft computing approach is preferred over hard computing. Soft computing techniques provide an optimized route instead of the best route within a given time constraint.

In WMN's routing, matrices play a crucial role in path selection and route optimization. As defined in the literature, ETX [2] is termed as the "expected no of transmission nodes," which are required when transmitting the information from a source node to a destination node. To calculate ETX, each node spreads an inquiry packet which is having the number of received inquiries from every neighbor. The Route EXT sums up all the ETX links which come in between the route. With the help of source routing and ETX/ETT metrics, the Local On-Demand Link State (LOLS)[3] protocol executes the route-discovery process. WCETT was proposed [4] to minimize the number of nodes on the route of a flow that transmits information on the same channel. It is a combination of end-to-end delay and channel diversity. The MR-LQSR (Multi-Radio Link Quality Source Routing Protocol) [5] follows LQSR to work over multiple channels and interfaces with the help of the WCETT metric. ETT (Expected Transmission Time) [6] tackles the problem of low performance presented in ETX by considering the differences in link transmission rates. ETT adjusts ETX to different PHY rates and data-packet sizes. Apart from all these, we have some more metrics like Per-Hop Packet Pair Delay (PktPair)[7], Expected Transmission on a Path (ETOP) [8], Effective Number of Transmission (ENT)[9], and Modified Expected Number of Transmissions (mETX) [10], Metric of Interference and Channel Switching (MIC) [11], Bottleneck Link Capacity (BLC) path metric [12], cross-layer link quality and congestion aware (LQCA) metric [13]. A novel interference aware low overhead routing metric was proposed [14]. In our proposed approach we use the fuzzy-based integrated link cost (ILC) method [15].

Some more interesting study on Soft computed based WMN's can be seen in [16]-[20].

Integrated link cost (ILC) = f(throughput, delay, jitter, node residual energy)

Proposed SGO Approach in Routing in Wireless Mesh Network for Dynamic Optimal Cost Route Evaluation

Begin
/* SGO starts
Initialize the SGO parameter. Generate N populations each with randomly generated NC candidate Solutions, Dimension D of each population, termination criteria self-introspection parameter c.
while! = T C **do**
/* TC is a termination criterion */
for i = 1: N **do**
Calculate ILC of every link of the network.
Using ILC to evaluate the fitness calculate the fitness value of each person in the population
Sort the ith populations from best to worst based on the values of ILC;
The best-fit individual of the ith population is chosen as the gbest of the person;
End **for**
Record the best route from amongst all available routes. The best fit is called elite;
Initiate an improving Phase to update the knowledge of persons with the help of the best route
for i = 1: N **do**
Update the i^{th} route by integrating it with the best route in populations.
End for
%Acquiring Phase
Select a route 'r' randomly from the available population of routes
Compute fitness of 'r'
for i = 1: N **do**
if r is better than the ith route
Update the i^{th} route by integrating it with the 'r' route in populations.
End for
Compute fitness of the entire route population.
Calculate the best fit route from the current population and call it temp_fit
Update elite with temp_fit (if required)
Stop

IV. Architectural Details of Various Client WMN Scenario

For implementation purposes, we consider different client WMN scenarios. The architectural detail of each scenario is shown in Table I.

TABLE I. NETWORK ARCHITECTURE OF CLIENT MESH NETWORK

No. of Nodes Area	Area(m×m)	Radio Range	Timing Constraint (in Seconds)
500	1000 *1000	250	1.3,1.5,1.7,2.1,2.3,2.5,2.7,3.1,3.3,3.5
1000	1000*1000	250	1.2,1.4,1.6,1.8,2.0,2.2,2.4,2.6,2.8,3.0
2000	2000*2000	250	2.0,2.2,2.4,2.6,2.8,3.0,3.2,3.4,3.6,3.8
3000	3000*3000	250	4.0,4.5,5.0,5.5,6.0,6.5,7.0,7.5,8.0,8.5
4000	4000*4000	250	5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, 9.0, 9.5, 10.0

V. Result and Discussion

To assess the appropriateness of the suggested SGO-based approach to route evaluation in WMNs, we implemented it in MATLAB along with 6 other algorithms and ran simulations using dynamic scenarios. We evaluated 500, 1000, 2000,3000, and 4000 node client WMN designs for simulation purposes. For simplicity, we have eliminated the findings from WMNs with up to 1000 nodes because the proposed SGO technique performs better on larger networks. Table I shows the architectural design of several Client WMN node scenarios.

A. *Comparative Performance of 2000 Node Client Wireless Mesh Networks*

The performance of the Social Group Optimization-based new routing approach for the Wireless Mesh Network approach is evaluated along with other 6 algorithms on the 2000 node client WMN scenario. We have considered 2.0, 2.2,2.4,2.6,2.8,3.0,3.2,3.4,3.6, and 3.8 seconds timing constrain for performance analysis purpose. For each timing constrain, we conducted 10 trials. For testing purposes, we conducted a total number of 100 trials.

The performance of each approach with different time intervals on the 2000 nodes scenario is presented in Table II. For all time limits and given trials, the results achieved by SGO based routing approach outperformed ACO, DSR, AODV, BAT, BBO, and Firefly algorithms 56 times out of 100 trials. The optimal route generated by Firefly is 32 times and BAT 8 times. BBO generated better results 4 times in some time constraints including equal optimal route cost with BAT, Firefly, and SGO. With the given timing constraints, no route was discovered by ACO, DSR, and AODV. For the time constraints 2.0, 2.6 3.4, and 3.8, BAT and BBO were able to discover routes but were unable to produce optimal route cost. For the time constraints 2.4,3.0, and 3.2 BAT, BBO, and Firefly generated similar route costs. SGO came out as a winning approach with the highest route discovery and minimal route cost in the entire analysis. Fig. 3 shows the pictorial representation of the best performance of SGO with the other 6 algorithms.

TABLE II. ARCHITECTURAL DETAILS OF 2000 NODES CLIENT WMN SCENARIOS

Time Constraints										
Algo	**2.0**	**2.2**	**2.4**	**2.6**	**2.8**	**3.0**	**3.2**	**3.4**	**3.6**	**3.8**
ACO	------	------	------	------	------	------	------	------	------	------
AODV	------	------	------	------	------	------	------	------	------	------
DSR	------	------	------	------	------	------	------	------	------	------
BBO	0	0	1+A	0	0	1+A	1+A	0	1+A	0
BAT	0	2	1+A	0	2	1+A	1+A	0	1+A	0
Firefly	4	3	2	3	4	1+A	2	4	5	4
SGO	6	5	6	7	4	7	6	6	3	6
A=1 represents equal results generated										

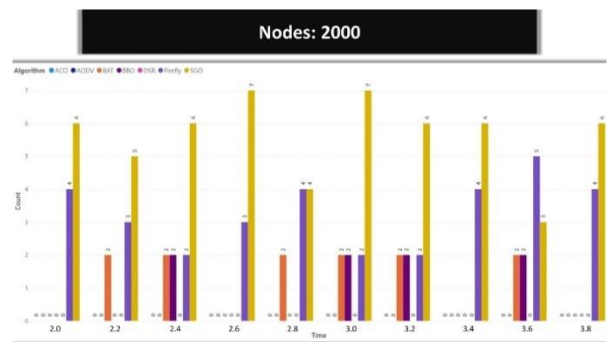


Fig. 3. Graphical Presentation of 2000 Nodes in Client Wireless Mesh Networks.

B. *Comparative Performance of 3000 Node Client Wireless Mesh Networks*

From Table III, it is observed that for time constraints of 4.0 seconds BAT and BBO produce minimum route cost and equal results for one time each, Firefly-based approach produced 3 times and SGO generated minimum cost route 5 times. Within time constraint 4.5 BBO and BAT and SGO produced enumerated equal best paths. For 5.0 seconds BAT, BBO and Firefly produced equal results once and generated minimal route path 1 time.

SGO on the other hand generated a minimal route cost path 7 times. For time constraints 6.0, 7.0,8.0, and 8.5 BBO and BAT-based approaches successfully discovered the route but did not produce the shortest path. For the same time constraints, the shortest path discovered by Firefly is 13 times and SGO is 22 times in total. Further, we observe that the overall performance of SGO based routing approach is much better than the other 6 approaches for each timing constraint. For a total of 100 trials, SGO produced a minimum cost path 55 times followed by Firefly 34 times, BAT 6 times, and BBO 5 times. For the given timing constraint all the algorithms except for AODV, ACO, and DSR have produced results. Fig. 4 represents the histogram of the frequency of best performance for each timing constraint.

TABLE III. ARCHITECTURAL DETAILS OF 3000 NODES CLIENT WMN SCENARIOS

Time Constraints										
Algo	**4.0**	**4.5**	**5.0**	**5.5**	**6.0**	**6.5**	**7.0**	**7.5**	**8.0**	**8.5**
ACO	------	------	------	------	------	------	------	------	------	------
AODV	------	------	------	------	------	------	------	------	------	------
DSR	------	------	------	------	------	------	------	------	------	------
BBO	1+A	1+A	1+A	1	0	0	0	1+A	0	0
BAT	1+A	1+A	1+A	1	0	1	0	1+A	0	0
Firefly	3	2	1+A	5+A	7	3	4	2	2	5
SGO	5	6+A	7	3+A	3	6	6	6	8	5
A=1 represents equal results generated										

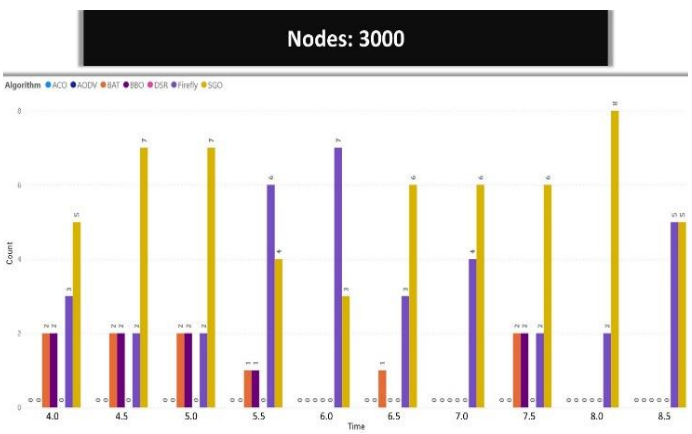


Fig. 4. Graphical Presentation of 3000 Nodes in Client Wireless Mesh Networks.

C. Comparative Performance of 4000 Node Client Wireless Mesh Networks

Table IV shows the architectural design of 4000 nodes of Client Wireless Mesh Network along with 6 algorithms in given time constraints of 5.0,5.5,6.0,6.5,7.0,7.5,8.0,8.5,9.0,9.5, and 10.0 seconds. During implementation and results analysis, SGO outperformed ACO, DSR, AODV, BAT, BBO, and Firefly. ACO, DSR, and AODV failed to identify the path as the number of nodes increased, suggesting that these two approaches cannot perform for bigger networks. For the time constraints 5.0,6.5,8.0 and 9.5 BBO, BAT and Firefly have produced equal minimal route cost paths for once. SGO in the same time constraints produced the best results 30 times. For a time constraint of 5.5 seconds, BBO and BAT were able to discover a route but did not produce the shortest path. On the other hand, the optimal path produced by SGO is 8 times and Firefly 2 times for the same time intervals i.e. 5.5 seconds. For time constraints 6.5 and 7.0 seconds SGO discovered route 9 times, firefly 1 time each. BAT and BBO, on the other hand, were able to discover a route but not an optimal one. For a total of 100 trials, SGO outperforms 76 times, followed by Firefly 28 times, BAT 4 times, and BBO 1 time excluding the same route discovery cost. It illustrates how, as SGO is implemented in a larger network with greater time constraints, its performance improves, yielding the optimal shortest path.

Fig. 5 shows the graphical representation of the simulation results implemented using MATLAB.

TABLE IV. ARCHITECTURAL DETAILS OF 4000 NODES CLIENT WMN SCENARIOS

Time Constraints											
Algo	**5.0**	**5.5**	**6.0**	**6.5**	**7.0**	**7.5**	**8.0**	**8.5**	**9.0**	**9.5**	**10.0**
ACO	------	------	------	------	------	------	------	------	------	------	
AODV	------	------	------	------	------	------	------	------	------	------	
DSR	------	------	------	------	------	------	------	------	------	------	
BBO	A	0	0	A	0	0	1+A	0	0	A	A
BAT	1+A	0	0	A	0	0	1+A	0	0	1+A	1+A
Firefly	1+A	2	4	1+A	1	6	2+A	3	3	1+A	4
SGO	8	8	6	9	9	4	6	7	7	7+A	5
A=1, represents equal results generated											

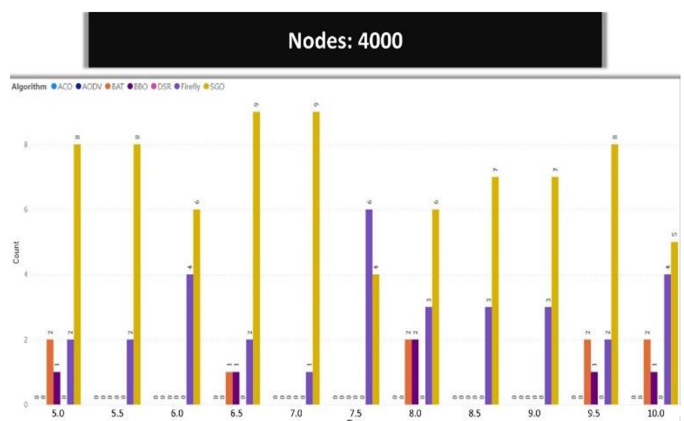


Fig. 5. Graphical Presentation of 4000 Nodes in Client Wireless Mesh Networks.

VI. CONCLUSION

Wireless Mesh Network is dynamic because nodes are free to move in any direction. As the node's changes, the structure of the network also changes along with them. Due to network changes, there comes a high probability of route variation between source and destination. This route variation causes a loss of information between nodes and however it is difficult to get the optimized results in less time. Thus, there is a need for a routing approach that can discover the optimum route quickly even if there is any fault in a specific path. To overcome this problem some soft computing-based approaches can be integrated into Wireless Mesh networks which can help to get optimal results within the given time constraints. This article proposes a new SGO based new Routing approach for wireless mesh networks. The proposed approach was tested and implemented in MATLAB on different network scenarios. The performance of the proposed approach was compared with 6 other algorithms (ACO, AODV, DSR, BAT, and Firefly) on the bases to calculate a minimal cost path. The result analysis shows that SGO based routing approach outperforms the other 6 algorithms on network architecture greater than 1000 nodes. Thus, with all the computational results and analysis, we conclude that SGO based routing approach is the best-suited dynamic near shortest path approach amongst all 6 algorithms discussed.

Acknowledgment

I would like to express my very great appreciation to Dr Amar Singh for their valuable and constructive suggestion during the planning and development of the research work. This paper and research behind it would not have been possible without their guidance. His enthusiasm and knowledge have always been an inspiration to me. I wish to acknowledge the help and support provided by Lovely Professional University and my colleagues.

References

[1] S. Sharma, S. Kumar, and B. Singh, "Hybrid Intelligent Routing in Wireless Mesh Networks: Soft Computing Based Approaches," *Int. J. Intell. Syst. Appl.*, 2014, DOI: 10.5815/ijisa.2014.01.06.

[2] D. S. J. De Couto, D. Aguayo, J. Bicket, and R. Morris, "A high-throughput path metric for multi-hop wireless routing," in *Wireless Networks*, 2005, vol. 11, no. 4, pp. 419–434, DOI: 10.1007/s11276-005-1766-z.

[3] X. Xu, C. Dong, and A. Liu, "Optimization of load balancing routing algorithm based on extended localized link states in low earth orbit satellite networks," *Int. J. Satell. Commun. Netw.*, 2021, DOI: 10.1002/sat.1403.

[4] A. P. Subramanian, M. M. Buddhikot, and S. Miller, "Interference aware routing in multi-radio wireless mesh networks," in *2006 2nd IEEE Workshop on Wireless Mesh Networks, WiMESH 2006*, 2006, pp. 55–63, DOI: 10.1109/WIMESH.2006.288620.

[5] Z. Che-Aron, A. H. Abdalla, and K. Abdullah, "The performance evaluation of AODV-based and DSR-based multi-radio routing protocols in cognitive radio Ad Hoc network," *Res. J. Appl. Sci. Eng. Technol.*, 2013, DOI: 10.19026/rjaset.6.3944.

[6] N. El Haoudar and A. Maach, "Routing metric for Wireless Mesh Networks," in *2nd International Conference on Innovative Computing Technology, INTECH 2012*, 2012, pp. 57–62, DOI: 10.1109/INTECH.2012.6457746.

[7] T. Kimura and S. Kamei, "QoS evaluation of diffserv-aware constraint-based routing schemes for multi-protocol label switching networks," 2004, doi: 10.1016/S0140-3664(03)00210-X.

[8] G. Jakllari, S. Eidenbenz, N. Hengartner, S. V. Krishnamurthy, and M. Faloutsos, "Link positions matter: A noncommutative routing metric for wireless mesh networks," *IEEE Trans. Mob. Comput.*, 2012, doi: 10.1109/TMC.2011.79.

[9] V. Gupta and R. Pandey, "An improved energy aware distributed unequal clustering protocol for heterogeneous wireless sensor networks," *Eng. Sci. Technol. an Int. J.*, 2016, doi: 10.1016/j.jestch.2015.12.015.

[10] R. Murugeswari, S. Radhakrishnan, and D. Devaraj, "A multi-objective evolutionary algorithm based QoS routing in wireless mesh networks," *Appl. Soft Comput. J.*, 2016, DOI: 10.1016/j.asoc.2015.12.007.

[11] M. Boushaba, A. Hafid, A. Belbekkouche, and M. Gendreau, "Reinforcement learning based routing in wireless mesh networks," *Wirel. Networks*, 2013, DOI: 10.1007/s11276-013-0592-y.

[12] H. Haile, K. J. Grinnemo, S. Ferlin, P. Hurtig, and A. Brunstrom, "End-to-end congestion control approaches for high throughput and low delay in 4G/5G cellular networks," *Computer Networks*. 2021, DOI: 10.1016/j.comnet.2020.107692.

[13] V. K. Sharma, L. P. Verma, and M. Kumar, "A fuzzy-based adaptive energy-efficient load distribution scheme in ad-hoc networks," *Int. J. Intell. Syst. Appl.*, 2018, DOI: 10.5815/ijisa.2018.02.07.

[14] L. Ma, Q. Zhang, and X. Cheng, "A power controlled interference aware routing protocol for dense multi-hop wireless networks," *Wirel. Networks*, 2008, DOI: 10.1007/s11276-006-9233-z.

[15] S. Sharma, S. Kumar, and B. Singh, "Routing in Wireless Mesh Networks: Three New Nature-Inspired Approaches," *Wirel. Pers. Commun.*, 2015, DOI: 10.1007/s11277-015-2588-7.

[16] Shakti Kumar, Amar Singh, Sukhbir Singh Walia, “Parallel Big Bang - Big Crunch Global Optimization Algorithm: Performance and its Applications to routing in WMNs ”, Wireless Personal Communications, Springer, 100(4), 2018, pp. 1601- 1618.

[17] Amar Singh, Shakti Kumar, Ajay Singh , Sukhbir Singh Walia, “Three-parent GA: A Global Optimization Algorithm”, Journal of Multiple-Valued Logic and Soft Computing, Volume 32, 2019, pp. 407 – 423.

[18] Amar Singh, Shakti Kumar, Sukhbir S. Walia, P3PGA: Multi-population 3 Parent Genetic Algorithm and its Application to Routing In WMNs, Implementations and Applications of Machine Learning, Springer, Springer, PP. 1-28, 2020.

[19] Amar Singh, Shakti Kumar, Sukhbir Singh Walia, “FW-AODV : An Optimized AODV Routing Protocol for Wireless Mesh Networks”, International Journal of Advanced Research in Computer Science, Volume, Volume 8, No. 3, March – April 2017, pp. 1131-1135.

[20] Bhanu Sharma, Amar Singh, “Routing Protocols for Wireless Mesh Networks: A Survey”, Advances and Applications in Mathematical Sciences, Vol. 18, Issue 8, June 2019, pp. 605-616.

A New Combination Approach to CPU Scheduling based on Priority and Round-Robin Algorithms for Assigning a Priority to a Process and Eliminating Starvation

Hussain Mohammad Abu-Dalbouh
Department of Computer Science, College of Science and Arts, Qassim University, Unaizah, Saudi Arabia

***Abstract*—The main purpose of an operating system is to control a group of processes, through a method known as CPU scheduling. The performance and efficiency of multitasking operating systems are determined by the use of a CPU scheduling algorithm. Round-robin scheduling is the best solution for time-shared systems, but it is not ideal for real-time systems as it causes more context shifts, longer wait times, and slower turnaround times. Its performance is mostly determined by the time quantum. Processes cannot have priorities set for them. Round-robin scheduling does not give more critical work greater consideration, which may affect system performance in solving processes. On the other hand, a priority algorithm can resolve processes' priority levels. This means that each process has a priority assigned to it, and processes with highest priority are executed first. If which process should come first and the process waiting time in CPU are not considered, this can cause a starvation problem. In this paper, a new CPU scheduling algorithm called the mix PI-RR algorithm was developed. The proposed algorithm is based on a combination of round-robin (RR) and priority-based (PI) scheduling algorithms for determining which tasks run and which should be waiting. The disadvantages of both round-robin and priority CPU scheduling algorithms are addressed by this novel algorithm. When using the proposed mix PI-RR algorithm, the performance measures indicated improved CPU scheduling. Other processes should not be affected by the CPU's requirements. This algorithm helps the CPU to overcome some of the problems of both algorithms.**

Keywords*—Average turnaround time; average waiting time; utilization; performance measures; operating system; process

I. Introduction

People can shop, learn, arrange appointments, play games, and more due to technological advancements such as mobile phones and computers. Because humans are typically unable to utilize and maintain these devices due to their complexity, operating systems have emerged to address these issues [1], [2]. They are best described as a link between the user and the computer hardware that makes managing and controlling the computer system easier. Both the user and the system benefit from the services provided by operating systems [3]. On the user side, they provide user interfaces and assist in the implementation of programs, file management, and information exchange with other computers, while on the system side, they allow multiple users to share resources and protect system resources [4], [5].

In a multitasking context, CPU scheduling is a critical duty for an operating system. A ready queue is maintained when more than one procedure needs to be executed. In a two-processor system, each processor has its own ready queue. The operating system chooses a process from a list of those in the ready queue, and assigns the CPU to it based on an algorithm [6], [7]. To ensure fairness and avoid hunger while allocating CPU to processes, close attention is essential. When making scheduling decisions, the aim is to keep the average waiting time, average turnaround time, and number of context flips as low as possible.

The operating system is in charge of managing the computer's hardware and software resources, as well as performing many functions. Processor scheduling is regarded as a fundamental task. All resources are scheduled before they are used, so they are available to processes when they are needed and at a new stage in the process life cycle [8], [9]. A short-term scheduler (STS) [10] selects a process from the ready queue for implementation, and scheduling is the essential function of the operating system in a computer system. These algorithms are used to schedule tasks in the CPU; each one outperforms the others in some performance metrics, and has its own set of benefits and drawbacks [11].

The job of a CPU scheduler is to select a process from a memory list of ready-to-run processes. In the following situations, the CPU scheduling choice for a scheduler must be made:

- Switch a process from running to ready state.
- Switch a process from waiting to ready state.
- Send a process to terminate state.

The success of the scheduler is decided by an algorithm. High-quality CPU scheduling algorithms rely on maximize usage rate, throughput, turnaround time, waiting time, and response time. In multi-processing systems, the user executes multiple applications at the same time, each of which contains multiple processes that require the CPU to complete its responsibilities, but only one process can acquire the CPU at a time. As a result, CPU scheduling is required, which allows one function to use the CPU while another waits for other resources, improving management reliability and efficiency [12]. One of the most significant components of the device is the CPU. Because most operations rely on it, we must maximize its usage and throughput, while reducing turnaround time, waiting time, and response time. CPU scheduling techniques, which control how processes enter the CPU, can meet all of these requirements [13]. There are numerous scheduling algorithms, each of which is implemented in a unique way. The FCFS algorithm, for example, assigns the CPU to the first person who arrives. The SJF algorithm allocates the CPU to the shortest task. The round-robin algorithm assigns a time quantum to each process, calculates its working time in the CPU, then leaves the process and permits another to run. According to their priority, the priority algorithm [8] determines which processes are allowed to access the CPU. Many issues might arise during the execution of scheduling algorithms [14].

This paper proposes a mix priority and round-robin algorithm (mix PI-RR algorithm) for assigning a priority to a process and eliminating starvation. This algorithm has the optimal advantages of both priority and round-robin algorithms. The rest of this paper is organized as follows: Section II presents the literature review. In Section III, the proposed algorithm is discussed and a flowchart is presented. In Section IV, present the discussion and conclusion in the final section.

II. Literature Review

To discover the best CPU algorithm for a given procedure, it is possible to compare the three CPU algorithms based on their waiting times. Each algorithm has been extensively tested and the outcomes compared. In [15], the researchers developed an improved round-robin scheduling approach based on the clustering algorithm, which combined the advantages of prioritizing short operations with low round-robin scheduling overheads to reduce the average waiting time and turnaround time. Using the k means technique, similar processes were clustered. These researchers employed the CPU scheduling approach in [16] to create a fast system with fewer resources. They were able to improve the algorithm's efficiency and reduce its runtime. The various scheduling algorithms were designed and implemented by them. In [17], the researchers recommended scheduling techniques to increase the operating system's real-time performance.

In [18], the round-robin scheduling algorithm's time quantum concerns were addressed. The researchers developed the smart job first dynamic round-robin technique. Using a dynamic time-quantum technique, the program required the CPU schedule to sort processes in ascending order based on burst time, assign system priority, and calculate a smart priority factor (SPF) for each process. The team created a simulator to evaluate the proposed algorithm.

In [19], the researchers released an enhanced version of the Fittest Job First Dynamic Round-Robin algorithm (FJFDRR), which incorporates the process arrival time as an algorithmic element that various queues handle. The suggested approach was compared against current scheduling algorithms in four test cases using the ATAT, AWT, AR, and CS metrics. Based on the number of processes provided, the statistics demonstrated that the suggested technique had the best appropriate context switch rate.

The researchers presented the Enhanced Round-Robin (ERR) algorithm in [20], which aims to increase CPU performance by reducing the average waiting time and turnaround time. In three separate scenarios, the suggested algorithm was compared to the RR and IRR algorithms. The findings revealed that this approach performed better by reducing the average WT and average TAT.

The researchers suggested the Modified Priority Preemptive Scheduling Technique as a novel CPU scheduling algorithm in [21]. Priority pre-scheduling is implemented in a cyclical manner by the algorithm. The results indicated that the novel technique handled the starving problem, while also improving the speed of the standard preemptive algorithm.

In [22], the researchers proposed an approach for managing loads and prioritizing selection of tasks.

The SJF algorithm must be used to rank jobs, followed by the RR algorithm for execution. The findings demonstrated that the ad hoc algorithm prioritizes higher priority jobs and executes them rapidly, while contextual switching is reduced for low priority processes, reducing the options between RR and SJF.

In the Cyber-Physical System, the researcher presented a scheduling strategy for high-priority random jobs [23]. A fog group is used in idle time to process the most recent available time and execution time before assigning the system to a random high-priority task. This method speeds up the dispatch of high-priority random jobs, allowing them to be performed more quickly.

A number of CPU scheduling algorithms have been developed in recent years to ensure predictable processor allocation. Often, the best features of each algorithm have been combined to create the ideal algorithm for a given situation. The upgraded round-robin (IRR) CPU scheduling algorithm, invented by Mishra, is an improved round-robin scheduler. It is comparable to round-robin (RR), but is a little better [24]. IRR chooses the first process from the ready queue and gives it the CPU for up to one QT. When a process completes its QT, it checks the remaining CPU burst time of the presently executing process.

III. Proposed Algorithm

The round-robin algorithm does not consider the importance and significance of processes; it simply solves the queue, so the process order that the CPU performs causes a decrease in the CPU efficiency. The priority algorithm has some disadvantages because it solves the process priority for which process come first, so it performs the process with the highest priority and does not consider the arrival time or size of the process. It also ignores CPU bursts that can cause starvation. Therefore, this study created a mix of the two previous algorithms to decrease the disadvantages for the operating system environment. Mix PI-RR algorithm and flowchart of the proposed Mix PI-RR algorithm is shown in “Fig. 1 and 2”.

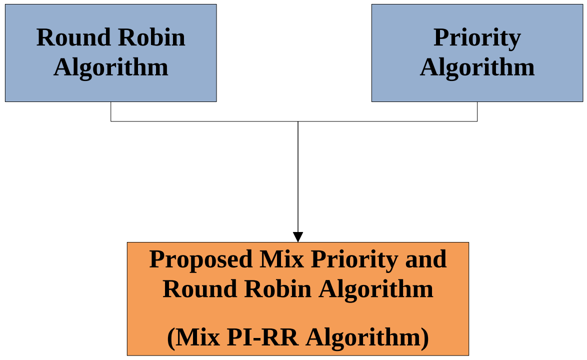


Fig. 1. Mix PI-RR Algorithm.

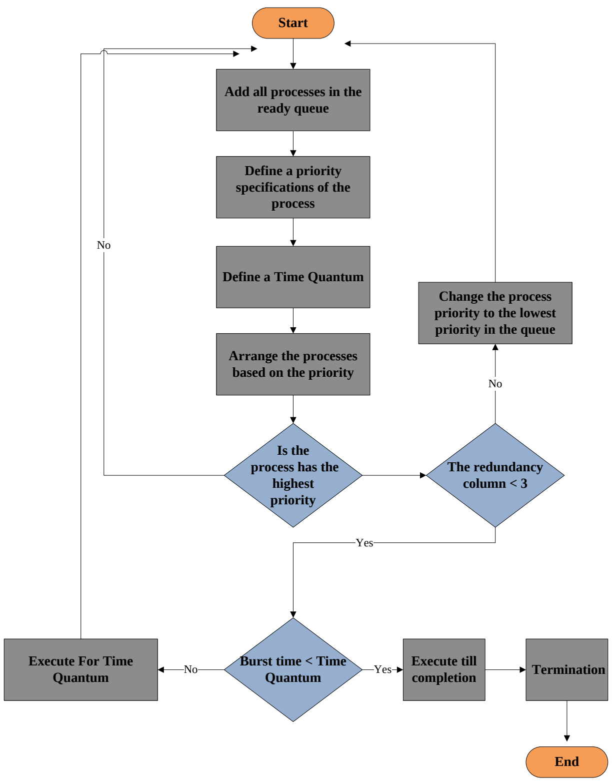


Fig. 2. Flowchart of the Proposed Mix PI-RR Algorithm.

The proposed algorithm adds two more columns to the original columns in the round-robin as follows:

Priority column: The CPU gives some processes higher priority than others. The number of times to repeat this is listed in the second column. If the procedure is prioritized three times then, in the central processing unit, the original priority is updated and the lowest priority is assigned. Following this, the processes will be carried out in accordance with the arrival time of the central processing unit (CPU). Finally, when the CPU receives the lowest priority process, the priority will be activated again. Then, it will be implemented three times before being assigned the lowest priority in the central processing unit (CPU).

NOTE 1: If more than one process had the same priority after a change, we looked at the redundancy column, and then we implemented the process with the fewest repetitions.

NOTE 2: If they were equal in the redundancy column, the arrival time to the central processing unit was taken. To break ties, processes with equal priority were completed on a FCFS basis.

NOTE 3: If a process was running and another process arrived with the highest priority, it did not interrupt the work of the process, but waited until the end of the quantum time.

IV. Discussion

Preparing a successful proposal for a newly inspired algorithm in such a pure field is not a simple undertaking. Given the difficulties of this research topic, it was necessary to propose a new optimization technique with novel aspects. Apart from the innovation, the authors' findings were supported by a dataset and a comparison of some criteria between algorithms. As a result, the three samples from the literature were utilized as examples in this study, and the average TAT and average WT of both the proposed and presented algorithms were compared. The average TAT and average WT are examined and compared to the current round-robin scheduling algorithm in different cases. Some scenarios were demonstrated, the results of each iteration were studied, and the final outputs were compared using the round-robin algorithm to verify the quality and efficiency of the suggested mix PI-RR algorithm.

Sample 1: The first dataset, which contained seven processes, was used in this sample from the benchmark datasets used in the studies. In addition to explaining how the suggested algorithm works, this research developed a Gantt chart for the method. For the following collection of processes, Table I shows the length of the CPU-burst period in milliseconds. We assigned the time quantum as 3 ms for each process. Tables II and III show gantt chart of sample 1 and turnaround time and waiting time of sample 1, respectively.

TABLE I. Processes of Sample 1

Process	**CPU Burst time**	**Arrival time**	**Priority**	**Number of times repeat**
P1	**9**	**0**	**3**	***///***
P2	**9**	**2**	**2**	***///***
P3	**12**	**4**	**4**	***///=7***
P4	**8**	**5**	**1**	***///***
P5	**7**	**6**	**5**	***///***
P6	**9**	**6**	**7**	***///***
P7	**12**	**6**	**6**	***///=7***

TABLE II. Gantt Chart of Sample 1

P1	**P2**	**P4**	**P4**	**P4**	**P2**	**P2**	**P1**	**P1**
0-3	**3-6**	**6-9**	**9-12**	**12-14**	**14-17**	**17-20**	**20-23**	**23-26**

P3	**P3**	**P3**	**P5**	**P5**	**P5**	**P7**
26-29	**29-32**	**32-35**	**35-38**	**38-41**	**41-42**	**42-45**

P7	**P7**	**P6**	**P6**	**P6**	**P3**	**P7**
45-48	**48-51**	**51-54**	**54-57**	**57-60**	**60-63**	**63-66**

TABLE III. TURNAROUND TIME AND WAITING TIME OF SAMPLE 1

Process	Turnaround Time (ms)	Waiting time
P1	26	17
P2	18	9
P3	31	19
P4	9	1
P5	36	29
P6	54	45
P7	45	33
Average	31.285	21.875

Sample 2: For the following collection of processes, Table IV shows the length of the CPU-burst period in milliseconds. We assigned the time quantum as 5 ms for each process, which contained eight processes. Tables V and VI show gantt chart of sample 2 and turnaround time and waiting time of sample 2, respectively.

TABLE IV. PROCESSES OF SAMPLE 2

Process	CPU Burst time	Arrival time	Priority	Number of times repeat
P1	15	0	4	///
P2	18	0	1, 8	/// , /
P3	15	0	2	///
P4	14	0	3	///
P5	12	0	7	///
P6	20	0	8	
P7	7	0	5	//
P8	8	0	6	//

TABLE V. GANTT CHART OF SAMPLE 2

P2	P2	P2	P3	P3	P3	P4	P4
0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
P4	P1	P1	P1	P7	P7	P8	P8
40-44	44-49	49-54	54-59	59-64	64-66	66-71	71-74
P5	P5	P5	P2	P6	P6	P6	P6
74-79	79-84	84-86	86-89	89-94	94-99	99-104	104-109

TABLE VI. TURNAROUND TIME AND WAITING TIME OF SAMPLE 2

Process	Turnaround Time (ms)	Waiting time
P1	59	46
P2	89	71
P3	30	15
P4	44	30
P5	86	74
P6	109	89
P7	66	59
P8	74	64
Average	69.625	56

TABLE VII. PROCESSES OF SAMPLE 3

Process	CPU Burst time	Arrival time	Priority	Number of times repeat
P1	9	0	10	//
P2	12	0	9	///
P3	15	0	6	///
P4	17	0	3	///=10
P5	12	0	2	///
P6	14	0	8	///
P7	16	0	7	///,=10,/
P8	13	0	4	///
P9	6	0	5	//
P10	15	0	1	///

TABLE VIII. GANTT CHART OF SAMPLE 3

P10	P10	P10	P5	P5	P5	P4	P4
0-5	5-10	10-15	15-20	20-25	25-27	27-32	32-37
P4	P8	P8	P8	P9	P9	P3	P3
37-42	42-47	47-52	52-55	55-60	60-61	61-66	66-71
P3	P7	P7	P7	P6	P6	P6	P2
71-76	76-81	81-86	86-91	91-96	96-101	101-105	105-110
P2	P2	P1	P1	P4	P7		
110-115	115-117	117-122	122-126	126-128	128-129		

TABLE IX. TURNAROUND TIME AND WAITING TIME OF SAMPLE 3

Process	Turnaround Time (ms)	Waiting time
P1	126	117
P2	117	105
P3	76	61
P4	128	111
P5	27	15
P6	105	91
P7	129	113
P8	55	42
P9	61	55
P10	15	0
Average	83.9	71

Sample 3: For the following collection of processes, Table VII shows the length of the CPU-burst period in milliseconds. We assigned the time quantum as 5 ms for each process, which contained ten processes. Tables VIII and IX show gantt chart of sample 3 and turnaround time and waiting time of sample 3, respectively.

Scheduling is a fundamental operating system feature. Almost all computer resources are pre-programmed before they are used. One of the most important computer resources is the central processing unit (CPU). Its scheduling is crucial to the architecture of an operating system. Which processes run and which processes wait are determined by CPU scheduling. CPU scheduling is critical because it has a significant impact on resource usage, system performance, and CPU efficiency. Process execution is made up of a cycle of CPU execution (CPU burst) and I/O wait (I/O burst), with CPU burst coming first, then I/O burst, then another I/O burst, and so on. The most recent CPU explosion ends with a system request to stop the process.

Due to their significant waiting time, long response time, large turnaround time, and low throughput, existing round-robin CPU scheduling algorithms cannot be used in real-time operating systems. Furthermore, existing priority CPU scheduling algorithms are inadequate for real-time operating systems since they create starvation, and do not take into consideration which processes come first and the time spent waiting for them to run in the CPU.

The proposed mix priority and round-robin algorithm (mix PI-RR algorithm) is an algorithm that obtains the optimal advantages of both priority and round-robin algorithms. Round-robin scheduling does not give any process priority or additional consideration based on other processes, and processes cannot have priorities set for them. Therefore, delayed execution of important processes may affect the performance of the whole system. On the other hand, in the priority algorithm, each process is assigned a priority. Processes with highest priority are executed first. However, this occurs without taking into account which process comes first and the time the process has been waiting in CPU to run, and this can cause starvation. Therefore, this paper introduced the mix PI-RR algorithm to assign a priority to important processes, without causing starvation.

The average waiting and turnaround times depend on the number of processes in the ready queue; as number of processes increases, time cost increases. In addition, long burst times of the processes increase the time cost. To emphasize the efficiency of the proposed algorithm, samples datasets varying in number and burst times of processes are used. The proposed mix PI-RR algorithm enhances CPU performance in general, the results revealed that wait time and turnaround time were reduced. Furthermore, the CPU algorithms enabled the user to obtain good results without increasing the time. Tables X and XI show the average turnaround time and average waiting time of the proposed mix PI-RR algorithm and current round-robin algorithm. As observed from the average turnaround time and average waiting time, the performance of the proposed mix PI-RR algorithm was better than the current round-robin algorithm. It is clearly observed that average turnaround time and average waiting time of the processes are optimum for proposed Mix PI-RR algorithm compared to round robin fundamental algorithm. The comparison between the proposed mix PI-RR algorithm and the current round-robin algorithm is shown in “Fig. 3 and 4”.

TABLE X. COMPARING AVERAGE TURNAROUND TIME

	Average Turnaround Time (ms)		
	Sample 1	**Sample 2**	**Sample 3**
Proposed Mix PI-RR Algorithm	**31.285**	**69.625**	**83.9**
Round Robin	**48.8571**	**90**	**106.9**

TABLE XI. COMPARING AVERAGE WAITING TIME

	Average Waiting Time (ms)		
	Sample 1	**Sample 2**	**Sample 3**
Proposed Mix PI-RR Algorithm	**21.875**	**56**	**71**
Round Robin	**39.428**	**67.375**	**94**

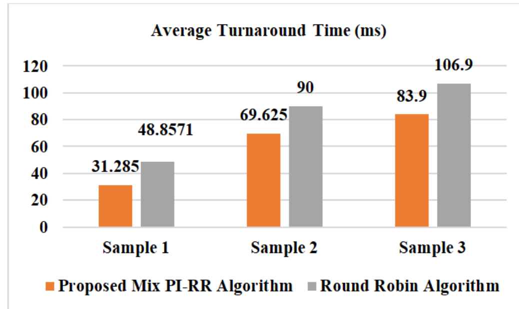


Fig. 3. Comparing Average Turnaround Time.

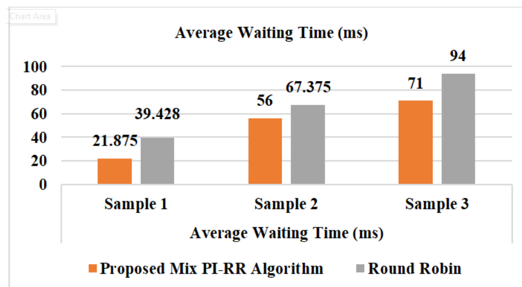


Fig. 4. Comparing Waiting Time.

V. CONCLUSION

The algorithm presented here outperforms several other algorithms; in general, it outperforms the RR and priority-based methods. No algorithm is ideal in every circumstance. It is impossible to watch a precise scheduling algorithm in action, yet precise performance can be viewed in real-time operating system operations. Several elements, such as changeable capacity, have a substantial impact on performance. This study introduced a real-time operating system and real-time tasks. We highlighted RR and priority drawbacks like high average turnaround, high context switching, high response time, high turnaround time, and low throughput, as well as the failure to take into account the process that should be first and how long processes have been waiting. After analyzing RR and priority algorithm’s performances and drawbacks, we proposed a new algorithm, named mix priority and round-robin (the mix PI-RR algorithm), which deals with the drawbacks of simple round-robin and priority algorithms. This new approach performed better than a simple RR and priority, by taking the best features of each algorithm and combining them to create the ideal algorithm for a given situation in terms of average waiting time and average turnaround time. This study justified the mix between priority and round-robin to help the CPU overcome indefinite blocking or starvation (leaving some lower priority processes waiting in CPU) in priority algorithms, and using queue up to solve the processes regardless of the importance and priority of the process for the CPU in the round-robin algorithm. The results of this evaluation highlight ways in which instructional material should be clarified. It will be important for the proposed algorithm to be more efficient and effective than current CPU scheduling algorithms. Finally, further research is needed to compare it with other algorithms. In future work, simulations of CPU scheduling strategies are recommended. The most efficient way to evaluate a scheduling algorithm is to code it and include it in an operating system; then, the algorithm's correct working capabilities can be determined in real-time systems. Further research and studies in the future should be carried out to discover other scheduling algorithms that are optimal in certain situations and, hence, deliver the highest level of user satisfaction.

Acknowledgment

The researcher would like to thank Qassim University, Kingdom of Saudi Arabia. This study was supported in part by a grant from Deanship of Scientific Research, Qassim University.

References

[1] P. B. Galvin, G.Gagne and A. Silberschatz, Operating system concepts. John Wiley & Sons. 2003.

[2] U. Shafi, M.A. Shah, A. Wahid, K. Abbasi, Q. Javaid, M. Asghar and M. Haider, A novel amended dynamic round robin scheduling algorithm for timeshared systems. Int. Arab J. Inf. Technol., 17(1), 90-98. 2020.

[3] M. Aijaz, R. Tariq, M. Ghori, S.W. Rizvi and E.F. Qazi, Efficient Round Robin Algorithm (ERRA) using the Average Burst Time. In 2019 International Conference on Information Science and Communication Technology (ICISCT) (pp. 1-5). IEEE. 2019.

[4] S. Mody and S. Mirkar, Smart Round Robin CPU Scheduling Algorithm For Operating Systems. In 2019 4th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT) (pp. 309-316). IEEE. (2019, December).

[5] H. B. Parekh and S. Chaudhari, Improved Round Robin CPU scheduling algorithm: Round Robin, Shortest Job First and priority algorithm coupled to increase throughput and decrease waiting time and turnaround time. In 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC) (pp. 184-187). IEEE. (2016, December).

[6] Amit Kumar Sain, "Dynamical Modified R.R. CPU Scheduling Algorithm", International Journal of Computer Trend and Technology, Volume 4, Issue 2, PP. 90-93, ISSN:2231-2803. 2013.

[7] M.A. Alworafi, A. Dhari, A. Al-Hashmi and A.B. Darem, An improved SJF scheduling algorithm in cloud computing environment. In 2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT) (pp. 208-212). IEEE. (2016, December).

[8] A. Joshi and s. Gosswami, Modified Round Robin algorithm by using Priority Scheduling. Advances in Computational Sciences and technology, 10(6), 1543-1549. 2017.

[9] S. Zouaoui, L. Boussaid and A. Mtibaa, Priority based round robin (PBRR) CPU scheduling algorithm. International Journal of Electrical & Computer Engineering (2088-8708), 9(1). 2019.

[10] A. Najim and Al-Tahhan, Hybrid Algorithm for CPU Scheduling by Using Dynamic Time Quantum", Future Research Journal, Al-Hadba university Collage, PP. 99-136, ISSN 1680-9300. Iraq . (April 2014),".

[11] Neelakantagouda Patil (October 2015)," A Knapsack Based CPU Process Scheduling Using Neelsack Algorithm", (IJSEAS) International Journal of Scientific Engineering and Applied Science,Volume-1, pp. 138-144, Issue-7, ISSN:2395-3470, India.

[12] William Stallings, "Operating Systems Internal and Design Principles", 5thEdition, ISBN-10: 0-13-230998, 2006.

[13] Silberschatz, A., Peterson, J. L., and Galvin, B., "Operating System Concepts", Addison Wesley, 7th Edition, ISBN-10: 0471694665, 2006.

[14] E.O. Oyetunji, A. E. Oluleye," Performance Assessment of Some CPU Scheduling Algorithms", Research Journal of Information Technology, 1(1), pp. 22-26, 2009.

[15] M Mostafa, S., & Amano, H. (2020). Dynamic Round Robin CPU Scheduling Algorithm Based on K-Means Clustering Technique. Applied Sciences, 10(15), 5134.

[16] Farooq, M. U., Shakoor, A., & Siddique, A. B. (2017, March). An efficient dynamic round robin algorithm for cpu scheduling. In 2017 International Conference on Communication, Computing and Digital Systems (CCODE) (pp. 244-248). IEEE.

[17] Zouaoui, S., Boussaid, L., & Mtibaa, A. (2019). Priority based round robin (PBRR) CPU scheduling algorithm. International Journal of Electrical & Computer Engineering (2088-8708), 9(1).Technologies and Optimization (Trends and Future Directions)(ICRITO) (pp. 397-400). IEEE.

[18] Gupta, A. K., Yadav, N. S., & Goyal, D. (2016). Design and Performance Evaluation of Smart Job First Dynamic Round Robin (SJFDRR) Scheduling Algorithm with Smart Time Quantum. American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS), 26(4), 66-78.

[19] Manuel, J. I., Baquirin, R. B., Guevara, K. S., & Tandingan, D. (2019, February). Fittest Job First Dynamic Round Robin (FJFDRR) scheduling algorithm using dual queue and arrival time factor: a comparison. In IOP Publishing Ltd, IOP Conf. Ser.: Mater. Sci. Eng (Vol. 482, p. 012046).

[20] Khatri, J. (2016). An enhanced Round Robin CPU scheduling algorithm. IOSR Journal of Computer Engineering (IOSR-JCE), 18(4), 20-24.

[21] Chandiramani, K., Verma, R., & Sivagami, M. (2019). A Modified Priority Preemptive Algorithm for CPU Scheduling. Procedia Computer Science, 165, 363-369.

[22] Tripathi, S., Prajapati, S., & Ansari, N. A. (2017, May). Modified optimal algorithm: for load balancing in cloud computing. In 2017 International Conference on Computing, Communication and Automation (ICCCA) (pp. 116-121). IEEE.

[23] Zhang, J., Chen, C., Zheng, H. K., & Luo, Q. Y. (2019, June). A High Priority Random Task Fuzzy Scheduling Algorithm for CPS. In 2019 Chinese Control And Decision Conference (CCDC) (pp. 482-487). IEEE.

[24] Y. Berhanu, A. Alemu and M.K. Mishra, Dynamic time quantum based round robin CPU scheduling algorithm (Doctoral dissertation). 2017.

Balancing a Practical Inverted Pendulum Model Employing Novel Meta-Heuristic Optimization-based Fuzzy Logic Controllers

Dao-Thi Mai-Phuong[1], Pham Van-Hung[2], Nguyen Ngoc-Khoat[3], Pham Van-Minh[4]
Faculty of Electrical Engineering, Hanoi University of Industry, Hanoi, Vietnam[1, 2, 4]
Faculty of Control and Automation, Electric Power University, Hanoi, Vietnam[3]

***Abstract*—This paper concentrates on proposing a newly effective control approach to ensure the balance of an inverted pendulum (IP) system consisting of a free-rotational rod and a small cart. The novel idea is to create an effective integration between a PD-like fuzzy logic architecture and a modified genetic algorithm (*m*GA). The *m*GA is executed with an appropriately fast optimization as an initial phase in dealing with optimizing scaling factors of the fuzzy logic controller. There are totally six meaningful scaling factors corresponding to two fuzzy logic controllers applied in the balance control system of the IP. These six scaling coefficients are highly significant to strongly affect control quality of the system applying such a PD-like fuzzy logic control methodology. Excellent results obtained in terms of numerical simulations compared with those of both the conventional PID and existed fuzzy logic counterparts as well as practical experiments implemented in a real IP topology verify the promising applicability of the new control methodology proposed in this study.**

***Keywords*—mGA; inverted pendulum (IP); balance control; scaling factors; optimization**

I. Introduction

It should be obvious that controlling an inverted pendulum (IP) system to be balanced at a desirable position of the rod embedded in the model is one of the most traditional control problems [1-3]. This control approach is highly significant to a various number of nonlinear control strategies such as control of transport machines needing to balance things, robot technology, and even launch direction of missiles. Naturally, the IP system has been usually used to testify the applicability of new control strategies.

Recently, intelligent controllers, i.e. fuzzy logic – based regulators have been increasingly developed to replace with conventional counterparts such as PI (proportional integral) and PID (proportional – integral – derivative) in dealing with nonlinear and/or uncertain control systems [4-10]. Fuzzy logic technique has been considered to be highly suitable for these types of control plants due to its operation principle depends only on knowledge and experiences of experts to handle a system. Obviously, the better understanding of the system we have, the more effective operation of a control strategy based on fuzzy logic model we can obtain [11-13].

This work concentrates on applying fuzzy logic – based control strategy which was presented in [5] to design intelligent controllers, carrying out the balance of an IP system. Since an IP system is a multi-input multi-output (MIMO) system and the balance control problem should be solved by suitable controllers, the PD-like fuzzy logic control methodology has been selected. Technically, such a fuzzy logic controller has two inputs, error and derivative of error, and one output. Corresponding to these inputs and output, there are three scaling factors which strongly affect control quality of the system need to be determined in an effective manner. This mission is able to be successfully executed by using an appropriate optimization method. Metaheuristic optimization techniques [14-18], such as GA and PSO, have been completely capable of dealing with the determination of three scaling factors.

When applying a number of aforementioned biology-inspired optimization techniques, e.g. GA, it should be clear this is a time-consuming procedure. To find an optimal solution corresponding to a set of scaling factors for the fuzzy logic controller effective for the IP's balancing control, this is a fact that the duration is frustrating to operators. The new contribution of the current paper is to find a possible way to reduce this time period. The authors propose a hybrid procedure to initialize the GA mechanism by means of a faster optimization technique [15], e.g. Salp Swarm Optimization Algorithm (SSOA). This optimization mechanism is used to initially create a set of closed-optimal parameters as first phase of the GA. It is denoted modified GA (mGA) procedure and will be used to determine scaling coefficients of the two controllers embedded in the balance control for the IP: cart's position and pendulum's rotational angle ones. Simulation processes and experiments in a real model will be implemented to demonstrate the feasibility and applicability of the control scheme proposed in this study.

The rest of this paper is organized as follows. Section II presents an overview of related works to the IP's balance control. Section III will briefly create a mathematical model of an IP system which is helpful to design balance control strategies. Next, Section IV focuses on proposing a novel GA technology modified from a typical one. A control strategy applying such a modified GA mechanism will also be provided in Section V for solving the IP's balancing problem. Section VI concentrates on testifying the control strategy proposed in both simulation and experiment scenarios. Eventually, the conclusion and future work raised from this study will also be provided in the last section.

This work was supported by Hanoi University of Industry, Project Granted number 16-2021-RD/HD – ĐHCN.

II. Related Work

Studies on IP systems, especially on the balancing control problem, have been conducted by a lot of researchers around the world. In [1], the authors investigated several traditional control strategies i.e. PID, LQR, Pole placement and Fractional PID to control the position of the pendulum's rod as well as cart via a number of simulations implemented in MATLAB software. Reference [2] introduced a comprehensive illustration from the past to current approach of nonlinear control theory for the IP system, focusing on robust control methodology. The classical PID regulators had been used for a linear inverted pendulum system in cooperation with the GA mechanism which was embedded to tune PID's parameters [3]. In [19], the conventional PID, LQR together with MPC (model predictive controller) –based schemes were applied in dealing with the IP's balance control. The authors in [20] proposed a Sugeno - based fuzzy parallel distributed compensation regulator to control a nonlinear IP model with demonstration on MATLAB simulations. In almost reports conducted on the IP balance control, the control methodologies might not be sufficient in terms of their applicability and effectiveness concerning both theory and practice. It means that if the studies used conventional regulators such as PID and LQR, the control qualities were not good enough to ensure a high quality control system. In contrast, if the intelligent control approaches e.g. fuzzy logic – based schemes were applied, they had only been employed for the theoretical IP system testified by a proper software such as MATLAB. This work will be addressed by an idea of integration theoretical and practical IP topology. In addition, a novel hybrid intelligent control strategy using fuzzy logic technology integrated with a modified GA in comparison with conventional regulator as well as initial smart controllers will be proposed in this paper.

III. A Mathematical Model of the IP System

A dynamic model representing a typical inverted pendulum system is shown in Fig. 1. This model consists of two main parts, i.e. a cart and a pole. The cart is actually a wagon driven by a small DC motor. The pole looks like a rod attached to a weight of mass m. This pole can be freely rotated by a cart – mounted rotary joint. The system as drawn in Fig. 1 is manipulated by only one force F taken out by the DC motor.

In order to establish the dynamic model of the inverted pendulum system, taking the second Newton's law of motion the following equation can be obtained:

$$\begin{cases} (M+m)\ddot{x} - \left(ml\sin\theta\right)\dot{\theta}^2 + \left(ml\cos\theta\right)\ddot{\theta} = F \\ m\ddot{x}\cos\theta + ml\ddot{\theta} = mg\sin\theta \end{cases} \quad (1)$$

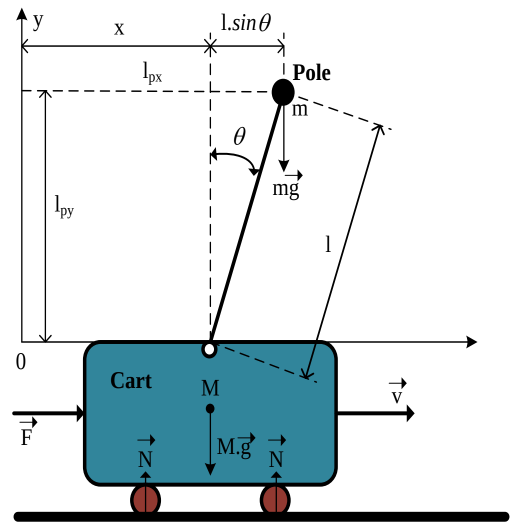


Fig. 1. The Typical Model of an Inverted Pendulum System.

Where F or u is the force of motion generated from the dc motor mounted on the cart. The fact that this force is only the control signal to keep the rod to be stable at the equilibrium. From (1), the following can be deduced:

$$\begin{cases} \ddot{x} = \dfrac{F + \left(ml\sin\theta\right)\dot{\theta}^2 - mg\cos\theta\sin\theta}{M + m\sin^2\theta} \\ \ddot{\theta} = \dfrac{F\cos\theta - \left(M+m\right)g\sin\theta + ml\left(\sin\theta.\cos\theta\right)\dot{\theta}}{ml\cos^2\theta - (M+m)l} \end{cases} \quad (2)$$

The two equations indicated above are usually used to represent the dynamics of the IP system.

IV. The Modified Genetic Algorithm - MGA

Considered to be one of the most effective optimization techniques, the genetic algorithm (GA) has been applied for a huge number of optimization problems, including a significant layer of control issues. The balance control problem of the IP system is exactly suitable in using this optimization mechanism.

Typically, the conventional GA technique is presented in Fig. 2. The core idea for the GA came from the principles of biology relating to the theory of genetics and natural selection, an alternative name of the theory of natural evolution invented by Charles Darwin. This theory mainly focuses on a natural phenomenon in which the reproduction of the fittest individuals which are selected to replace the other ones must be implemented. The production of new generations, so-called offsprings, reflects the evolution of nature. Developed by John Holland and his colleagues, the GA has been inspired by step-by-step of the genetics. They consist of several main steps as shown in Fig. 3. Remember that the first step, normally defined as initialization of parameters for the GA mechanism, is highly important. If this phase is executed in a reasonable manner, it should be possible to dramatically reduce the duration of whole procedure.

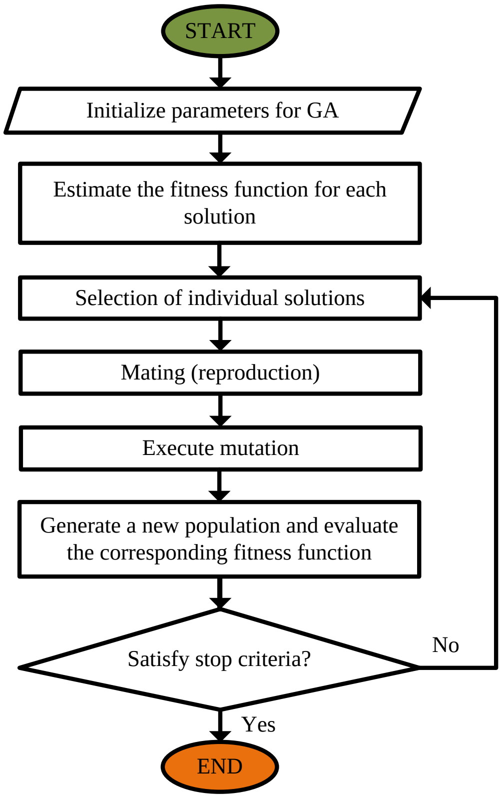


Fig. 2. The Flow Chart of a Conventional GA Technique.

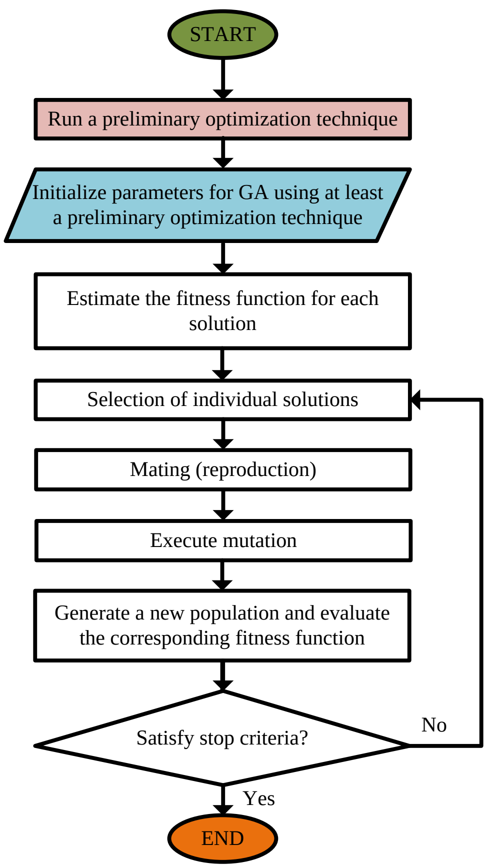


Fig. 3. The Flowchart of the Proposed Modified GA Technique.

In this paper, a modified GA (mGA) mechanism is proposed in order to minimize the period of time regarding the GA execution. An auxiliary phase is added as an initialization of parameters for the GA procedure. As shown in Fig. 3, the mGA is initialized by running a preliminary optimization mechanism. Such an optimization method aims to create initial parameters for the GA applying a possible optimization technique. A feasible method is selection of a faster optimization procedure, i.e. PSO (particle swarm optimization) and Salp Swarm Optimization Algorithm (SSOA) [15]. Theoretically, the PSO or SSOA is a local optimization technique may obtain faster execution time in comparison with the GA mechanism. This statement is also examined and confirmed by numerical simulation experiments. The mission of such a preliminary optimization technique in this aspect is to temporarily find a possible solution to the problem. The result obtained is a set of parameters which is near to local and/or global optimization point. From point of view, this idea can significantly minimize duration performing the GA procedure. Since the GA is a global optimization method, using a preliminary optimization technique to create initial parameters has a meaningful role, reducing significantly executed time. This feasible procedure also ensures the quality as well as convergent rate of the GA mechanism. The next section will verify the effectiveness of the proposed mechanism.

V. Design of an MGA – Based Control Strategy to Balancing the IP System

The balance control of an IP system, considered to be a typical nonlinear control problem, can be successfully solved by applying intelligent controllers such as fuzzy logic regulators. This study is an extended work proposed in [5]. The authors integrate the fuzzy logic architectures presented in [5] with the mGA proposed in the previous section. The novel control methodology is described in Fig. 4.

As shown, the goals of the control issue are to bring the system back to the stable state within acceptable tolerances. This means both the position of the cart and the bias angle of the pole must be manipulated to ensure the balance of the IP system. It is noted that there is only one control signal, u, which is created from two controllers regarding two feedback parameters, i.e. position x and rotational angle θ. Technically, this control signal must be fed to an actuator which is normally a DC motor. The control signal also needs to be limited because the motor can only operate within certain constraints of parameters such as power, voltage and current. In this work, a permitted range of voltage is applied to the control signal u.

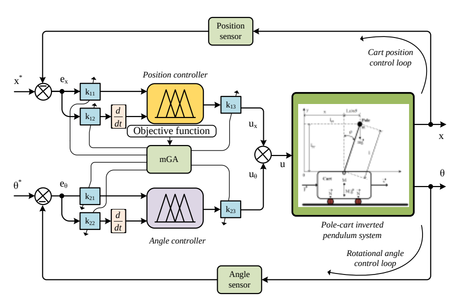


Fig. 4. Proposed Control Strategy for Balancing the IP System using mGA-based Fuzzy Logic Controllers.

The two controllers, position and angle regulators are employing the fuzzy logic architectures as presented in [5]. Since this is a nonlinear balance control problem, the PD-like fuzzy logic model is more suitable than the PI-based one. Remember that each of these PD-like fuzzy logic controllers has two inputs and one output. The two inputs are errors between desired and feedback signals. This study focuses on the vertical position of the pole, so that the desired value for the rotational angle is set to zero. In contrast, the setpoint corresponding to the desired position can be set at any value depending upon a particular control purpose. In this work, several references in accordance with the desired positions will be provided for verification via numerical simulations.

Each of two inputs and one output for an individual fuzzy logic controller as illustrated in Fig. 4 has a particular scaling factor. The fact that is each scaling factor in this aspect absolutely affects control quality of the system, leading to a mandatory requirement of determining it in a good enough manner. This is successfully implemented by means of a strong optimization technique. The modified GA proposed in the previous section is fully suitable in this perspective. Given a possible candidate of objective functions as indicated in the expression below, the proposed mGA can obtain the control quality as desired:

$$f_{obj} = \int \left(e_x^2(t) + \rho . e_\theta^2(t) \right) dt \tag{3}$$

Where $e_x(t)$ and $e_\theta(t)$, expressed in the following two equations, are position error and angle error, respectively.

$$e_x(t) = x^*(t) - x(t) \tag{4}$$

$$e_\theta(t) = \theta^*(t) - \theta(t) \tag{5}$$

The factor ρ added in (3) denotes the weighting coefficient for executing the optimization mechanism.

As shown in Fig. 4, there are totally 6 parameters with regard to 6 scaling factors of two fuzzy logic controllers need to be optimized. The new GA optimization technique proposed in the previous section is utilized to determine these scaling factors. The control performance will be testified in the next section through both numerical simulation processes and practical experiments.

VI. Numerical Simulations and Experiments

Performing numerical simulations on appropriate software as well as conducting experiments on real models is extremely necessary steps to prove the correctness and feasibility of the proposed control strategy. In this section, simulations applying MATLAB/Simulink will be executed at first to theoretically testify applicability of the new control methodology compared to those of conventional PID regulators [19] as well as fuzzy logic controllers [20]. Simulation parameters are given in Table I and Table II. Two PD-type fuzzy logic controllers used for the IP system have the same structures with the membership functions as well as sets of fuzzy rules. Fig. 5 describes membership functions for two inputs and one output of each fuzzy logic architecture, while Fig. 6 illustrates fuzzy rules in form of a 3-D graph. The rule base can be specifically found in [6]. Applying the control methodology proposed in Section IV, after executing the mGA procedure, the optimal scaling factors obtained are indicated in Table III.

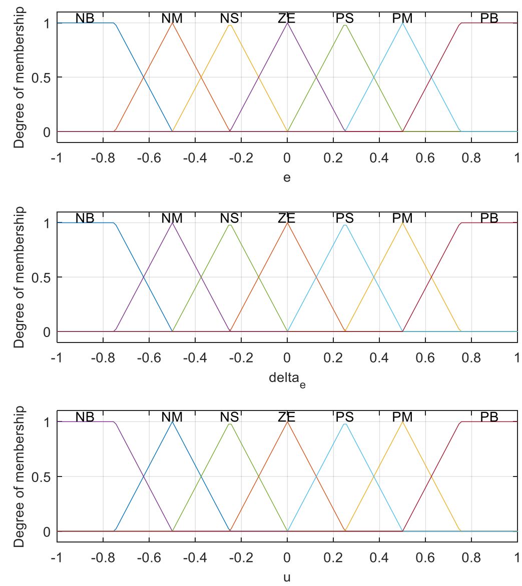


Fig. 5. Membership Functions of the PD – Like Fuzzy Logic Control Architecture.

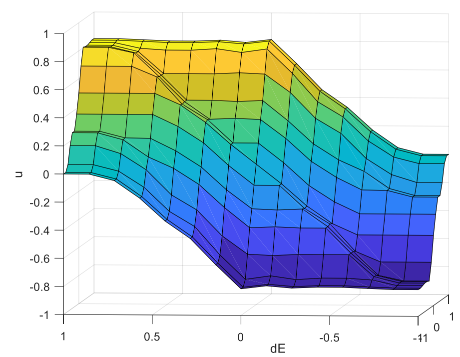


Fig. 6. A 3-D Illustration of the PD-Like Fuzzy Logic Model in Accordance with Rule base given in [6].

A. Simulation Results using MATLAB / Simulink

With six optimal factors presented in Table III, to verify the feasibility of the proposed control scheme, this work assumes the following three scenarios:

1) *Scenario 1*: The desired position of the cart is kept to be constant at 0.1 (m).

2) *Scenario 2*: The desired position of the cart is varied in four levels, i.e. 0(m), 0.1(m), 0.3(m) and 0.2(m) at different time instants.

3) *Scenario 3*: This should be the most difficult case when the cart's position is supposed to periodically change in the form of a square pulse between 0.1(m) and 0.3 (m).

All three perspectives of desirable cart's positions are plotted in Fig. 7. The results concerning these simulation cases are shown in Fig. 8 to 11. It is noted that in each illustration, both position of the cart and deviated rotational angle are considered to ensure the balance of the system. Moreover, the proposed intelligent fuzzy control strategy is compared with widely-used conventional PID regulator – based schemes [19] and fuzzy logic controller – based counterparts [20] in each simulation scenario.

In the first simulation case, due to the simplest position of the cart when considered as reference of the control system, the settling times for both the angle and position are quite excellent as shown in Fig. 8. Obviously, these durations resulting from the proposed control strategy are much less than two seconds. The results of the conventional PID controllers [19] are apparently greater than three seconds. The control quality resulting from the proposed FLC is also better than that of the FLC presented in [20] as illustrated in Fig. 8. This description demonstrates more feasible performance of the new control strategy proposed in this study.

Fig. 7. Three Scenarios of Desirable Positions to the Cart.

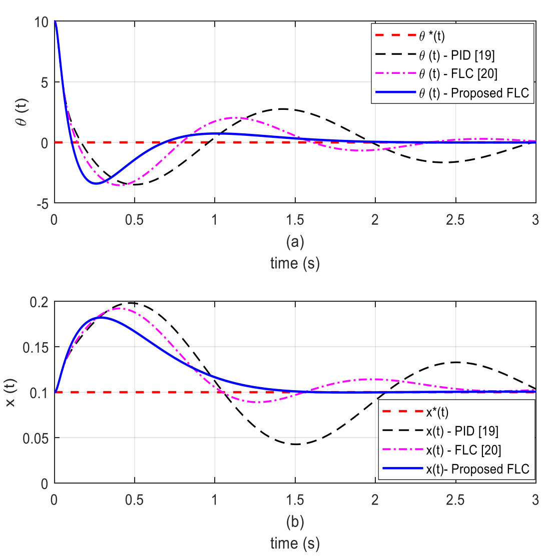


Fig. 8. Simulation Results for the First Scenario.

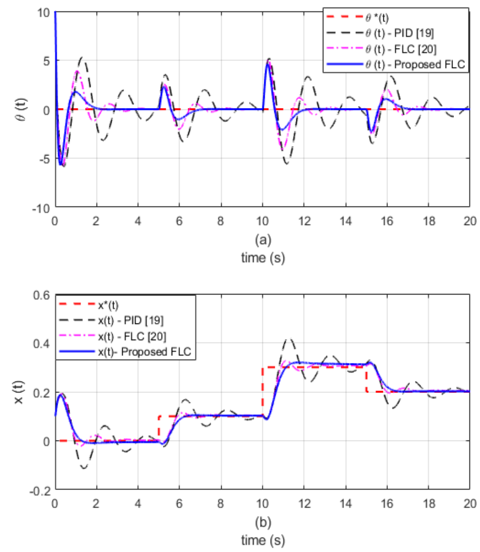


Fig. 9. Simulation Results for the Second Scenario.

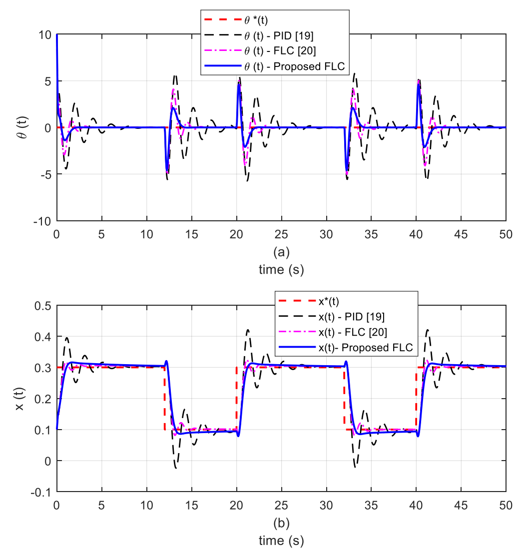
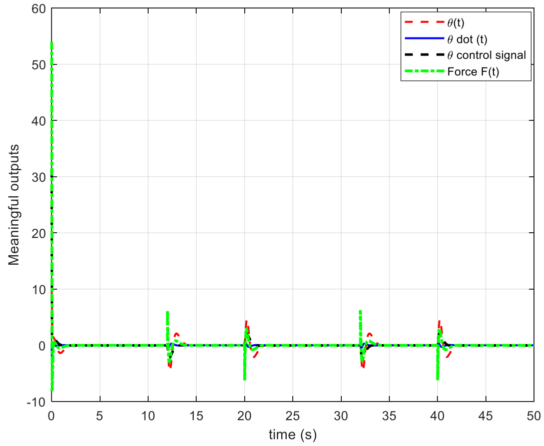
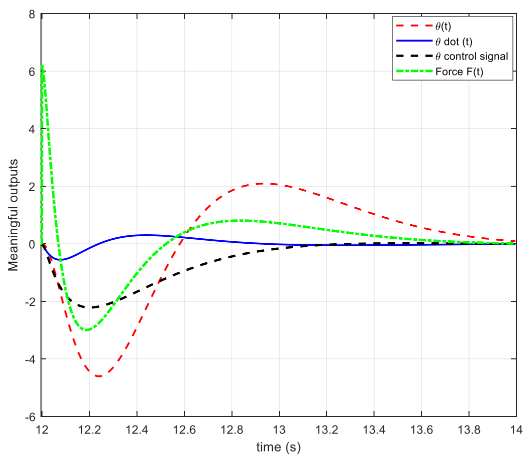


Fig. 10. Simulation Results for the Third Scenario.

In the last two simulation scenarios as shown in Fig. 9 to 11 which are more complicated than the first case, it is clear that results of the proposed fuzzy logic controllers are still much better than those of the PID controllers [19] and FLC [20]. Even when the desired position x*(t) changes periodically as illustrated in Fig. 10, the actual position of the cart is able to track well with good control performances outperforming the conventional PID counterpart. Remember that in this perspective, the deviation angle of the pendulum must be still balanced at the vertical axis, ensuring the original goal of the control problem. In addition to the deviation angle, Fig. 11 shows several control signals such as rotational acceleration, angle control signal, and control force F fed to a motion DC motor for the third simulation scenario. It should be obvious that these signals tend to be extinguished to zero values, ensuring a stable goal of vertical equilibrium for the pendulum.



(a) With Entire Simulation Time.



(b) An enlarged Part.

Fig. 11. Several Simulation Signals for the Third Simulation Scenario Applying the Proposed FLC.

B. Simulation Results via Python Software

To make a more obvious view verifying control performance of the proposed control strategy, the authors have employed Python software to simulate the balance of the inverted pendulum system in a highly intuitive way as shown in Fig. 12.

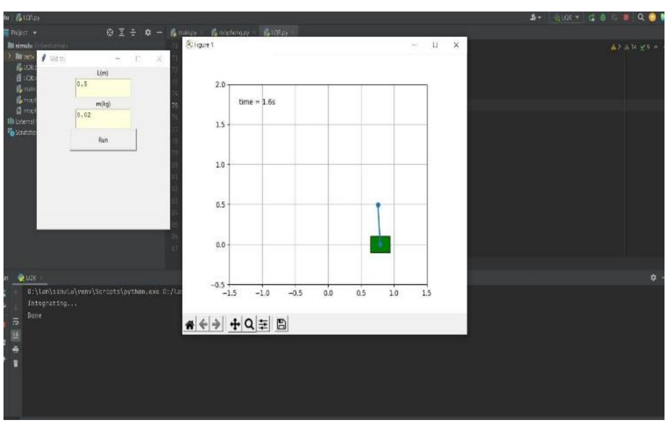


Fig. 12. Python – based Simulation Verifying Control Performances.

C. Practical Experiments

In an effort to examine the proposed control strategy, a practical IP system has successfully been built. Fig. 13 provides artwork for this practical prototype. With a set of optimal scaling factors obtained from the mGA as indicated in Table III, we address practical experiments on the IP model. Fig. 14 shows the real signals of both the position and angle of the IP system when considering the deviations of the position as presented if the previous simulation scenario. It was found that after transient period, both real signals, in form of transmitted into voltage values, are damped to be closed to zero, verifying the capability of the proposed control methodology.

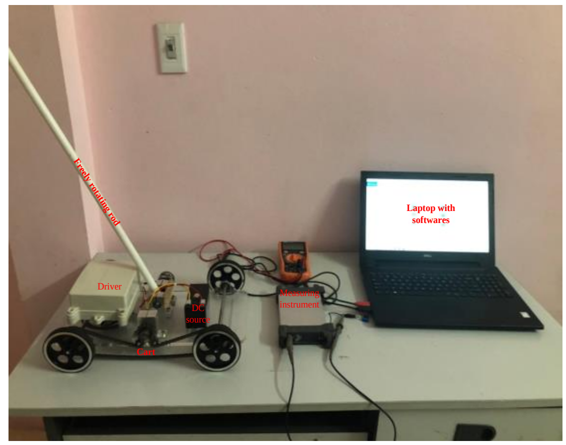


Fig. 13. The Real Model of the Inverted Pendulum System.

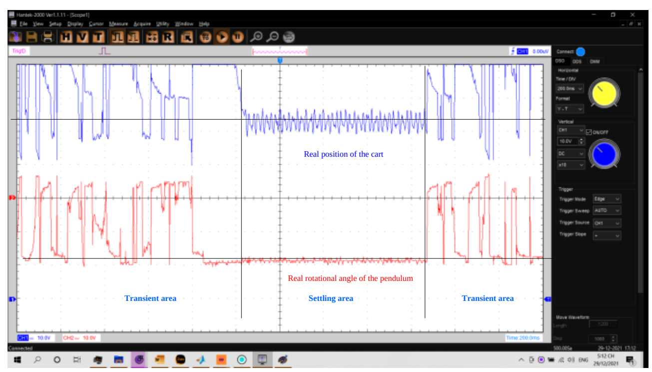


Fig. 14. Real Illustration of the Two Output Variables.

VII. Conclusion and Future Work

This study proposed a novel control strategy based on fuzzy logic architecture integrated with an mGA optimization technique to balance an IP system. The GA presented in this work used a faster optimization mechanism to accelerate global convergence, improving control quality of the system. Together with various simulation scenarios executed in MATLAB/Simulink platform, an actual IP system has also been built to demonstrate the efficiency of the control methodology proposed in this paper. Simulation results obtained by means of the proposed fuzzy logic control approach are much better than those of the conventional PID regulators as well as a single fuzzy logic controller, demonstrating the feasibility of this study. One direction to further improve this work as future work is to continue optimizing membership functions of two fuzzy logic controllers as well as rule base sets. Technically, the mGA proposed will be completely able to carry out this mission. In this aspect, a promising intelligent fuzzy logic – based control strategy can be obtained in dealing with a huge number of complicated control problems, including the IP balancing issue.

REFERENCES

[1] K. Singhal and A. Swarup, "Application and comparison of few control methods to inverted pendulum dynamics", 2018 4th International Conference on Computing Communication and Automation (ICCCA), pp. 1-6, 2018.

[2] Boubaker, Olfa; Iriarte, Rafael (ed.), "The inverted pendulum in control theory and robotics: From theory to new innovations", Control, Robotics & amp; Sensors, 2017, IET Digital Library.

[3] Sravan Bharadwaj C., Sudhakar Babu T., Rajasekar N, "Tuning PID controller for inverted pendulum using genetic algorithm", In: Konkani A., Bera R., Paul S. (eds) Advances in Systems, Control and Automation. Lecture Notes in Electrical Engineering, vol 442. Springer, Singapore, 2018.

[4] Fayçal CHABNI, Rachid TALEB, Abderrahmen BENBOUALI and Mohammed Amin BOUTHIBA, "The application of fuzzy control in water tank level using Arduino" International Journal of Advanced Computer Science and Applications (IJACSA), 2016, vol. 7(4), pp. 261 – 265.

[5] Ngoc-Khoat Nguyen, Duy-Trung Nguyen, "A Comparative Study on PI – and PD – Type Fuzzy Logic Control Strategies" International Journal of Engineering Trends and Technology, 2021, vol. 69(7), pp. 101-108.

[6] B. K. Bose, Modern Power Electronics and AC Drives. Prentice Hall, October, 2001.

[7] Sudheer Kumar Nagothu, Kurra Upendra Chowdary, Ponduri Siva Prasad, R Sirandhini, "A novel wound assessment system using fuzzy logic", International Journal of Engineering Trends and Technology, 2022, vol. 70(3), pp. 118-125.

[8] NgocKhoat Nguyen, "Study on a novel speed control strategy of hydraulic turbines applying artificial neural network and metaheuristic optimization technique", International Journal of Electrical Engineering and Technology (IJEET), 2020, vol. 11(10), pp. 311-321.

[9] D. N. Utama et al. Decision support model for evaluating the level of the implementation effectiveness of 3in1 or ERP systems using fuzzy logic method, International Journal of Emerging Trends in Engineering Research, 2020, vol. 8(7), pp. 3799 - 3803, 2020.

[10] T-M-P. Dao, Y. Wang, N-K. Nguyen, "Applying meta – heuristic optimization methods to design novel adaptive PI – type fuzzy logic controllers for load – frequency control in a large – scale power grid", Turkish Journal of Electrical Engineering & Computer Sciences, 2016, vol. 24(6), pp. 4900-4914.

[11] Guanrong Chen, Trung Tat Pham. Introduction to fuzzy sets, fuzzy logic, and fuzzy control systems. CRC Press, 2001.

[12] Simões, Marcelo Godoy, Artificial intelligence for smarter power systems: Fuzzy logic and neural networks, Energy Engineering, IET Digital Library, 2021.

[13] Madhusudhan M, H. Pradeepa, "Fuzzy based controller to enhance transient and steady state stability for multimachine power system", International Journal of Engineering Trends and Technology, 2022, vol. 70(2), pp. 103-114.

[14] Mahmoud Baklizi, "FLACC: Fuzzy logic approach for congestion control" International Journal of Advanced Computer Science and Applications (IJACSA), 2019, vol. 10(7), pp. 43 - 50.

[15] Murat Cikan, Bedri Kekezoglu, "Comparison of metaheuristic optimization techniques including equilibrium optimizer algorithm in power distribution network reconfiguration", Alexandria Engineering Journal, 2022, vol. 61(2), pp. 991 – 1031.

[16] Allawi Z.T.; Ibraheem I.K.; Humaidi A.J, "Fine-tuning meta-heuristic algorithm for global optimization", Processes, 2019, vol. 7(10), pp. 1-14.

[17] Fernando Martinez Santa, Fredy H. Martinez Sarmiento and Holman Montiel Ariza, "Path optimization for mobile robots using genetic algorithms" International Journal of Advanced Computer Science and Applications (IJACSA), 2022, vol. 13(2), pp. 657 – 662.

[18] Thi-Mai-Phuong Dao, Yaonan Wang, Ngoc-Khoat Nguyen, "A novel control approach for switched reluctance motors based on fuzzy logic and particle swarm optimization techniques". Journal of Scientific Bulletin Series C: Electrical Engineering and Computer Science, 78(3), pp. 85-96, 2016.

[19] Elisa Saa Varghese et al, "Optimal control of inverted pendulum system using PID controller, LQR and MPC", IOP Conf. Ser.: Mater. Sci. Eng. 263, 2017.

[20] Ahmad Ilyas Roose, Samer Yahya, Hussain Al-Rizzo, "Fuzzy-logic control of an inverted pendulum on a cart", Computers & Electrical Engineering, 2017, vol. 61, pp. 31-47.

APPENDIX

TABLE I. NOMENCLATURE

Symbol	**Meaning**	**Value, [unit]**
M	Mass of the cart	0.1 [kg]
m	Mass of the pole	0.02 [kg]
l	Length of the rod (mass omitted)	0.35 [m]
$x(t)$	Position of the cart	[m]
$\dot{x}(t)$	Velocity of the cart	[m/s]
$\ddot{x}(t)$	Acceleration of the cart	[m^2/s]
$\theta(t)$	Pendulum angle (upright position)	[rad], [degree]
$\dot{\theta}(t)$	Rotational velocity of the pole	[rad/s]
$\ddot{\theta}(t)$	Rotational acceleration of the pole	[rad^2/s]
F	Variable force exerted on the cart	[N]
k11, k12, k13	Scaling factors of the fuzzy logic – based position controller	N/A
k21, k22, k23	Scaling factors of the fuzzy logic – based angle controller	N/A

TABLE II. PARAMETERS TO EXECUTE THE MGA TECHNIQUE

No.	**Symbol**	**Meaning**	**Value**
1	*Nmax*	Maximum value of generations	1000
2	*Npop*	Population size	20
3	*Nvar*	Number of variables	6
4	*Hc*	Upper constraint	100
5	*Lc*	Lower constraint	0
6	*cr*	Crossing factor	0.6
7	*mt*	Mutating factor	0.4

TABLE III. OPTIMAL SCALING FACTORS FOR TWO FUZZY LOGIC CONTROLLERS APPLYING MGA TECHNIQUE

Symbol	**Type of controller**
k11 = 12.154; k12 = 4.572; k13 = 1.245	Angle fuzzy logic controller
k21 = 6.798; k22 = 7.923; k23 = 0.865	Position fuzzy logic controller

Application of Affective Computing in the Analysis of Advertising Jingles in the Political Context

Gabriel Elías Chanchí Golondrino[1], Manuel Alejandro Ospina Alarcón[2]
Faculty of Engineering, Systems Engineering Program
University of Cartagena
Cartagena de Indias, Colombia

Luz Marina Sierra Martínez[3]
Faculty of Electronic Engineering and Telecommunications, Systems Engineering Program
University of Cauca
Popayán, Colombia

***Abstract*—Affective computing is an emerging research area focused on the development of systems with the ability to recognize, process and simulate human emotions in order to improve a user's experience in an interactive system. One of the possible fields of application of affective computing is marketing and advertising, where the application of emotion analysis techniques on the opinions made by users in different contexts has been evidenced, being a challenge the analysis of different types of multimedia content, such as advertising jingles. Thus, in this article we propose as a contribution the development of an emotion analysis study on the advertising jingles of the main candidates for mayor of Cartagena-Colombia in the period 2020-2023. For the development of this study, the acoustic properties of arousal and valence present throughout the audio track of each advertising jingle were taken into consideration, in such a way that through these properties it is possible to classify an audio fragment into an emotion belonging to the circumflex model or Russell model. To perform the segmentation of the audio track and the extraction of the acoustic properties of arousal and valence, we developed the MUSEMAN tool, which allows determining the emotions fluctuation in the advertising jingle audio track.**

Keywords—Affective computing; advertising jingles; emotion analysis; arousal; valence

I. Introduction

One of the most effective means of promoting products, services and even political campaigns is advertising jingles. These are broadcast on radio networks, television channels, web portals and social networks in order to promote the use of a product or service or motivate the intention to vote for a candidate [1]. Given that advertising jingles comprise an audio component, it is possible to analyze the emotionality present during the audio track using the methods and techniques provided by affective computing [2]. Research in affective computing focuses on systems being able to interpret the emotional state of a human and obtain or provide an appropriate response in real time to the identified emotion [3]–[7], which can be used to improve the interaction of a user with a service, content, or specific product [8].

A common approach to detecting emotions in acoustic content is the circumflex model, or Russell's model, which proposes to represent emotional states in two dimensions [9]–[11]: The horizontal dimension (x-axis), called valence, expresses the degree of positivity or negativity of the emotion (pleasure/displeasure), and the vertical dimension, called arousal, indicates the degree of excitement of the emotion (arousal-relaxation) [12]. Emotions can therefore be defined in regions within the emotional plane as a combination of valence and activation [13], [14]. In detecting emotions in music, the arousal value represents the intensity and activity in the course of an audio track, so that fast tracks and loud sounds will have a high arousal value, while tracks with a softer sound and slower tempo will have a low arousal value. Valence thus represents the acoustic property that describes the positivity present in an audio track, i.e. tracks with a positive valence are associated with emotions such as joy, euphoria, and happiness, while tracks with a negative valence correspond to emotions such as sadness, depression, or anger [15]. In this way, from the musical properties of arousal and valence of an audio track, by mapping them in the emotional plane it is possible to obtain the trigonometric angle and the emotion associated with the Russell model, thereby avoiding taxonomic ambiguity [16].

Based on the above and taking advantage of the advantages provided by the acoustic properties of arousal and valence, in this article we proposed as a contribution the development of a study based on emotion analysis on the advertising jingles of the candidates for mayor of Cartagena-Colombia for the period 2020-2023, in which the extraction and segmentation of the audio of each jingle was performed in order to obtain the emotion associated with each segment taking into account the Russell model and with the aim of obtaining the trace of emotions of each jingle. To carry out the above process, the MUSEMAN (Music Emotion Analyzer) tool was developed in Java language, which allows automating the above-mentioned analysis process using free libraries and technologies (FFmpeg, openEAR and JFreeChart). The FFMpeg tool allows the extraction and segmentation of the audio track into fragments of 5 seconds or less. The openEAR tool was used in the background and allows obtaining the acoustic variables of arousal and valence for a given audio fragment, which allow classifying each fragment in Russell's two-dimensional model. Finally, the JFreeChart library was used to generate the emotion trace graph for each jingle, as well as to generate the spider web graph with the distribution of emotions for each jingle. Thus, the present proposal has as a contribution with respect to other works of the state of the art the application of affective computing in the analysis of advertising jingles in the political context, as well as the automation of the process of analysis of emotions in multimedia audio contents. The approach proposed in this article is intended to support content creators in the objective design of the emotions communicated in multimedia advertising content, in order to generate greater interest in the audience.

The article is organized as follows: Section 2 describes a set of related works that were taken into account for the development of the present research; Section 3 presents the materials and methods used for carrying out this research study; In Section 4, the results of the analysis performed on the advertising jingles of the candidates for mayor of Cartagena de Indias using the MUSEMAN tool, are presented; and finally, Section 5 reports the conclusions and future work derived from this research.

II. Related Work

In previous work, the following proposals are found. In 2021, Sánchez-Barragán et al. [17] propose a music content recommendation system, which has as input the affective profile of the user determined from the sentiment analysis of the comments made through the social network twitter. In 2020, Russo et al. [16] presented a system based on images of cochleograms for the detection of affective musical content, first using Russell's model in combination with a convolutional neural network to extract the relevant musical characteristics. In 2019, Chmulik et al. [12], evaluated different selected acoustic characteristics using Russell's model (2-D emotional model). For comparison, they applied two-dimensional cepstrum (TDC), MIRToolbox (Matlab), Linear Prediction (LP) analysis, on a support vector regression (SVR) model. Also in 2019, Panwar et al. [18], studied the impact of emotions transmitted by music broadcast by radio channels, by mapping a function to transform music from radio channels into emotions. For the above, Thayer's plane of arousal and valence was used [19] and a linear regression model was applied to predict the values of music on the emotional plane. In addition, the mapping of emotions was performed at different times in various cities in the United States, obtaining a window for marketing with music and thus directing the different advertising messages in specific slots. In 2019, García et al. [20] , an emotion analysis study based on data mining was carried out on the tweets generated in Mexico related to the 2017 earthquake. In addition to detection of the primary emotions, the study sought to identify the evolution of emotions over time and the patterns of propagation by comparing two datasets, one containing tweets with the word "earthquake" and the other that included a mention of the emergency mechanisms. The study was able to conclude that the sentiments most widely propagated independent of the dataset were anger and joy. In 2019, Chanchí and Cordoba [2], a study of emotions and sentiments was carried out on the speech of the signing of the peace agreement. The analysis was performed by extracting the emotions during the audio track, making use of the arousal and valence variables in each subsegment of the track. Sentiment analysis was performed by studying the text of the speech through the service provided on the ParallelDots platform. This study was able to determine concordance between the motionality of the audio track and the polarity of the text of the speech. In 2018, Mäntylä et al. [21], a review of the literature was carried out on 6996 Scopus papers on the subject of sentiment analysis, making use of data mining techniques. The study concluded that work carried out at the beginning of the 21st century was focused on studies of public opinion. However, the theme of sentiment analysis acquired a greater diffusion from 2004 with the evolution of the web and the content generated and published by users in this. According to the results of the study, sentiment analysis has currently been dispersed to different topics such as the analysis of opinions of products or services, analysis in social networks, reactions to disasters, voting intentions in elections, medicine, and engineering software. In 2013, Mosquera-Cabrera [22], an exploration of music as a phenomenon of psychological interest and its participation in the development of emotional experiences is carried out. Through the exploration of different sources, the authors make a description of works that integrate neuroscience, the study of higher psychological functions and musical stimuli to elucidate the mechanisms that allow defining the role of music in the manifestation of positive emotions. The work concludes that thanks to the benefits of music, it becomes a tool to support the human being in different settings, in which it seeks to positively influence personal, social, and intellectual development. In this sense, among the main benefits of music in human beings, the authors found: facilitating the expression of emotions, energizing the body and mind, relieving fears and anxieties, psychophysical relaxation, and decreased perception of pain.

From the review of previous works, it is possible to identify that at the level of affective computing, sentiment analysis techniques have been widely spread in the study of opinions made in contents belonging to social networks in different application contexts, being a challenge the application of affective computing in the analysis of multimedia advertising content. In this sense, within this article it was proposed as a contribution the application of affective computing techniques in the analysis of emotions on advertising multimedia content in the political context, taking advantage of the advantages provided by the acoustic variables of arousal and valence in the determination of emotions from Russell's two-dimensional model. Similarly, considering that the existence of automated tools for the analysis of emotions in multimedia content has not been evidenced, this article proposes an automated tool called MUSEMAN, which allows the segmentation and analysis of the audio tracks of advertising jingles to obtain the emotions transmitted in each multimedia content.

III. Materials and Methods

This section presents the jingles considered for the present study, as well as the conceptualization about the musical recognition of emotions and the methodology used for the conduction of the developed study.

A. *Jingles Considered*

The present study was carried out considering the jingles published on social networks of the candidates who obtained the 6 highest votes in the elections of October 27, 2019; with William Dau, the winner, followed by William García, Yolanda Wong, Sergio Londoño, Fernando Araujo and Jesús Hernández. The jingles were obtained from YouTube and Facebook and transformed into audio and video formats using online encoding services. Table I shows the description of the videos from which the audio tracks considered for the present study were taken.

B. Music Emotion Recognition

Music emotions recognition (MER) uses several approaches: 1) The categorical one predicts an emotion using a classification with labels; 2) The dimensional approach establishes the emotions of a song, representing them on a dimensional plane with numerical values in X and Y [11], [19]; and 3) A dynamic process, to predict the emotion of each short segment of a song, thus obtaining the variation of musical emotion [23]. The steps followed to conduct a recognition of musical emotions may vary, depending on the classification process to be followed, but in general terms it can be summarized as follows [24]: 1) Speech acquisition the data is acquired and makes sure to have the corpus to use for the training datasets. 2) Noise removal - improving the quality of the data to be processed is vital for proper data processing, for this you can use filtering algorithms, spectral restoration and speech-structure based. 3) End point detection - the separation of non-speech sounds: recognition depends also on this separation being carried out; 4) Windowing - voice segments (word) are limited in a window, to help in the extraction of characteristics, for which the Mel frequency cepstral coefficient (MFCC) method is used; 5) Feature Extraction - the relevant characteristics for the prediction model to be built are selected, methods such as Hidden Markov Models and Vector Support Machines are used.

In the present work, the MUSEMAN tool was developed in the Java language, which allows the segmentation and extraction of the acoustic properties of arousal and valence associated with the segments of an audio track. These acoustic properties allow the classification of a certain audio segment within the circumflex model (Russell model) in such a way that the tool obtains the fluctuation of emotions throughout an audio track. Specifically, for the classification of emotions, the trigonometric angle θ formed with respect to the x-axis is obtained according to Equation 1 [14]. The angle θ obtained determines the quadrant and the emotion to which the analyzed audio segment belongs.

$$\theta = \arctan\left(\frac{arosual}{valence}\right) \tag{1}$$

The extraction of acoustic characteristics is carried out in the background by using the openEAR library [25], which comprises a set of tools with functionalities for the recognition of emotions, such as audio recording, feature extraction, classification, result evaluation and previously trained models. It is free software available under the GNU General Public License.

C. Experimental Methodology

To carry out this work, Iterative Research Pattern – IRP [26] were used, which has four phases as presented in Fig. 1. Next, Table II shows the decomposition of each phase into activities.

The different activities of the methodology are presented in Table II.

TABLE I. STATISTICAL RESULTS OBTAINED

Id	Jingle	Description
1	**Title:** I'm with Dau - Let's save Cartagena. **Duration:** 38 seconds **Facebook ID:** 926585794379373	Commercial for the candidate William Dau. Without background music, images allusive to the city with a voice in the background talking about corruption and inviting to vote for a new option.
2	**Title:** Commercial William García Tirado for mayor of Cartagena. **Duration:** 30 seconds **Youtube ID:** zXKFsvAvnIg	Commercial for the candidate William García Tirado. Song adapted for the candidacy with the images of the candidate walking with the people through the streets of Cartagena.
3	**Title:** Yolanda Wong – The strength of our young people. **Duration:** 57 seconds **Youtube ID:** eTtIQJYs2Qw	Commercial for the candidate Yolanda Wong. Song adapted for the candidacy and choreography of young people in the historic center.
4	**Title:** Sergio Londoño – Gone are the days of getting it wrong. **Duration:** 59 seconds **Youtube ID:** NNalpG-R60A	Commercial for the candidate Sergio Londoño. Without background music, the candidate appears speaking about corruption in the city, about his government plans and inviting citizens to vote.
5	**Title:** Fernando Araujo - I will be the first person to police Cartagena. **Duration:** 20 seconds **Youtube ID:** CXhW6sKe7cU	Commercial for the candidate Fernando Araujo. Without background music, the candidate appears walking through the streets of Cartagena talking about his proposals.
6	**Title:** Jesús Hernández Amín **Duration:** 38 seconds. **Youtube ID:** hxidFtMPD-I	Commercial for the candidate Jesús Hernández Amín. Without background music, the candidate appears reflecting on the problems of Cartagena with images alluding to them.

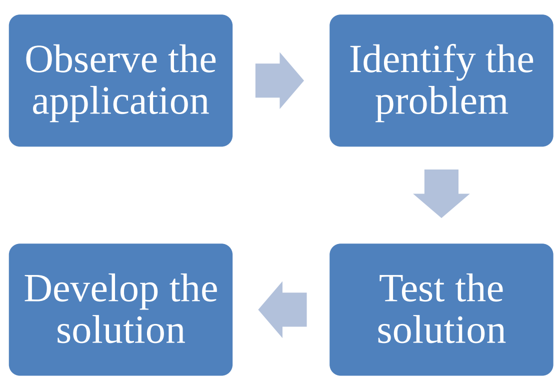


Fig. 1. Methodology Phases.

TABLE II. RESULTS OF DOMINANT POLARITIES

Phases	Activities description
Phase 1 - Observe	Activity 1.1: Review of the state of the art on Music Emotion Recognition Activity 1.2: Definition of the model of emotions, an adaptation of Russell's model of emotions was defined, which is based on five emotions: happiness, excitement, sadness, anger and relaxation, taking into account the proposal presented in [15].
Phase 2 - Identify	Activity 2.1: Characterization of the improvements to be made to the adapted and improved from tool [27]. Activity 2.2: Definition of the architecture of the MUSEMAN tool.
Phase 3 – Develop	Activity 3.1: Design and development of the tool, based on the model defined, the MUSEMAN tool was designed and implemented in the Java language, which allows dividing an audio track into segments and obtaining, through the acoustic properties, the emotion associated with each audio segment. Activity 3.2: Design and execution of unit and integration tests of the implemented features.
Phase 4 – Test	Activity 4.1: A study of the emotionality of the advertising jingles, emotional analysis was carried out on each of the audio tracks of the chosen advertising jingles, using the MUSEMAN tool. Activity 4.2: Analysis of results, the analysis and discussion of the results is carried out, obtaining the main conclusions using the MUSEMAN tool.

IV. RESULTS

This section presents the software architecture of the MUSEMAN tool and the results of the emotion analysis on the selected advertising jingles using the proposed tool.

A. MUSEMAN Software Tool

This section describes the software architecture of the tool for the classification of emotions from analysis of the acoustic properties of arousal and valence, adapted and improved from [27]. In addition, the definitive interfaces of the MUSEMAN tool are presented. Fig. 2 outlines the functional architecture of the MUSEMAN tool, which was developed in the Java language and has three functional layers: View, Analysis and Storage.

In the view layer is the graphical interface of the tool, where the audio file to be analyzed is loaded, which can be in .mp3 or .wav format. Once the file has been loaded, the emotionality analysis begins from the analysis layer, in which the file is segmented into 5-second fragments with the help of the FFmpeg library, in such a way that temporary files are generated audio on the storage layer. As the audio fragments are generated, the acoustic properties of arousal and valence are obtained, making use of the free openEAR library. With the acoustic properties of arousal and valence obtained, it is possible to classify each fragment within the circumflex model or Russell model, from obtaining the trigonometric angle formed by these two variables. In this work the model of emotions proposed in [14] was used, which has five emotions: happy, excited, angry, sad and relaxed, each of which have a coverage range of 72. As emotions are calculated, they are presented in the view layer and kept in the storage layer in a vector (time stamp and emotion), in such a way that the vector will have as many emotions as fragments are obtained from the loaded audio track. When the analysis process on the audio track has finished, it is possible to generate from the view layer a report with the results of the analysis in a csv file, consulting the emotions and the time stamps in the emotions vector. Additionally, when the report is generated, two types of graphs are also obtained: one that shows the fluctuation of emotions and their distribution along the track analyzed. These graphs were generated with the help of the Java JFreeChart library.

Fig. 3 presents the graphical interface of the MUSEMAN tool, which was developed in the Java language (using the swing library components) and has three tabs: "Emotions Analysis", "Emotions Trace" and "Emotions Distribution".

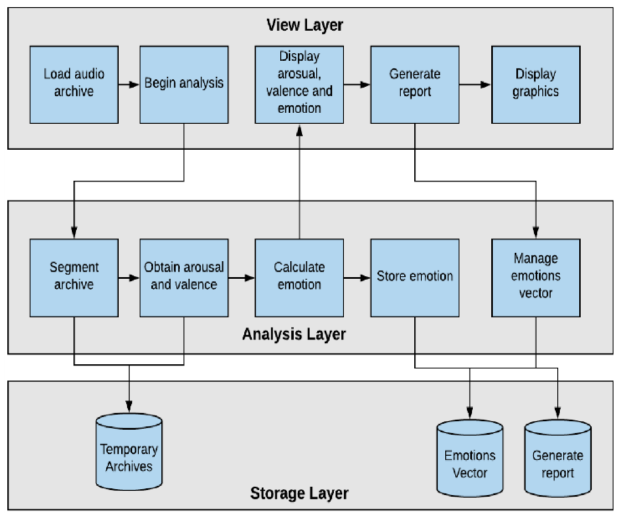


Fig. 2. Architecture of the MUSEMAN Software Tool.

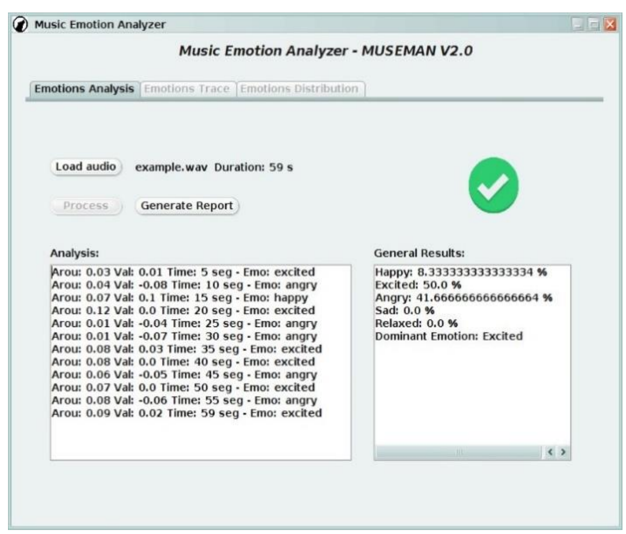


Fig. 3. Main Interface of the MUSEMAN Software Tool.

From Fig. 3, in the "Emotions Analysis" tab, the audio file is loaded using the "Load audio" button, then when pressing the "Process" button, the segmentation of the audio track starts (FFmpeg library), to obtain the variables arousal and valence (openEAR library), and calculate the emotion associated with each fragment. In the text area on the left, in Fig. 3, the value of the variables arousal and valence is displayed in real time, as well as the emotion of each 5-second segment throughout the audio track. When the analysis process is complete, the percentages for each emotion and the dominant emotion on the audio track are displayed in the text area on the right. Also in Fig. 3, it can be seen that, for the case of the loaded audio file, it is obtained that the "excited" emotion is the one with the highest percentage in the audio track with a value of 50%. In Fig. 3, the "Generate Report" button can be seen, with which a .csv file is obtained with the time stamp of each segment and the emotion corresponding to them. Once the report has been generated, in the tab "Trace Emotions" a graph is obtained that shows the fluctuation of emotions along the audio track, as can be seen in Fig. 4. This graph is made by using from the JFreeChart library.

In Fig. 4, the fluctuation of emotions in the course of the loaded audio track in Fig. 3 is shown, where it can be seen that the track fluctuates mostly between excited (2) and angry (3) emotions, with a shift towards happy emotion (1).

In Fig. 5 the "Emotions Distribution" tab of the MUSEMAN tool is shown, where a radial graph shows the distribution of emotions in the audio track, that is, the number of times an emotion is present in the audio. Thus, for example, it is observed that for the track loaded in Fig. 3, there are six instances corresponding to the excited emotion, five instances for the angry emotion and one instance associated with the happy emotion.

Finally, it is appropriate to specify that the tool requires the use of the FFmpeg and openEAR libraries in the background. It was therefore necessary to deploy the tool on the Linux operating system, in its Lubuntu distribution.

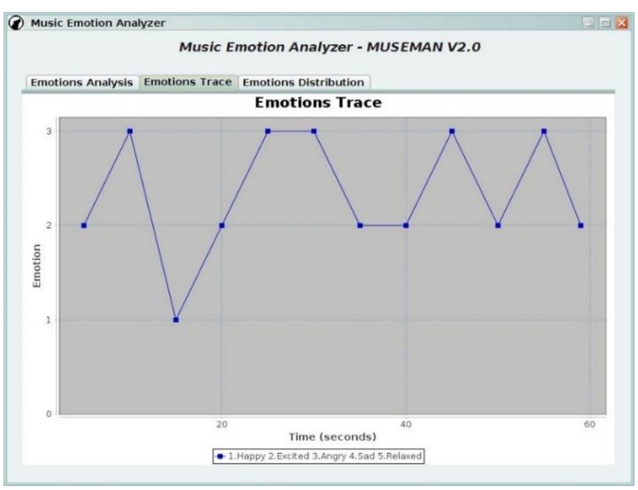


Fig. 4. Graph of Fluctuation of Emotions.

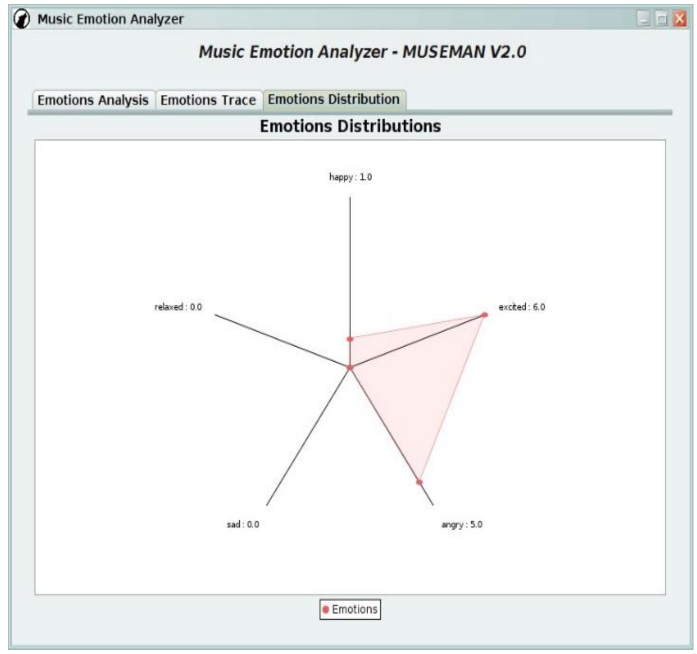


Fig. 5. Distribution of Emotions in the Audio Track.

B. *Emotions Analysis Carried Out on Advertising Jingles*

From the previously constructed tool presented, the analysis of the emotions on the audio associated with the advertising jingles of the candidates for mayor of Cartagena de Indias for the period 2020-2023 was carried out. From the audios associated with the videos presented in Table I and using the MUSEMAN tool, the graph of the fluctuation of emotions of the jingles of the candidates was obtained. In Fig. 6, it can be seen that the horizontal axis (x) shows the duration in seconds of the track, while the vertical axis (y) shows the emotions identified at a certain moment of the track: 1 happy, 2 excited, 3 angry, 4 sad, and 5 relaxed.

In Fig. 6, it can be seen that the horizontal axis (x) shows the duration in seconds of the track, while the vertical axis (y) shows the emotions identified at a certain moment of the track: 1 happy, 2 excited, 3 angry, 4 sad, and 5 relaxed. In Fig. 6 it can be seen that the jingles associated with:

- Candidates Araujo, García and Wong remain or broadcast their entire duration in an excited state (2).
- Candidate Dau spends half the time in the excited state (2) and alternates the rest of the time between happy (1), sad (4) and relaxed (5) emotions.
- Candidate Londoño is associated with the angry emotion (3) most of the time with two fluctuations towards the excited (2) and sad (4) states.
- Candidate Hernández remains most of the time in the sad state (4), with a fluctuation towards the relaxed state (5).

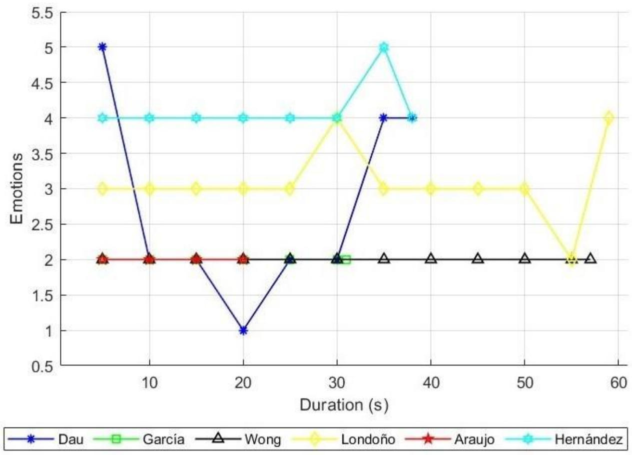


Fig. 6. Emotion Analysis Results.

The above results can be seen more clearly in Table III, which shows the way in which emotions are distributed as a percentage in each audio track.

Additionally, in Table III, it can be seen that the track of:

- Candidate Dau's dominant emotion is the excited state (50%) with fluctuations towards happy, relaxed and sad states, taking into account that in a large part of the track there is an emphasis on some acts of corruption in the city, ending with an invitation to vote for the candidate.

TABLE III. RESULTS OF DOMINANT POLARITIES

Track	Percentage of distribution of emotions
1 - Dau	%Happy = 12.5 %Excited = 50 %Angry = 0 %Sad = 25 %Relaxed = 12.5
2 – García	%Happy = 0 %Excited = 100 %Angry = 0 %Sad = 0 %Relaxed = 0
3 – Wong	%Happy = 0 %Excited = 100 %Angry = 0 %Sad = 0 %Relaxed = 0
4 - Londoño	%Happy = 0 %Excited = 8.33 %Angry = 75 %Sad = 16.66 %Relaxed = 0
5 – Araujo	%Happy = 0 %Excited = 100 %Angry = 0 %Sad = 0 %Relaxed = 0
6 - Hernández	%Happy = 0 %Excited = 0 %Angry = 0 %Sad = 87.5 %Relaxed = 12.5

- For candidates García and Wong, the dominant emotion is associated with the excited or euphoric state (100%). This reflects that throughout the track the message is transmitted by the lyrics of a song with a high arousal level.
- For candidate Londoño, the dominant emotion is that of the angry state (75%) with fluctuations towards the excited and sad states. The above highlights that for most of the track the candidate denounces the acts of corruption of past administrations, ending with an invitation to vote for his proposals.
- Candidate Araujo has the excited state (100%) as the dominant emotion, explaining the fact that the candidate throughout the jingle is accompanied by people talking in a euphoric way about their proposals for change.
- Candidate Hernández's dominant emotion is the sad state (87.5%) with a fluctuation towards the relaxed state, which is explained by the fact that for most of the track the candidate reflects on the situation of the city with a slow tone of voice and ending with the invitation to vote for his proposals.

V. DISCUSSION OF THE RESULTS

In this article we presented as a contribution the application of affective computing techniques in the emotional analysis of advertising jingles in the political context, taking advantage of the benefits of multimedia content and specifically audio and music to communicate emotions. In this sense, the acoustic properties of arousal and valence were used to determine the emotion associated with an audio fragment belonging to a jingle in Russell's two-dimensional model [11], [15]. Thus, the approach proposed in this work provides an objective contribution to the creators of advertising content for the design of the emotions to be communicated in an advertising jingle, enriching the traditional methods of content creation.

For the development of the proposal presented in this article, an automated tool was built for the analysis of emotions in advertising jingles called MUSEMAN, which allows the fragmentation of the audio of a jingle in segments of 5 seconds, to each of which the variable arousal and valence is determined, so that the tool allows to obtain statistically the percentage distribution of emotions in the jingle and graphically the variation of emotions over time. It is worth mentioning that in order to obtain the acoustic properties of each of the audio fragments; the proposed tool makes use of the openEar library in the background. Thus, with respect to other state-of-the-art proposals, including the openEar library itself, the MUSEMAN tool automates the process of analysis of advertising jingles, having as an added value the segmentation processes, as well as the statistical and graphic analysis of the emotions present in the analyzed jingle, which allows supporting content creators in the design of the emotions to be communicated in an advertising jingle.

One of the limitations of this study is that it focuses on the emotional analysis of advertising jingles in the political context from the analysis of acoustic variables, so that this study can be complemented in the future by analyzing the text corresponding to the message expressed in the audio, for example by using sentiment analysis techniques, in order to determine the level of polarity of the message. Thus, a possible challenge in the analysis of emotions in advertising jingles could be the use of fuzzy logic based systems that receive as input the polarity values of the text and the emotionality of the audio to determine as output an enriched emotionality level for the analyzed jingle.

VI. Conclusion

In this work, an analysis of emotions was carried out on the advertising jingles of the candidates for mayor of Cartagena de Indias for the period 2020-2023. The study was carried out from the analysis of the audios associated with each of the jingles, by determining the musical properties of arousal and valence at different moments of the audio track, in such a way that from these properties it was possible to establish an emotion in the plane of the circumflex, or Russell, model. This study intends to serve as a reference regarding the application of affective computing in the design of the messages and emotions transmitted in advertising jingles.

The approach proposed in this work represents a contribution to the state of the art regarding the application of affective computing in the analysis of commercial advertising jingles and in the political context, taking advantage of the benefits provided by the acoustic properties of arousal and valence, in terms of determining the emotions of an audio fragment according to Russell's model. In this sense, the study developed in this article was supported by a tool called MUSEMAN, which allows the automation of the process of analysis of emotions in acoustic multimedia content, through the segmentation of the advertising jingle analyzed in a set of fragments to each of which the associated emotion is determined from the acoustic properties of arousal and valence, in order to obtain the trace of emotions that make up the jingle. Thus, the tools and libraries considered in the proposed tool can serve as a reference for the extrapolation of this research in the analysis of multimedia content in other application areas.

The analysis of emotions from the study of musical properties such as arousal and valence corresponds to a non-subjective technique, since it depends on acoustic characteristics such as time, intensity, volume, rhythm, etc. This makes this method useful both for studying the design of the messages transmitted in advertising jingles and for analyzing the perception or level of user satisfaction.

The MUSEMAN tool proved to be adequate in terms of the analysis of the emotions present in an audio track, since it allows to carry out the segmentation processes, obtaining the arousal and valence properties, calculating the emotions associated with each segment, generating reports and display of follow-up and distribution graphs. In this sense, the MUSEMAN tool takes advantage of the advantages provided by the FFmpeg and openEAR libraries, acting in the background on the Linux operating system.

The study carried out showed that the candidates used different communication strategies in their advertising messages. On the one hand, three candidates transmitted the emotion of euphoria in their messages through the use of adapted songs (García and Wong) and through the socialization of the proposals in the streets of Cartagena (Araujo). On the other hand, the three remaining candidates recounted stories with a reflection on the problems of corruption and the problem of inequality in the city, relying on emotions such as anger and sadness.

As a future work derived from the present research, we intend to complement and contrast the results obtained by applying sentiment and text analysis techniques to the messages included in the advertising jingles.

Acknowledgment

The authors would like to thank the Universidad of Cartagena-Colombia and the Universidad of Cauca-Colombia for their support in the development of this research.

References

[1] A. López-Giraldo, “La efectividad de los jingles y su poder de recordación,” Razón y Palabra, vol. 21, no. 4_99 SE-Monográfico, pp. 197–211, Dec. 2017, [Online]. Available: https://revistarazonypalabra.org/index.php/ryp/article/view/1086.

[2] G. Chanchí and A. Córdoba, “Análisis de emociones y sentimientos sobre el discurso de firma del acuerdo de paz en Colombia,” Rev. Ibérica Sist. e Tecnol. Informação, no. E22, pp. 95–107, 2019.

[3] M. Zheng, Y. She, F. Liu, J. Chen, Y. Shu, and J. Xiahou, “BabeBay-A Companion Robot for Children Based on Multimodal Affective Computing,” ACM/IEEE Int. Conf. Human-Robot Interact., vol. 2019-March, pp. 604–605, Mar. 2019, doi: 10.1109/HRI.2019.8673163.

[4] J. Yang, Y. Xue, Z. Zeng, and W. Guo, “Research on multimodal affective computing oriented to online collaborative learning,” in Proceedings - IEEE 19th International Conference on Advanced Learning Technologies, ICALT 2019, Jul. 2019, pp. 137–139, doi: 10.1109/ICALT.2019.00045.

[5] O. O. Rudovic, “Machine learning for affective computing and its applications to automated measurement of human facial affect,” in 2016 International Symposium on Micro-NanoMechatronics and Human Science (MHS), Nov. 2016, pp. 1–1, doi: 10.1109 / MHS.2016.7824242.

[6] P. Baranyi and Á. Csapó, “Definition and Synergies of Cognitive Infocommunications,” Acta Polytech. Hungarica, vol. 9, no. 1, pp. 67–83, 2012, [Online]. Available: http://acta.uni-obuda.hu/Baranyi_Csapo_33.pdf.

[7] D. C. Ong, H. Soh, J. Zaki, and N. D. Goodman, “Applying Probabilistic Programming to Affective Computing,” IEEE Trans. Affect. Comput., vol. 12, no. 2, pp. 306–317, Apr. 2021, doi: 10.1109/TAFFC.2019.2905211.

[8] G. Reyes Zambrano and J. M. Banchón, “Affective computing and analysis of consumer behavior,” Int. J. Innov. Appl. Stud., vol. 20, no. 2, pp. 551–559, 2017, [Online]. Available: http://www.ijias.issr-journals.org/abstract.php?article=IJIAS-16-304-02.

[9] J. Dharmapriya, L. Dayarathne, T. Diasena, S. Arunathilake, N. Kodikara, and P. Wijesekera, “Music emotion visualization through colour,” in 2021 International Conference on Electronics, Information, and Communication, ICEIC 2021, Jan. 2021, pp. 1–6, doi: 10.1109/ICEIC51217.2021.9369788.

[10] R. Niese, A. Al-Hamadi, M. Heuer, B. Michaelis, and B. Matuszewski, “Machine vision based recognition of emotions using the circumplex model of affect,” in 2011 International Conference on Multimedia Technology, ICMT 2011, 2011, pp. 6424–6427, doi: 10.1109/ICMT.2011.6001887.

[11] J. A. Russell, “A circumplex model of affect,” J. Pers. Soc. Psychol., vol. 39, no. 6, pp. 1161–1178, Dec. 1980, doi: 10.1037/H0077714.

[12] M. Chmulik, R. Jarina, M. Kuba, and E. Lieskovska, “Continuous music emotion recognition using selected audio features,” in 2019 42nd International Conference on Telecommunications and Signal Processing, TSP 2019, Jul. 2019, pp. 589–592, doi: 10.1109/TSP.2019.8768806.

[13] P. Bustamente, N. López, and O. L. Quintero, "Reconocimiento y Regionalización de las Emociones en el Plano Excitación-Valencia," 2015, doi: 10.13140/RG.2.1.3188.6164.

[14] K. R. Tan, M. L. Villarino, and C. Maderazo, "Automatic music mood recognition using Russell's twodimensional valence-arousal space from audio and lyrical data as classified using SVM and Naïve Bayes," IOP Conf. Ser. Mater. Sci. Eng., vol. 482, no. 1, pp. 1–6, Feb. 2019, doi: 10.1088/1757-899X/482/1/012019.

[15] L. A. Solarte Moncayo, M. Sánchez Barragán, G. E. Chanchí Golondrino, D. F. Duran Dorado, and J. L. Arciniegas Herrera, "Dataset de contenidos musicales de video, basado en emociones," Ing. USBMed, vol. 7, no. 1, pp. 37–46, Jun. 2016, doi: 10.21500/20275846.2460.

[16] M. Russo, L. Kraljević, M. Stella, and M. Sikora, "Cochleogram-based approach for detecting perceived emotions in music," Inf. Process. Manag., vol. 57, no. 5, p. 102270, Sep. 2020, doi: 10.1016/J.IPM.2020.102270.

[17] M. Sánchez-Barragán, G. Chanchí-Golondrino, and W. Campo-Muñoz, "Sistema de recomendación para contenidos musicales basado en el análisis afectivo del contexto social," Rev. Ibérica Sist. e Tecnol. Informação, no. 39, pp. 100–113, 2020.

[18] S. Panwar, P. Rad, K.-K. R. Choo, and M. Roopaei, "Are you emotional or depressed? Learning about your emotional state from your music using machine learning," J. Supercomput. 2018 756, vol. 75, no. 6, pp. 2986–3009, Jul. 2018, doi: 10.1007/S11227-018-2499-Y.

[19] R. Thayer, The Biopsychology of Mood and Arousal. Oxford University Press, 1990.

[20] J. García García, P. Henríquez-Coronel, J. Pincai Ponce, and J. Herrera-Tapia, "Analítica de Twitter para el estudio de las emociones primarias durante el terremoto de México 2017," Rev. Ibérica Sist. e Tecnol. Informação, no. E19, pp. 479–492, 2019.

[21] M. V. Mäntylä, D. Graziotin, and M. Kuutila, "The evolution of sentiment analysis—A review of research topics, venues, and top cited papers," Comput. Sci. Rev., vol. 27, pp. 16–32, Feb. 2018, doi: 10.1016/j.cosrev.2017.10.002.

[22] I. Mosquera-Cabrera, "Influencia de la música en las emociones: una breve revisión," Realitas, vol. 1, no. 2, pp. 34–38, 2013, [Online]. Available: https://dialnet.unirioja.es/servlet/articulo?codigo=4766791.

[23] Y.-H. Yang and H. Chen, "Machine Recognition of Music Emotion," ACM Trans. Intell. Syst. Technol., vol. 3, no. 3, pp. 1–30, May 2012, doi: 10.1145/2168752.2168754.

[24] S. Rashmi, M. Hanumanthappa, and V. Kavitha, "An Invasion to Human - Computer Interaction: Stages of Speech Recognition Process using Speech Processing Techniques," in ICSNS 2018 - Proceedings of IEEE International Conference on Soft-Computing and Network Security, Dec. 2018, pp. 1–4, doi: 10.1109/ICSNS.2018.8573654.

[25] F. Eyben, M. Wöllmer, and B. Schuller, "OpenEAR - Introducing the Munich open-source emotion and affect recognition toolkit," in Proceedings - 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, ACII 2009, 2009, pp. 1–6, doi: 10.1109/ACII.2009.5349350.

[26] K. Pratt, "Design Patterns for Research Methods: Iterative Field Research," 2009. [Online]. Available: http://kpratt.net/wp-content/uploads/2009/01/research_methods.pdf.

[27] G. Chanchí, L. M. Sierra, and W. Y. Campo, "Propuesta de un analizador automático de emociones para contenidos musicales," Rev. Ibérica Sist. e Tecnol. Informação, no. E26, pp. 558–570, 2020.

An Intelligent Approach based on the Combination of the Discrete Wavelet Transform, Delta Delta MFCC for Parkinson's Disease Diagnosis

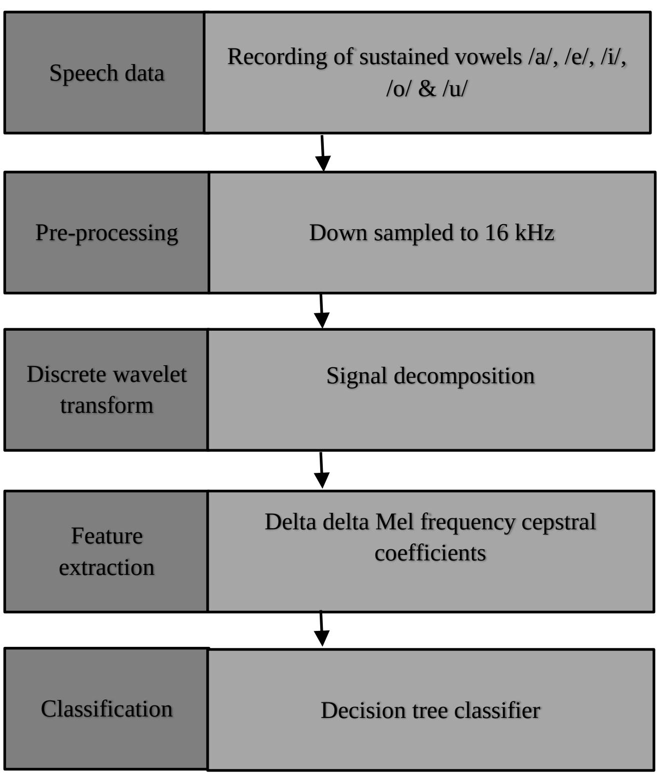
BOUALOULOU Nouhaila[1], BELHOUSSINE DRISSI Taoufiq[2]
Laboratory Electrical and Industrial Engineering, Information Processing, Informatics, and Logistics (GEITIIL), Faculty of Science Ain Chock. University Hassan II, Casablanca, Morocco

NSIRI Benayad[3]
Research Center STIS, M2CS, National Higher School of Arts and Craft, Rabat (ENSAM)
Mohammed V University in Rabat
Morocco

Highlights

- Collection of several types of speech recordings of the vowels /a/, /e/, /i/, /o/ and /u/.
- The decomposition of each voice signal by DWT using the different types of wavelets.
- Extraction of the delta delta MFCC from all voice samples.
- Classification using decision tree classifier along with holdout scheme

Graphical Abstract:



***Abstract*—To diagnose Parkinson's disease (PD), it is necessary to monitor the progression of symptoms. Unfortunately, diagnosis is often confirmed years after the onset of the disease. Communication problems are often the first symptoms that appear earlier in people with Parkinson's disease. In this study, we focus on the signal of speech to discriminate between people with and without PD, for this, we used a Spanish database that contains 50 records of which 28 are patients with Parkinson's disease and 22 are healthy people, these records contain five types of supported vowels (/a/, /e/, /i/, /o/ and /u/), The proposed treatment is based on the decomposition of each sample using Discrete Wavelet Transform (DWT) by testing several kinds of wavelets, then extracting the delta delta Mel Frequency Cepstral Coefficients (delta delta MFCC) from the decomposed signals, finally we apply the decision tree as a classifier, the purpose of this process is to determine which is the appropriate wavelet analyzer for each type of vowel to diagnose Parkinson's disease.**

Keywords—Parkinson's disease; discrete wavelet transform; delta delta MFCC; decision tree classifier

I. Introduction

Parkinson's disease is a severe health problem. According to the American Parkinson's Disease Association (APDA) [1], more than 10 million people worldwide are affected by Parkinson's disease. Due to the loss of certain groups of brain cells that produce neurotransmitters, including dopamine, causes symptoms such as impaired speech, movement, and sleep, as well as panic and anxiety attacks.

Speech disorders include reduced speech intensity, fluctuating fundamental frequency, and irregular speech articulation which are signs that appear early in people with Parkinson's disease, allowing many studies to use the speech signal for the identification of Parkinson's disease [2-5].

However, the speech signal is one of the most complex signals to characterize, which makes it difficult to develop a system to understand different diseases such as Parkinson's disease [6-8], Alzheimer's disease [9-11], and COVID 19 [12-15], etc. This complexity of the speech signal comes from a combination of several factors, the redundancy of the acoustic signal, the high inter-and intra-speaker variability, the effects of coarticulation in continuous speech, and the recording conditions. To overcome these difficulties, many methods and mathematical models have been developed, including neural networks [16, 17], Support Vector Machines (SVM) [18, 19], stochastic Markov models, and in particular Hidden Markov Models (HMM) [20, 21].

These methods and models work from information extracted from the speech signal considered relevant. This extraction is performed by an acoustic analysis which leads to gathering this information under the term of the vector of acoustic parameters whose dimension and nature are determinant to reach good performances of the knowledge system of Parkinson's disease which is of interest to us in this article. The different types of acoustic parameters commonly cited in the literature are the coefficients: LPC [22], LPCC [23], LFCC [24], PLP [23], MFCC [6, 23], etc. Generally, the MFCC coefficients are the most used acoustic parameters in speech feature extraction [25-27].

However, for these systems of recognition of Parkinson's disease, research work has studied the improvement of the performance of this system by combining the MFCC coefficients with other types of acoustic parameters such as LPCC [28], PLP [28], energy [29], wavelets [29-31] and Empirical Mode Decomposition (EMD) [32].

In this study, far from static coefficients, for correct detection of Parkinson's disease through voice our contribution consists in proposing a new method of selection of relevant acoustic parameters based on the use of dynamic delta delta MFCC coefficients combined with wavelets, these differential coefficients referred as dynamic parameters provide useful information on the temporal trajectory of the speech signal. This information extracted by discrete wavelet transform and delta delta MFCC will be used in the classification block using a decision tree classifier. To evaluate the performance of this model we applied it to a database of five vowels (/a/, /e/, /i/, /o/, /u/), each vowel includes 28 people with Parkinson's disease and 22 are healthy people.

The structure of the rest of the article is as follows: section II concerns a definition of the methods used, section III presents a description of the process used and the results obtained, and finally the conclusion in section IV.

II. METHOD

A. Feature Extraction

For feature extraction, we are interested in the joint use of DWT and the second derivative of the MFCC (delta delta MFCC). This step allows extracting features that will be used by the decision tree classifier.

1) Discrete wavelet transform: For the discrete wavelet transform, it is the discrete version of the Continuous Wavelet Transform (CWT) that used Mallat's algorithm, it is based on the principle of multi-resolution that allows the separation of details and approximations of signals by using a pair of filters H and G that constitute respectively a low-pass filter and a high-pass filter.

With the high-pass filters, we obtain the coefficients of the discrete wavelet decomposition (the details), and with the low-pass filters, we obtain the approximation coefficients. This operation is applied again to the approximation, generating another detail and a new approximation as shown in Figure 1. There are several types of wavelets, in our case; we used the wavelets presented in the table I:

2) Delta delta MFCC: Temporal changes in the cepstrum play an important role in human perception and it is through the derivatives of MFCC coefficients that we can measure these changes.

TABLE I. WAVELET FAMILIES USED

Wavelets families	**Wavelets**
Daubechies	db1, db2, and db3
Coiflets	coif1, coif2, and coif3
Symplets	sym1, sym2, and sym3
Discrete Meyer	dmey

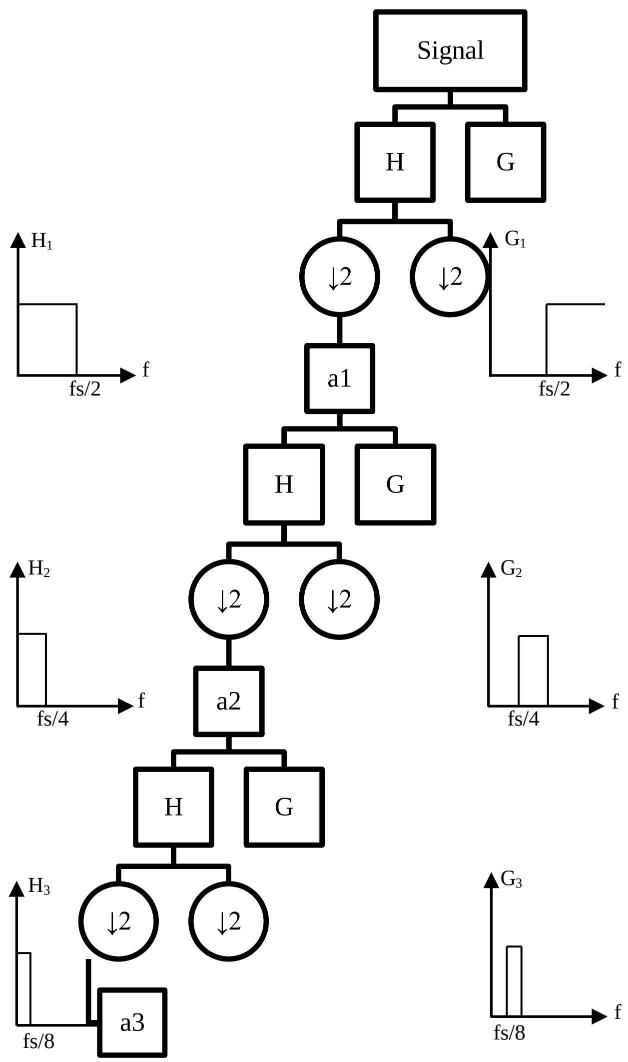


Fig. 1. Representation of the Mallat Algorithm for the Implementation of the DWT Decomposition, H is a Low Pass Filter and G is a High Pass Filter.

Usually, MFCC coefficients are referred to as static parameters, since they contain only the information about a given frame. To improve the frame representation, it is often proposed to introduce new parameters into the parameter vector. The reference [33] proposed the use of dynamic parameters that present cepstral transition information in the speech signal. In particular, it proposed second-order differential coefficients, also called delta delta coefficients, derived from cepstral coefficients. Let d_t be the first-order differential coefficient (delta MFCC) of frame t, then the corresponding second-order differential coefficient (delta delta MFCC) dd_t is calculated by the following formula:

$$dd_t = \frac{\sum_{n=1}^{N} n\left(d_{t+n} - d_{t-n}\right)}{2\sum_{n=1}^{N} n^2} \tag{1}$$

The coefficients delta delta MFCC, also called acceleration coefficients are obtained using the second derivative of static Mel Frequency Cepstral Coefficients (MFCC), the latter is a representation defined as the discrete cosine transform of the logarithm of the spectrum of the energy of the speech segment. The spectral energy is calculated by applying a bank of evenly spaced filters on a modified frequency scale, called the Mel scale. The Mel scale redistributes the frequencies in a non-linear scale that simulates human perception of sounds. Figure 2 illustrates the steps involved in obtaining the delta delta MFCC coefficients. Based on the results obtained in [26], we use only the first 12 delta delta MFCC coefficients.

B. Feature Classification

1) The decision tree classifier: A decision tree is one of the most popular techniques in machine learning. Indeed, decision tree learning is part of supervised learning, it is generally a classifier presented in the form of a tree structure [34].

A decision tree consists of a set of rules allowing a segment of a data set into homogeneous groups. Each rule associates the conjunction of tests with the descriptive variables. The first vertex is called the root of the tree, the following variables, which correspond to non-terminal nodes, are segmentation variables; each branch corresponds to a modality of the variable considered at this level of the tree. This process is repeated on each node of the tree, the nodes that are not pure are segmented until pure leaves are obtained.

Here we try to classify a population of individuals containing healthy people and people with Parkinson's disease pronounce these five vowels (/a/, /e/, /i/, /o/, /u/) into two classes with respect to a label {1 (healthy), 0 (sick)} from the recordings. The decision tree-learning algorithm is described below:

Algorithm 1: Decision tree

Data: a sample Ω of m labeled records

Initialization: empty tree; current node: root; current sample: Ω

- Repeat
- Decide whether the current node is terminal
- If the current node is terminal then
- Label the current node with a leaf
- Otherwise
- Select a test and create the subtree
- End if
- Current node: a node not yet studied
- Current sample: sample reaching the current node
- Until a decision tree is produced

Output: decision tree

---

2) Holdout method: All classification results are obtained using the "holdout" method. This method is a commonly used practice for evaluating machine-learning models. It works by first dividing the data randomly into two parts; one of larger size is used for training and the other part is reserved for error rate estimation. Another version of this method, called "data shuffle", consists in repeating L times the random division of the data into two parts; one for training and the other for testing, and then calculating the average of the L estimates of the error rates evaluated on the test data parts. The advantage of this method is that all data are used for both training and testing. The holdout is a simple method to understand and generally results in a less biased model estimate than other methods.

To assess classifier performance, the following parameters are used:

$$\text{Accuracy} = \frac{TN + TP}{TN + TP + FP + FN} \tag{2}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{3}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{4}$$

Where TP: true positive; TN: true negative; FP: false positive; FN: false negative.

III. Methodology and Result

We have implemented our algorithm in Matlab version R2019a. The tests were performed on a PC with the following configuration:

CPU: Intel(R) Core(TM) i3-3110M CPU @ 2.40GHz

Memory: 4,00 Go

Operating System: Windows 10 64bit.

A. Data Collection

This study takes into account two subgroups of this group: the Healthy Control (HC) group with 22 speakers and the Parkinson's disease (PD) group with 28 speakers. All utterers included in this Italian corpus were registered in Bari (Puglia region), Italy. Each recording session took place in a controlled environment, taking into account factors such as room temperature, distance from the microphone, time of day, and having a conversation with the subject to warm up their vocal muscles. The sampling frequency was 16 kHz; more information is available in [35]. Table II includes the demographic information of the corpus.

B. Methodology

The objective of this part of Parkinson's disease detection is to design and experiment with a system based on extensive use of wavelet types combined with acceleration coefficients (delta delta MFCC) as features and the decision tree as a classifier, as illustrated in Figure 3.

For the feature selection and extraction domain, which is an important step in the recognition process of Parkinson's disease, we consider the use of the wavelet transform, which is a time and frequency analysis tool that allows obtaining variable temporal and frequency resolutions.

TABLE II. AGE DEMOGRAPHIC DATA FROM THE ITALIAN CORPUS, ARRANGED BY GENDER AND CLASS

	Female		**Male**		**Total**	
	PD	***HC***	***PD***	***HC***	***PD***	***HC***
Subjects	9	12	19	10	28	22
Age range	40-80	60-72	50-77	60-77	40-80	60-77

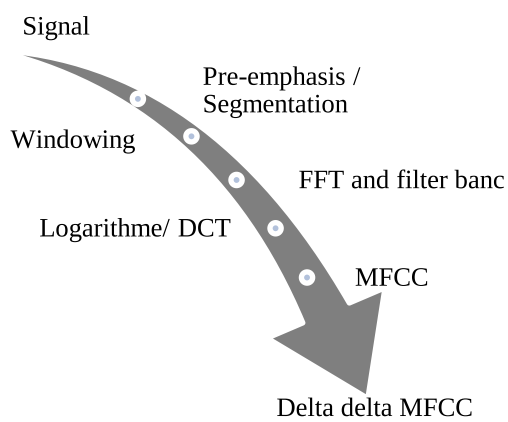


Fig. 2. Delta Delta MFCC Extraction Process.

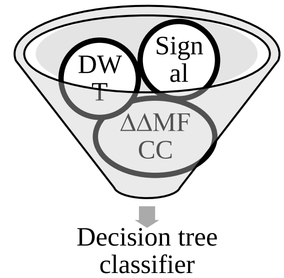


Fig. 3. The Process used to Diagnose Parkinson's Disease.

Our model for the diagnosis of Parkinson's disease will be applied to a corpus containing 28 subjects with Parkinson's disease and 22 healthy subjects for each vowel (/a/, /e/, /i/, /o/, /u/). It is based firstly on the decomposition of the signal into two sub-bands of approximate frequency and details by the use of a wide range of wavelets (Debauchies, Coiflets, Symlets, and Discrete Meyer) on the first seven scales.

Susceptible to better characterize our system, extracting the first 12 coefficients delta delta MFCC that are parameters widely used to encode the dynamic information of cepstral parameters. This extraction is only performed on the low-frequency band (approximation) for each type of wavelet and from the first scale up to the seventh scale.

The main objective of the feature selection step is its direct contribution to the performance of the overall system this step allowed us to extract the features that will be then used by the decision tree classifier with the cross-validation method "holdout" that serves to decompose the base of each vowel (/a/, /e/, /i/, /o/, /u/) into 80% as a learning base and the whole base as a test base.

C. Experiments and Results

In this part of our work, we implemented our Parkinson's disease detection system where we introduced the DWT method and the delta delta MFCC coefficients as a feature vector, we changed the wavelet types for each first seven scales, then we measured the impact of each of these changes on the Parkinson's disease recognition performance by the decision tree classifier performance evaluation metrics.

A comparative study between the extracted features based on delta delta MFCCs is performed to visualize the distinctive behavior of the subject with Parkinson's disease and the healthy subject for the five vowels. Figures 4-8 show the 12 delta delta MFCC coefficients as extracted features for a healthy person and a person with Parkinson's disease. The feature patterns are obtained from 50 speech samples for each vowel (/a/, /e/, /i/, /o/, /u/) from people with Parkinson's and healthy people. Almost all speech signals follow the same pattern. Figure 4 clearly distinguishes the variation in the speech pattern of a healthy person and a person with Parkinson's disease for the vowel /a/ based on the wavelet of discrete Meyer at scale 5 the latter is the one that gave us a better Accuracy. Figures 5, 6, 7, and 8 present the delta delta MFCC coefficients respectively for the vowels /e/, /i/, /o/, and /u/, we based on these wavelets with these scales respectively discrete Meyer at scale 5, Coiflets level 3 at scale 6, discrete Meyer at scale 7 and discrete Meyer at scale 6 because it is those that gave us important results that are displayed in Tables III, IV, V, VI, and VII a notable variation is observed between the healthy speech signal and that of Parkinson's disease. Thus, the proposed feature-based on delta delta MFCC can be a good marker for the prediction of Parkinson's disease.

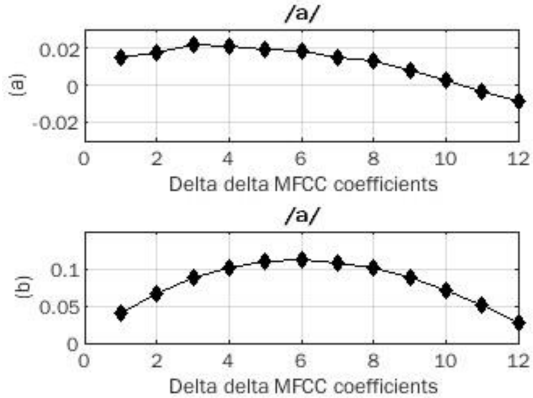


Fig. 4. Delta Delta Coefficients Extracted from the Vowel /a/ for: (a) Healthy Person, (b) Person with Parkinson's Disease.

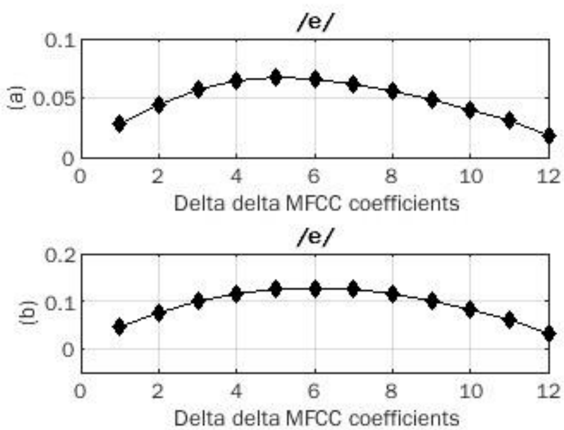


Fig. 5. Delta Delta Coefficients Extracted from the Vowel /e/ for: (a) Healthy Person, (b) Person with Parkinson's Disease.

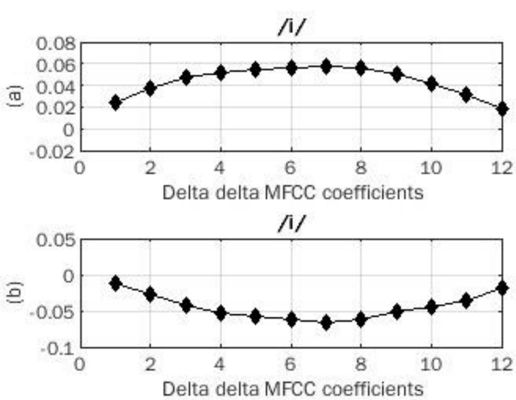


Fig. 6. Delta Delta Coefficients Extracted from the Vowel /i/ for: (a) Healthy Person, (b) Person with Parkinson's Disease.

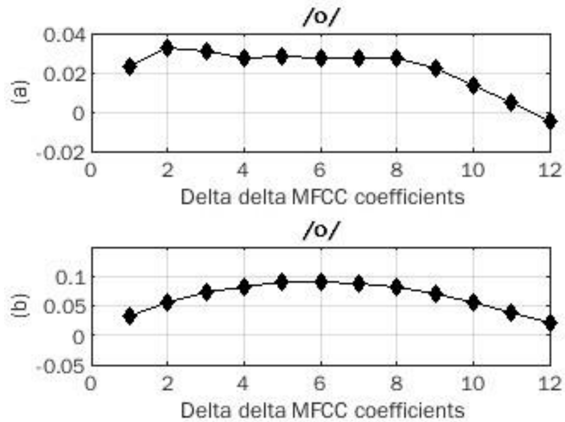


Fig. 7. Delta Delta Coefficients Extracted from the Vowel /o/ for: (a) Healthy Person, (b) Person with Parkinson's Disease.

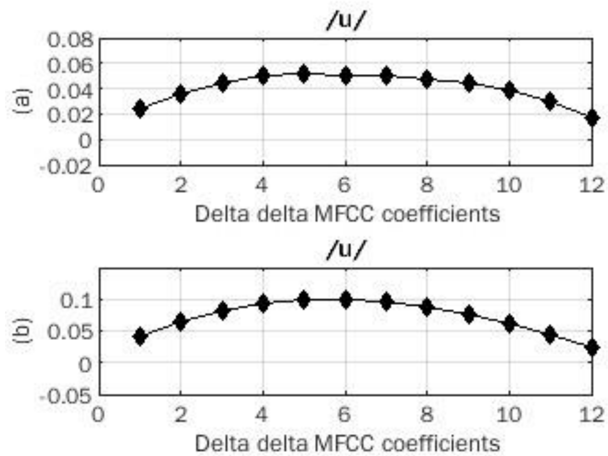


Fig. 8. Delta Delta Coefficients Extracted from the Vowel /u/ for: (a) Healthy Person, (b) Person with Parkinson's Disease.

The following tables show the different results obtained for the five vowels according to accuracy, specificity, and sensitivity.

For the vowel /a/, the results obtained in Table III show that the wavelet by using discrete Meyer at scale 5 offers better modeling for the classification of healthy people and people with Parkinson's disease with high accuracy of 97.5%. While for the vowel /e/, Table IV is still with the same wavelet, and on the same scale of 5, the best accuracy of 92.5% is achieved.

In contrast to the experiment for the vowel /i/, in Table V we register a slight performance low of the accuracy of 87,5% by using the wavelet of Coiflets level 3 at scale 6 and discrete Meyer at scale 7.

For the vowel /o/, according to the results shown in Table VI, we find that the wavelet of discrete Meyer at scale 7 gives accuracy up to 92,5% for the discrimination between healthy patients and patients with Parkinson's disease. For the vowel /u/, Table VII shows an important result with an accuracy of 90% by using the wavelet of discrete Meyer at scale 6.

TABLE III. ACCURACY, SPECIFICITY, AND SENSITIVITY FOR THE VOWEL /A/ USING THE DECISION TREE CLASSIFIER

	Accuracy							**Specificity**							**Sensitivity**						
Scale wavelet	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Db1	67,5	67,5	62,5	77,5	85	83	80	67,9	32,1	53,6	82,1	71,4	63,6	50	4.5	72,7	45,5	63,6	63,6	75	82,1
Db2	65	67,5	70	77,5	90	72,5	72,5	31,8	31,8	68,2	25	72,7	72,7	45,5	68,9	45,5	39,3	77,3	67,9	60,7	75
Db3	67,5	65	70	77,5	90	87,5	77,5	72,7	40,9	63,6	45,5	81,8	68,2	59,1	32,1	32,1	53,6	78,6	35,7	71,4	64,3
Coif1	70	65	75	80	90	89,3	77,5	50	9,1	40,9	36,4	72,7	68,2	75	50	89,3	82,1	82,1	78,6	78,6	50
Coif2	65	65	77,5	77,5	95	87,5	90	68,2	68,2	63,6	13,6	59,1	72,7	36,4	35,7	46,4	67,9	85,7	89,3	78,6	92,9
Coif3	70	65	70	80	95	85	80	45,5	68,2	59,1	40,9	63,6	68,2	72,7	53,6	42,9	64,3	85,7	92,9	67,9	57,1
Sym1	67,5	67,5	65	77,5	90	77,5	72,5	45,5	45,5	63,6	4,5	36,4	36,4	68,2	64,3	71,4	35,7	75	85,7	25	60,7
Sym2	67,5	67,5	70	77,5	90	87,5	70	59,1	72,7	36,4	50	68,2	77,3	68	53,6	42,9	67,9	82,1	82,1	67,9	72
Sym3	67,5	65	70	80	90	80	80	40,9	72,7	59,1	13,6	63,6	68,2	77,3	67,9	39,3	60,7	85,7	85,7	50	28,6
dmey	65	65	67,5	82,5	97,5	82,5	95	27,3	54,5	36,4	54,5	54,5	68,2	59,1	71,4	50	75	82,1	96,4	71,4	89,3

TABLE IV. ACCURACY, SPECIFICITY, AND SENSITIVITY FOR THE VOWEL /E/ USING THE DECISION TREE CLASSIFIER

	Accuracy							**Specificity**							**Sensitivity**						
Scale wavelet	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Db1	65	62,5	62,5	62,5	87	80	77,5	18,2	72,7	59,1	77,3	68,2	77,3	68,2	75	35,7	35,7	42,9	67,9	64,3	
Db2	62,5	60	60	60	85	80	67,5	45,5	72,7	50	54,5	81,8	81,8	77,3	93,3	28,6	42,9	53,6	46,6	75	39,3
Db3	62,5	57,5	65	62,5	82,5	80	70	9,1	4,5	50	18,2	72,7	77,3	77,3	78,6	89,3	46,4	64,3	42,9	67,9	42,9
Coif1	65	60	57,5	62,5	85	80	70	13,6	50	50	13,6	81,8	77,3	77,3	85,7	53,6	50	89,3	60,7	67,9	35,7
Coif2	62,5	60	60	62,5	77,5	77,5	70	40,9	72,7	40,9	28,6	72,7	63,6	45,5	75	35,7	53,6	72,7	64,3	64,3	57,1
Coif3	65	62,5	62,5	62,5	82,5	82,5	77,5	77,3	54,5	72,7	31,8	68,2	81,8	77,3	32,1	64,4	39,3	60,7	74,1	46,4	53,4
Sym1	65	60	60	60	87,5	80	77,5	9,1	40,9	27,3	72,7	4,5	72,7	63,6	75	53,6	64,3	35,7	82,1	71,4	32,1
Sym2	62,5	57,5	67,5	62,5	82,5	80	67,5	54,5	72,7	50	18,2	72,7	72,7	54,5	42,9	32,1	46,5	70	64,3	67,9	60,7
Sym3	65	57,5	65	62,5	80	80	70	40,9	72,7	59,1	68,2	77,3	63,6	68,2	67,9	35,7	46,4	32,1	42,9	46,4	64,3
dmey	70	70	62,5	60	92,5	82,5	80	45,5	45,5	68,2	54,5	72,7	36,4	68,2	78,6	78,6	39,3	32,5	35,7	85,7	89,3

TABLE V. ACCURACY, SPECIFICITY, AND SENSITIVITY FOR THE VOWEL /I/ USING THE DECISION TREE CLASSIFIER

	Accuracy							**Specificity**							**Sensitivity**						
Scale wavelet	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Db1	77,5	75	65	72,5	77,5	77,5	72,5	54,5	54,5	63,6	63,6	72,7	59,1	77,3	67,9	71,4	53,6	60,7	57,1	57,1	42,9
Db2	77,5	75	65	70	77,5	65	75	59,1	13,6	18,2	77,3	54,5	45,5	68,2	53,6	85,7	75	21,4	67,9	53,6	60,7
Db3	77,5	75	62,1	65	75	72,5	67,5	31,8	4,5	50	59,1	77,3	63,6	45,5	78,6	67,9	46,4	57,1	21,4	53,7	60,7
Coif1	75	67,5	70	75	70	75	75	54,5	63,6	77,3	72,7	54,5	81,8	77,3	71,4	60,7	39,3	50	57,1	53,6	42,9
Coif2	80	75	67,5	60	72,5	77,5	65	59,1	13,6	18,2	77,3	54,5	63,6	54,5	53,6	85,7	75	21,4	67,9	46,4	57,1
Coif3	77,5	75	67,5	67,5	70	**87,5**	77,5	45,5	50	36,4	50	50	81,8	63,6	89,3	67,9	57,1	53,6	46,4	42,9	85,7
Sym1	77,5	75	65	70	77,5	80	77,5	50	9,1	54,5	63,6	77,3	81,8	63,6	85,7	60,7	57,1	67,9	39,3	35,7	67,9
Sym2	77,5	75	65	70	77,5	72,5	65	59,1	50	22,7	68,2	77,3	72,7	81,8	75	67,9	67,9	46,4	39,3	53,6	46,4
Sym3	77,5	75	62,5	67,5	75	75	67,5	59,1	50	27,3	50	72,7	22,7	13,6	75	67,9	71,4	67,9	46,4	67,9	71,4
dmey	81,33	75	67,5	65	77,5	77,5	**87,5**	13,6	63,6	63,6	68,2	59,1	40,9	68,2	82,1	60,7	39,3	39,3	67,9	50	78,6

TABLE VI. ACCURACY, SPECIFICITY, AND SENSITIVITY FOR THE VOWEL /O/ USING THE DECISION TREE CLASSIFIER

	Accuracy							**Specificity**							**Sensitivity**						
Scale wavelet	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Db1	82,5	75	70	67,5	75	77,5	77,5	63,6	63,6	18,2	68,2	68,2	81,8	50	64,3	57,1	67,9	46,4	60,7	32,1	75
Db2	80	75	70	70	75	82,5	77,5	77,3	54,5	81,8	63,6	68,2	54,5	63,6	35,7	71,4	53,6	50	57,1	64,3	67,9
Db3	80	72,5	70	65	77,5	70	75	22,7	59,1	59,1	63,6	59,1	36,4	77,3	85,7	64,3	39,3	25	53,6	75	46,4
Coif1	77,5	72,5	70	67,5	75	77,5	72,5	63,6	54,5	81,1	54,5	77,3	72,7	81,8	60,7	32,1	35,7	50	35,5	60,7	32,1
Coif2	72,5	75	75	65	77,5	80	67,5	18,2	72,7	72,7	63,6	54,5	81,8	31,8	89,3	35,7	53,6	42,9	67,9	64,3	60,7
Coif3	75	90	72,5	72,5	75	82,5	79	63,6	40,9	63,6	68,2	72,7	77,3	63,6	60,7	85,7	50	60,7	50	53,6	67,9
Sym1	84	75	70	70	75	77,5	72,5	54,5	63,6	72,7	63,6	72,7	77,3	77,3	75	64,3	46,4	53,6	50	60,7	46,4
Sym2	80	75	70	70	77,5	82,5	77,5	54,5	77,3	45,5	59,1	40,9	77,3	59,1	64,3	21,4	64,3	46,4	64,3	67,9	64,3
Sym3	77,5	80	70	65	75	75	72,5	63,6	54,5	40,9	9,1	81,8	81,8	81,8	67,9	71,4	60,7	64,3	46,4	64,3	39,3
dmey	72,5	72,5	70	65	77,5	80	**92,5**	40,9	40,9	77,3	40,9	72,7	45,5	72,7	78,6	28,6	39,3	57,1	64,3	85,7	85,7

TABLE VII. ACCURACY, SPECIFICITY, AND SENSITIVITY FOR THE VOWEL /U/ USING THE DECISION TREE CLASSIFIER

	Accuracy							**Specificity**							**Sensitivity**						
Scale wavelet	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Db1	75	75	65	65	67,5	77,5	72,5	72,7	50	54,5	72,7	63,6	31,8	59,1	57,1	78,6	53,6	46,4	53,6	67,9	57,1
Db2	77,5	77,5	62,5	65	75	67,5	72,5	63,6	31,8	4,5	40,9	72,7	72,7	77,3	25	78,6	92,9	50	57,1	50	57,1
Db3	75	77,5	62,5	62,5	72,5	80	72,5	9,1	45,5	81,8	72,7	81,8	40,9	81,8	89,3	78,6	28,6	32,1	53,6	42,9	57,1
Coif1	77,5	77,5	62,5	65	72,5	72,5	70	50	54,5	40,9	68,2	50	81,8	45,5	78,6	35,7	57,1	53,6	57,1	57,1	60,7
Coif2	72,5	72,5	62,5	62,5	72,5	72,5	70	27,3	50	68,2	68,2	72,7	59,1	63,6	75	75	42,9	35,7	53,6	53,6	42,9
Coif3	72,5	67,5	62,5	62,5	70	80	72,5	63,6	59,1	68,2	68,2	72,7	81,8	77,3	60,7	39,3	50	35,7	57,1	53,6	57,1
Sym1	75	75	65	67,5	70	77,5	70	9,1	36,4	45,5	50	63,6	72,7	59,1	64,3	82,1	60,7	50	50	46,4	32,1
Sym2	77,5	77,5	62,5	65	72,5	67,5	72,5	59,1	54,5	68,2	4,5	52,2	72,7	27,3	67,9	60,7	46,4	82,1	85,7	35,7	71,4
Sym3	75	77,5	62,5	62,5	72,5	80	72,5	59,1	31,8	45,5	54,5	77,3	63,6	59,1	35,7	78,6	53,6	42,9	32,1	32,1	46,4
dmey	75	80	52,5	62,5	75	**90**	72,5	27,3	50	68,2	72,7	77,3	54,5	68,2	85,7	75	42,9	25	53,6	64,3	85,7

In Table VIII we have listed, but not exhaustively, the different works on Parkinson's disease recognition systems. The performances are difficult to compare because they vary according to several elements such as the type of data, the choice of the learning models, the way the parameters are obtained, etc. In this section, we will compare our results with those of some works.

TABLE VIII. COMPARISON OF PERFORMANCE WITH RECENTLY PUBLISHED WORK

Study	**Method**	**Accuracy**
Zayrit et al [29]	DWT-genetic algorithm and SVM	91,18 (k-fold cross-validation)
Benba et al [26]	MFCC- MLP kernel of SVM	82,50 (LOSO)
Sakar et al [36]	KNN+SVM	55,00 (LOSO on training database) 68,45 (LOSO on the testing database)
Belhoussine drissi et al [30]	DWT-MFCC-SVM	86,84
Eskidere et al [37]	Random subspace classifier	74,17 (10-fold CV)
Li et al [38]	SVM+FS	82,5
Ali et al [39]	Multimodal approach	70
Ali et al. [40]	LDA–NN–GA	95.00
Proposed method.	DWT-delta delta MFCC- decision tree classifier	vowel /a/ 97,5

Zayrit et al [29] evaluated the /a/ vowel in the Turkish corpus by a vector of 21 prosodic features including LPC, ZCR, energy, Shannon entropy, and MFCC. The recognition accuracy of PD using the SVM classifier was around 91.18%. The researchers in [26] also conducted experiments using a Turkish corpus where the accuracy was 82.50%. This accuracy rate was shown for cepstral features applied with the SVM classifier. Sakar et al [36] reported a PD recognition accuracy of 68.45% using prosodic features in an SVM-based classification. They used the vowel /a/ from the Turkish corpus. Belhoussine drissi et al [30] reported a recognition accuracy of 86.84% for PD using cepstral features in an SVM-based classification. They used a database of 38 recordings, 18 of which were from healthy individuals and 20 from patients with Parkinson's disease from the Turkic corpus. References [37], [38], [39], and [40] extracted the spectral, and prosodic features used respectively Random subspace classifier, Support Vector Machines (SVM) as classifiers, Multimodal approach and LDA-NN-GA classifiers, they achieved recognition accuracy of PD respectively 74.17%, 82.5%, 70%, and 95%. In this article, the results show that the joint use of the wavelets and the delta delta MFCC coefficients as features brings a significantly important improvement to the performance of the Parkinson's disease diagnostic system with an accuracy of up to 97,5%.

IV. CONCLUSION

The work presented in this article is part of a project to recognize Parkinson's disease from the voice, in an educational context. The objective is to detect the state of each person if he/she is healthy or suffering from Parkinson's disease. To achieve this goal, we first proposed a comprehensive and efficient system for automatic recognition of people's states from a Spanish corpus of the five sustained vowels (/a/, /e/, /i/, /o/, and /u/) produced by 28 subjects with Parkinson's disease and 22 healthy subjects, Our process starts with the transformation of the speech signals by several types of DWT based on the approximation of the first seven scales which will be injected into the delta delta MFCC block to extract the 12 coefficients at each time. These coefficients are applied in the classification using the decision tree classifier.

The results show that the proposed feature is superior, providing a maximum accuracy of 97.5% for the database that contains the vowel /a/. There is a significant improvement over recent studies. The complete study showed that the proposed combination of wavelets with delta delta MFCC could be used to effectively detect PD.

References

[1] Standaert, D. G., Saint-Hilaire, M. H., & Thomas, C. A. (2015). Parkinson’s disease handbook. American Parkinson Disease Association, New York, USA.

[2] Rusz, J., Tykalová, T., Novotný, M., Zogala, D., Růžička, E., & Dušek, P. (2022). Automated speech analysis in early untreated Parkinson's disease: relation to gender and dopaminergic transporter imaging. European Journal of Neurology, 29(1), 81-90.

[3] Mollaei, F., Shiller, D. M., Baum, S. R., & Gracco, V. L. (2022). Early auditory responses to speech sounds in Parkinson’s disease: preliminary data. Scientific reports, 12(1), 1-8.

[4] Tykalova, T., Novotny, M., Ruzicka, E., Dusek, P., & Rusz, J. (2022). Short-term effect of dopaminergic medication on speech in early-stage Parkinson’s disease. npj Parkinson's Disease, 8(1), 1-6.

[5] Ahmed, I., Aljahdali, S., Khan, M. S., & Kaddoura, S. (2022). Classification of parkinson disease based on patient’s voice signal using machine learning. Intelligent Automation and Soft Computing, 32(2), 705-722.

[6] Karan, B., Sahu, S. S., & Mahto, K. (2020). Parkinson disease prediction using intrinsic mode function based features from speech signal. Biocybernetics and Biomedical Engineering, 40(1), 249-264.

[7] Sakar, C. O., Serbes, G., Gunduz, A., Tunc, H. C., Nizam, H., Sakar, B. E., ... & Apaydin, H. (2019). A comparative analysis of speech signal processing algorithms for Parkinson’s disease classification and the use of the tunable Q-factor wavelet transform. Applied Soft Computing, 74, 255-263.

[8] Polat, K. (2019, April). A hybrid approach to Parkinson disease classification using speech signal: the combination of smote and random forests. In 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT) (pp. 1-3). Ieee.

[9] König, A., Satt, A., Sorin, A., Hoory, R., Toledo-Ronen, O., Derreumaux, A., ... & David, R. (2015). Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 1(1), 112-124.

[10] Pulido, M. L. B., Hernández, J. B. A., Ballester, M. Á. F., González, C. M. T., Mekyska, J., & Smékal, Z. (2020). Alzheimer's disease and automatic speech analysis: a review. Expert systems with applications, 150, 113213.

[11] Pappagari, R., Cho, J., Joshi, S., Moro-Velázquez, L., Zelasko, P., Villalba, J., & Dehak, N. (2021). Automatic detection and assessment of Alzheimer Disease using speech and language technologies in low-resource scenarios. In Proc. Interspeech (pp. 3825-3829).

[12] Deshpande, G., & Schuller, B. (2020). An overview on audio, signal, speech, & language processing for covid-19. arXiv preprint arXiv:2005.08579.

[13] Dash, T. K., Mishra, S., Panda, G., & Satapathy, S. C. (2021). Detection of COVID-19 from speech signal using bio-inspired based cepstral features. Pattern Recognition, 117, 107999.

[14] Deshpande, G., Batliner, A., & Schuller, B. W. (2022). AI-Based human audio processing for COVID-19: A comprehensive overview. Pattern recognition, 122, 108289.

[15] Quatieri, T. F., Talkar, T., & Palmer, J. S. (2020). A framework for biomarkers of COVID-19 based on coordination of speech-production subsystems. IEEE Open Journal of Engineering in Medicine and Biology, 1, 203-206.

[16] Hireš, M., Gazda, M., Drotár, P., Pah, N. D., Motin, M. A., & Kumar, D. K. (2022). Convolutional neural network ensemble for Parkinson's disease detection from voice recordings. Computers in biology and medicine, 141, 105021.

[17] Alissa, M., Lones, M. A., Cosgrove, J., Alty, J. E., Jamieson, S., Smith, S. L., & Vallejo, M. (2022). Parkinson’s disease diagnosis using convolutional neural networks and figure-copying tasks. Neural Computing and Applications, 34(2), 1433-1453.

[18] Moradi, S., Tapak, L., & Afshar, S. (2022). Identification of Novel Noninvasive Diagnostics Biomarkers in the Parkinson’s Diseases and Improving the Disease Classification Using Support Vector Machine. BioMed Research International, 2022.

[19] Mansour, A., & Lachiri, Z. (2017). SVM based emotional speaker recognition using MFCC-SDC features. International Journal of Advanced Computer Science and Applications, 8(4), 538-544.

[20] Khorasani, A., & Daliri, M. R. (2014). HMM for classification of Parkinson’s disease based on the raw gait data. Journal of medical systems, 38(12), 1-6.

[21] Roth, N., Küderle, A., Ullrich, M., Gladow, T., Marxreiter, F., Klucken, J., ... & Kluge, F. (2021). Hidden Markov Model based stride segmentation on unsupervised free-living gait data in Parkinson’s disease patients. Journal of neuroengineering and rehabilitation, 18(1), 1-15.

[22] Wyrobnik, M., van der Meer, E., & Klostermann, F. (2022). Altered event processing in persons with Parkinson’s disease. Psychophysiology, e14021.

[23] Orozco-Arroyave, J. R., Arias-Londoño, J. D., Vargas-Bonilla, J. F., & Nöth, E. (2013, June). Perceptual analysis of speech signals from people with Parkinson’s disease. In International Work-Conference on the Interplay Between Natural and Artificial Computation (pp. 201-211). Springer, Berlin, Heidelberg.

[24] Hlavnička, J., Čmejla, R., Tykalová, T., Šonka, K., Růžička, E., & Rusz, J. (2017). Automated analysis of connected speech reveals early biomarkers of Parkinson’s disease in patients with rapid eye movement sleep behaviour disorder. Scientific reports, 7(1), 1-13.

[25] Soumaya, Z., Taoufiq, B. D., Nsiri, B., & Abdelkrim, A. (2019, April). Diagnosis of Parkinson disease using the wavelet transform and MFCC and SVM classifier. In 2019 4th World Conference on Complex Systems (WCCS) (pp. 1-6). IEEE.

[26] Benba, A., Jilbab, A., Hammouch, A., & Sandabad, S. (2015, March). Voiceprints analysis using MFCC and SVM for detecting patients with Parkinson's disease. In 2015 International conference on electrical and information technologies (ICEIT) (pp. 300-304). IEEE.

[27] Benba, A., Jilbab, A., & Hammouch, A. (2016). Discriminating between patients with Parkinson’s and neurological diseases using cepstral analysis. IEEE transactions on neural systems and rehabilitation engineering, 24(10), 1100-1108.

[28] Oung, Q. W., Basah, S. N., Muthusamy, H., Vijean, V., & Lee, H. (2018, March). Evaluation of short-term cepstral based features for detection of Parkinson’s Disease severity levels through speech signals. In IOP Conference Series: Materials Science and Engineering (Vol. 318, No. 1, p. 012039). IOP Publishing.

[29] Soumaya, Z., Taoufiq, B. D., Benayad, N., Yunus, K., & Abdelkrim, A. (2021). The detection of Parkinson disease using the genetic algorithm and SVM classifier. Applied Acoustics, 171, 107528.

[30] Drissi, T. B., Zayrit, S., Nsiri, B., & Ammoummou, A. (2019). Diagnosis of Parkinson’s disease based on wavelet transform and mel frequency cepstral coefficients. Int. J. Adv. Comput. Sci. Appl, 10, 125-132.

[31] Soumaya, Z., Taoufiq, B. D., Benayad, N., Achraf, B., & Ammoumou, A. (2020). A hybrid method for the diagnosis and classifying parkinson's

patients based on time–frequency domain properties and K-nearest neighbor. Journal of Medical Signals and Sensors, 10(1), 60.

[32] Boualoulou, N., Nsiri, B., Drissi, T. B., & Zayrit, S. (2022). Speech analysis for the detection of Parkinson's disease by combined use of empirical mode decomposition, Mel frequency cepstral coefficients, and the K-nearest neighbor classifier. In ITM Web of Conferences (Vol. 43, p. 01019). EDP Sciences.

[33] Furui, S. (1986). Speaker-independent isolated word recognition using dynamic features of speech spectrum. IEEE Transactions on Acoustics, Speech, and Signal Processing, 34(1), 52-59.

[34] Quinlan, J. R. (1986). Induction of decision trees. Machine learning, 1(1), 81-106.

[35] Dimauro, G., Di Nicola, V., Bevilacqua, V., Caivano, D., & Girardi, F. (2017). Assessment of speech intelligibility in Parkinson's disease using a speech-to-text system. IEEE Access, 5, 22199-22208.

[36] Sakar, B. E., Isenkul, M. E., Sakar, C. O., Sertbas, A., Gurgen, F., Delil, S., ... & Kursun, O. (2013). Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. IEEE Journal of Biomedical and Health Informatics, 17(4), 828-834.

[37] Eskıdere, Ö., Karatutlu, A., & Ünal, C. (2015, September). Detection of Parkinson's disease from vocal features using random subspace classifier ensemble. In 2015 Twelve International Conference on Electronics Computer and Computation (ICECCO) (pp. 1-4). IEEE.

[38] Li, Y., Zhang, C., Jia, Y., Wang, P., Zhang, X., & Xie, T. (2017, October). Simultaneous learning of speech feature and segment for classification of Parkinson disease. In 2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom) (pp. 1-6). IEEE.

[39] Ali, L., Khan, S. U., Arshad, M., Ali, S., & Anwar, M. (2019, July). A multi-model framework for evaluating type of speech samples having complementary information about Parkinson's disease. In 2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE) (pp. 1-5). IEEE.

[40] Ali, L., Zhu, C., Zhang, Z., & Liu, Y. (2019). Automated detection of Parkinson's disease based on multiple types of sustained phonations using linear discriminant analysis and genetically optimized neural network. IEEE Journal of Translational Engineering in Health and Medicine, 7, 1-10.

Efficient Intrusion Detection System for IoT Environment

Rehab Hosny Mohamed[1], Faried Ali Mosa[2], Rowayda A. Sadek[3]

Information Technology Department, Faculty of Computers & Artificial Intelligence, Beni-Suef University, Cairo, Egypt[1, 2]
Information Technology Department, Faculty of Computers & Artificial Intelligence, Helwan University, Cairo, Egypt[3]

***Abstract*—These days, the Internet is subjected to a variety of attacks that can harm network devices or allow attackers to steal the most sensitive data from these devices. IoT environment provides new perspective and requirements for Intrusion detection due to its heterogeneity. This paper proposes a newly developed Intrusion Detection System (IDS) that relies on machine learning and deep learning techniques to identify new attacks that existed systems fail to detect in such an IoT environment. The paper experiments consider the benchmark dataset ToN_IoT that includes IoT services telemetry, Windows, Linux operating system, and network traffic. Feature selection is an important process that plays a key role in building an efficient IDS. A new feature selection module has been introduced to the IDS; it is based on the ReliefF algorithm which outputs the most essential features. These extracted features are fed into some selected machine learning and deep learning models. The proposed ReliefF-based IDSs are compared to the existed IDSs based correlation function. The proposed ReliefF-based IDSs model outperforms the previous IDSs based correlation function models. The Medium Neural Network model, Weighted KNN model, and Fine Gaussian SVM model have an accuracy of 98.39 %, 98.22 %, and 97.97 %, respectively.**

Keywords*—*Intrusion detection systems (IDSs); TON IoT dataset; machine learning; deep learning; ReliefF

I. Introduction

The Internet of Things (IoT) environment describes a network of physical objects that includes software, sensors, and other technologies to connect and transfer data with systems and devices on the Internet. IoT represents the interconnection of heterogeneous entities, where the term “entity” refers to a human being, a sensor, or possibly anything that can request or provide a service [1]. The Internet of Things (IoT) has become one of the most important technologies of the 21st century, so we need to protect data traveling between devices. IoT security is an ongoing research topic that's gaining increasing attention in governmental, industrial, and academic research. IoT security requirements are Confidentiality, Integrity, Availability, Authentication, and Authorization. There are multiple IDS systems that are used to protect networks from attacks, but some systems can’t identify the newest attacks. We provide in this paper an intrusion detection system to detect attacks, the system is developed using deep learning and machine learning techniques with an efficient feature selection using the ReliefF algorithm. The proposed model is based on the Windows 10 subset of the ToN-IoT dataset. A comparison is carried out with the existed work which selects features from Windows 10 dataset using the correlation function.

In this paper, we focus on Windows 10 dataset as the windows 10 operating system is widely used in personal computers. This dataset contains new attacks for IIoT networks which can’t be detected by traditional IDSs so this paper presents a new IDS system for IoT networks by using machine learning and deep learning algorithms. Many datasets have redundant or noisy data which may cause poor learning performance and consume time for training so the feature selection techniques are used to remove irrelevant and noisy attributes from the dataset. In this paper, we apply the ReliefF feature selection method on the windows 10 dataset.

The paper structure is as follows: Section II reviews the IDS backgrounds, Datasets, feature selection techniques, and Machine learning and deep learning algorithms. Section III discusses Related Works on IDS. The proposed model is explained in Section IV. Following that the Evaluation Metrics and Experimental Results are described in Section V. Finally, Section VI describes the conclusions of the future work.

II. Background

Intrusion detection systems (IDSs) are tools or software that can secure or defend your networks from intrusions. The purpose of IDS is to recognize various types of attacks and usage of the computer that cannot be known via firewalls. This is often very important to achieve a high level of protection versus action which compromises the availability, integrity, or confidentiality of a computer system [2]. IDS monitors network traffic for suspicious activities and issues alerts when such activities are detected. Some intrusion detection systems can do actions if anomalous traffic is detected such as blocking traffic sent from a suspicious IP address. IDS sometimes is classified based on its location; Host Intrusion Detection Systems (HIDS) and Network Intrusion Detection Systems (NIDS) [3]. HIDS monitors and analyzes host activities, application logs, system calls, and modifications that occur on system files to identify intrusions such as login to unprivileged data [4]. Network-based IDS monitors the network traffic by using techniques such as packet sniffing to collect network traffic and detect attacks such as DOS attacks or port scans [4]. There are two types of Network-based IDS Statistical anomaly IDS and Pattern matching IDS [3]. On the other hand, IDS is practically classified as signature-based IDS and anomaly-based IDS. Anomaly-based IDS is significantly used for detecting zero-day attacks. Efficient IDS development depends on selecting suitable machine learning or deep learning models as well as the feature selection technique in order to enhance the performance and reduce the computations.

Feature selection is an important step that plays a significant role in the development of an effective IDS. Feature selection is categorized into three methods like filter methods, wrapper methods, and embedded methods. The filter methods select the most discriminating features by data character. Features are ranked according to specific criteria and then the highest-ranked features are selected. There are many types of filter methods that have been used, such as ReliefF, F-statistic, mRMR, correlation, and information gain. Wrapper methods use the intended learning algorithm to evaluate the features. Examples of wrapper methods are Sequential forward selection and Sequential backward selection. Embedded methods perform feature selection in the model construction process examples of embedded methods are Elastic Net and Ridge Regression [5]. The most widely used feature selection methods are The ReliefF (Relief-F) [6], Correlation-Based Feature Selection [7, 8], and Information Gain Ratio-based feature selection [9, 10].

Various network datasets were prepared for the evaluation of IDS, such as KDDCUP99, NSLKDD [11], UNSW-NB15 [12], Ton-IoT, and ISCX [13].

There are several techniques of classification like decision trees, naïve Bayes (NB), neural networks (NN), support vector machines (SVM), Random Forest (RF), rule-based system, and nearest neighbor (KNN) [14,15]. Every algorithm utilizes learning methods to create a classification model. However, proper classification techniques should handle the training data and should be able to accurately identify a class of records that have not ever been seen before [4].

Deep Learning a branch of Machine Learning, has also begun to be widely used to implement IDS. There are many types of deep learning algorithms that can be used for intrusion detection systems such as recurrent neural network, deep belief network, Long Short Term Memory, restricted Boltzmann machine, deep auto-encoder, self-taught learning, convolutional neural network, deep migration learning, and replicator neural network [16, 17].

III. Related Work

There are lots of researches that use some methodologies and algorithms to overcome the most advanced attacks in the network. These researches trend to intrusion detection systems (IDSs) by using machine learning and deep learning algorithms with different feature selection methods to select the most important features from the datasets. The related works are explained in this section as follows:

Senthilnayaki Balakrishnan et al [18], the authors developed a new feature selection method based on Information Gain Ratio. They had named it an Optimal Feature Selection algorithm that could select an optimal number of features from the dataset. They used the KDD Cup dataset. They used Support Vector Machine and Rule-Based Classification algorithms to classify data as normal or as attacks data. The proposed feature selection and classification algorithms had achieved the best results.

S. Ramakrishnan et al. [19], the authors developed an IDS to detect attacks or normal data from the KDD Cup 99 dataset. They first selected the most important attributes by using the entropy-based feature selection method. The Fuzzy Control Language was used to classify data as normal or attack data. The results showed that the proposed system reduced the computational time and achieved high accuracy.

Peilun Wu et al. [20], in this paper, the authors developed an IDS system called Densely-ResNet. They selected the important attributes using the correlation function. The UNSW-NB15 was used to evaluate the performance of Densely-ResNet. The results showed that the Densely-ResNet had achieved high accuracy and reduced false alarm rate.

Shahid Latif et al. [21], the authors proposed an IDS that utilized a deep random neural network to protect IIoT systems. However, the system had been tested with the UNSW NB15 dataset. This dataset was used to observe its applicability or feasibility to IIoT. The results showed a better detection accuracy with a low false alarm rate.

Merna Gamal et al [14], the authors proposed a hybrid intrusion detection system using machine learning and deep learning techniques to benefit from the advantage of deep learning and machine learning. They used 10% of the KDDcup1999 dataset, and applied CNN to extract features from the KDDcup1999 dataset, then used machine learning techniques (SVM, KNN) to classify the data. The experiment results showed that the proposed model had achieved the best accuracy.

Prabhat Kumar et al [22], the authors developed an ensemble learning and fog-cloud architecture-driven cyberattack detection framework for IoMT networks. They used a ToN-IoT dataset that was collected from a large-scale and heterogeneous IoT network. Results showed that the proposed system could achieve high accuracy of 96.35%, a high detection rate of 99.98%, and a reduction in false alarm rate by 5.59%.

IV. Proposed Model

This section provides our proposed model. We select features from windows 10 datasets using the ReliefF algorithm. We apply the ReliefF algorithm to remove irrelevant variables from Windows 10 dataset. Windows 10 dataset contains 125 attributes and two attributes one called “Label” whose values are (0 or 1), 0 for normal data and 1 for attack data, and “type” attribute contains normal data and seven attack categories (DDOS, DOS, Injection, MITM, Password, Scanning, and XSS). Before selecting features from the windows 10 dataset, we do preprocess on a dataset. The “type” feature refers to attack categories, these categories are converted from text data to numeric data from 2 to 8 respectively, and normal data are converted to 1. Fig. 1 shows the number of rows for each type of data in the dataset. Then we divide the dataset into training and testing data using holdout validation, 80 % of the dataset is training data and 20 % of the dataset is used as testing data.

No of rows	Type
4608	ddos
525	dos
612	injection
24871	normal
1269	xss
3628	password
447	scanning
15	mitm

Fig. 1. Number of Rows for each Type of Data.

The ReliefF algorithm is used to select features from the windows 10 dataset. After selecting features, we apply deep learning and machine learning techniques to classify the data as normal or attacks data. The Matlab tool and Machine Learning Toolbox (Classification Learner) are used to simulate algorithms such as KNN, SVM, and NN. We apply different types of KNN algorithm (such as Weighted KNN and Medium KNN), SVM (like Linear SVM and Fine Gaussian SVM), and Neural Network (like Medium NN and Bilayered NN). Also, we apply the LSTM algorithm to the dataset. Fig. 2 shows the flowchart of our proposed model.

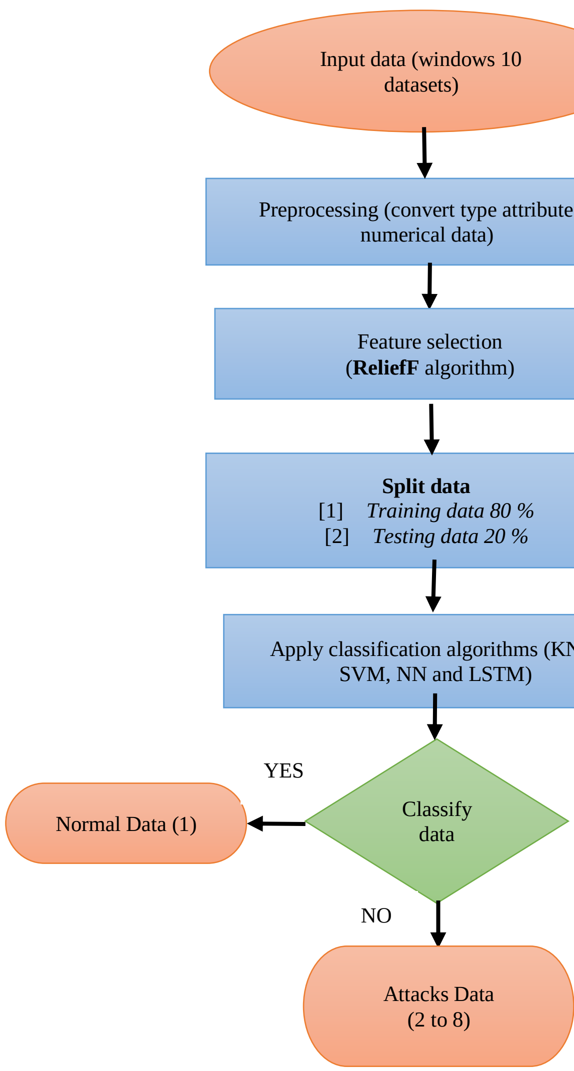


Fig. 2. Flowchart of Proposed Model.

V. Simulation and Experimental Results

Experiments are carried out to examine the performance of the proposed model.

A. Ton-IoT Datasets Description

Ton-IoT datasets were collected from Telemetry datasets of IoT services, Operating systems datasets of Windows and Linux, as well as datasets of Network traffic. The Windows datasets were generated using the virtual machines running Windows 7 and Windows 10 and incorporated the collections of data from multiple sources, including memory, process, processor, and hard drive of the systems. Windows 10 dataset contained 125 attributes and two attributes one called "Label" which values are (0 or 1), 0 for normal data and 1 for attack data, and the "type" attribute contained normal data and seven attack categories.

We apply our proposed model to Windows 10 dataset. We select features from it using the ReleifF algorithm then we apply deep learning and machine learning techniques to classify the data as normal or attack data. We divide the dataset using holdout validation as the following; 80 % of the dataset is used for training (28780 data sample records). 20 % of the dataset is used for testing (7195 data sample records).

B. Simulation Environment Properties

The proposed system is implemented using Matlab 2021a, Classification learner App is set to apply imported and preprocessed instances of the dataset the processor unit (CPU) is Intel(R) Core(TM) i5-6200U CPU @ 2.30GHz 2.40 GHz, main memory(RAM) is 16.0 GB and the operating system is Windows 10 Pro.

C. Evaluation Metrics

There are performance metrics are calculated to examine the model performance. These metrics are accuracy, recall, precision rate and, F1-Measure ought to be calculated for evaluation. These all performance metrics can be calculated by using a confusion matrix.

- Confusion matrix is a table or array that represents the performance of a classification model on testing data for which the true values are identified. Fig. 3 shows the confusion matrix.

Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

Fig. 3. Confusion Matrix.

- Accuracy is the ratio of true detection over the whole instances.

$$\boldsymbol{accuracy} = \frac{\text{TP+TN}}{\text{totalSample}} \tag{1}$$

- Recall is how often does it predict correctly. Also known as Sensitivity or True Positive Rate (TPR).

$$\boldsymbol{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{2}$$

- Precision indicates how often it is accurate when it is predicted to be accurate.

$$\boldsymbol{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

- F1-measure is the average of recall and precision weight. The mathematical representation of all measures can be deduced from the confusion matrix.

$$\boldsymbol{F - Mesure} = 2 * \frac{precision * Recall}{precision + Recall} \tag{4}$$

Area Under ROC Curve (AUC) is the performance metric for binary classification issues. AUC represents the work of the model for distinguishing between the negative classes and positive classes. An area of 1.0 acts as a model in which all predictions are made perfectly. An area of 0.5 acts as a model that is as good as random. The ROC can be classified into specificity and sensitivity.

D. Feature Selection

The paper in [23] selects features from Windows 7 and Windows 10 datasets using a correlation function that selects the most correlated features in datasets, and the authors suggest applying machine learning and deep learning algorithms to selected features. In our proposed model we suggest modifying the feature selection step, we apply the ReliefF algorithm to remove irrelevant variables from Windows 10 dataset. We compare our proposed model with the work implemented in the paper [23].

We use the ReliefF algorithm because it is a widely used filter-based feature selection method that finds the best feature subset by calculating the features' weights. The ReliefF algorithm's advantages are that more robust and can deal with incomplete and noisy data. It consumes less computational resources and is not limited to two-class problems, but it can work with multiple classes [6, 24]. The ReliefF algorithm first selects R samples randomly from the training set, and then it finds Near Hits in the same class, finds Near Miss in other classes, and updates each feature weight by rules. After repeating the above action m times, all feature weights will be obtained. The higher the feature weight is, the more useful the feature is for classification. The reverse is also true. The following steps are the pseudocode of the ReliefF algorithm:

Pseudocode *for* ReliefF *Algorithm:*

Input: a vector of attribute values and the class value for each training instance
Output: the vector W of estimations of the attributes qualifies.

Step 1: put up all of the weights *W[A]* := 0.0;
Step 2: **for** i:= 1 **to** *m* **do begin**
Step 3: randomly select an instance R_i ;
Step 4: find *k* nearest hits H_j ;
Step 5: **for** each class $C \neq \text{class}(R_i)$ **do**
Step 6: from class *C* find *k* nearest misses $M_j(C)$;
Step 7: **for** *A*:= 1 **to** a **do**

Step 8: $\text{W[A]} := \text{W[A]} - \sum_{j=1}^{K} \frac{\text{diff(A ,Ri ,Hi)}}{\text{m .K}} +$

Step 9: $\sum_{C \neq \text{class}(Ri)} [\frac{\text{P}(C)}{1-P(\text{class}(Ri))} \sum_{j=1}^{K} \frac{\text{diff(A ,Ri ,}Mj(C))}{\text{m .K}}$
Step 10: **end**;

E. Result and Discussion

ReliefF algorithm estimates the weight of each feature and sorts features according to their weights. Fig. 4 shows the most important weights of 20 selected features, the highest feature weight is equal to 0.257.

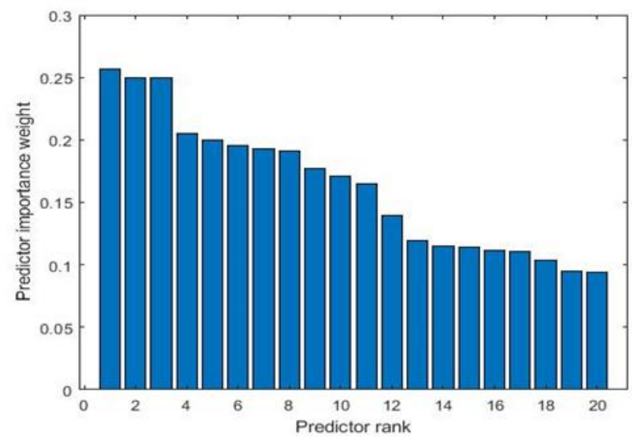


Fig. 4. Weights of Selected Features.

Tables I and II show features selected in the windows 10 dataset using the correlation function in [23] and features selected using the ReliefF algorithm in our proposed model, respectively. The selected features will be then used for training and testing machine learning and deep learning algorithms to evaluate their efficiency in classifying the dataset as normal or attacks data.

After selecting features, we apply deep learning and machine learning techniques to classify the data as normal or as attacks data. The algorithms applied are KNN, SVM, Neural Network, and LSTM. We apply different types of KNN algorithms (such as Weighted KNN and Medium KNN), SVM (like Linear SVM and Fine Gaussian SVM), and Neural Network (like Narrow NN and Bilayered NN). Table III shows the performance metrics of classification algorithms applied to the features selected in [23] in the windows 10 dataset. Table IV shows performance metrics of classification algorithms applied to the features selected using the ReliefF algorithm in windows 10 datasets.

TABLE I. FEATURES SELECTED FROM WINDOWS 10 DATASETS USING CORRELATION FUNCTION

Feature selection method	Selected Features
correlation function [23]	1. Network I.Intel.R 82574L GNC.Current Bandwidth 2. Network I.Intel.R 82574L GNC.Packets Sent Unicast.sec 3. Network I.Intel.R8 2574L GNC. Packets Sent.sec 4. LogicalDisk. Total Disk Read Bytes.sec 5. LogicalDisk Total. Avg.Disk.Bytes Transfer 6. Memory.Modified. Page List Bytes 7. Memory.Pool.Paged Bytes 8. Memory.Page Reads.sec 9. Process IO Data Operations sec 10. Processor pct Processor Time

TABLE II. FEATURES SELECTED FROM WINDOWS 10 DATASETS USING THE RELIEFF ALGORITHM

Feature selection method	Selected Features
ReliefF algorithm in proposed model	1. MemorySystemDriverResidentBytes 2. Process_Virtual_Bytes 3. Process_Virtual_BytesPeak 4. MemoryLong_TermAverageStandbyCacheLifetime_s 5. Process_Pool_PagedBytes 6. Process_HandleCount 7. Network_I_IntelR_82574L_GNC_TCP_APS 8. Process_WorkingSet 9. MemorySystemCacheResidentBytes 10. Process_ThreadCount 11. MemoryPct_CommittedBytesInUse 12. Process_Working_Set_Peak 13. MemoryPoolNonpagedBytes 14. MemoryStandbyCacheReserveBytes 15. MemorySystemDriverTotalBytes 16. Process_Working_Set_Private 17. MemoryCacheBytesPeak 18. MemoryCommitLimit 19. MemoryStandbyCacheCoreBytes 20. MemoryPoolPagedResidentBytes

TABLE III. PERFORMANCE METRICS OF CLASSIFICATION ALGORITHMS IN FEATURES SELECTED USING CORRELATION FUNCTION IN [23]

Classification algorithm	Accuracy (%)	Precision	Recall	F_measure	AUC
Medium KNN Model	93.47	0.664	0.957	0.785	0.99
Weighted KNN Model	94.02	0.686	0.956	0.799	0.99
Linear SVM Model	75.48	0.006	0.006	0.006	0.97
Fine Gaussian SVM Model	91.22	0.594	0.926	0.724	0.99
Bilayered NN Model	92.23	0.622	0. 956	0.753	0.99
Medium NN Model	94.12	0.689	0.962	0.803	0.99

TABLE IV. PERFORMANCE METRICS OF CLASSIFICATION ALGORITHMS IN FEATURES SELECTED USING THE RELIEF ALGORITHM IN THE PROPOSED MODEL

Classification algorithm	Accuracy (%)	Precision	Recall	F_measure	AUC
Medium KNN Model	98.15	0.994	0.979	0.987	0.99
Weighted KNN Model	98.22	0.991	0.983	0.987	0.99
Linear SVM Model	96.54	0.985	0.964	0.975	0.99
Fine Gaussian SVM Model	97.97	0.992	0.978	0.985	0.99
Bilayered NN Model	98.17	0.996	0.977	0.987	1.00
Medium NN Model	**98.39**	0.996	0.980	0.988	1.00

TABLE V. PERFORMANCE METRICS OF LSTM IN FEATURES SELECTED USING THE RELIEFF ALGORITHM AND CORRELATION FUNCTION

LSTM model	Accuracy	Precision	Recall	F_measure
ReliefF algorithm	70 %	0.72	0. 94	0.81
Correlation function	68.9 %	0.71	0.95	0.81

Long Short Term Memory algorithm (LSTM) is also applied to features of windows 10 which are selected from using the correlation function in the paper [23] and the ReliefF algorithm in our proposed model. Table V shows the comparison of performance metrics of the LSTM model applied to the features selected using the correlation function in the paper [23] and the ReliefF algorithm in our proposed model in windows 10 datasets.

The results showed that the accuracy of the LSTM model is 68.9 % of features selected from using the correlation function, and the accuracy of the LSTM model is 70 % of features selected from using the ReliefF algorithm and the Precision is 0.72 in our proposed model unlike in the paper [23] is 0.71. This indicates that our proposed model has achieved the best results.

The results of Table III and Table IV showed that our proposed model has achieved the best results, as in our proposed model the Medium NN Model has best results than in results in the paper [23]. The accuracy of the Medium NN Model in our proposed model is 98.39 and the Precision is 0.996 unlike in the paper [23] the Medium NN Model has achieved 94.12% and Precision is 0.689. Also Weighted KNN Model and Bilayered NN Model accuracy is 98.22%, and 98.17%, respectively. Unlike in [23], the Weighted KNN Model and Bilayered NN Model accuracy is 94.02%, and 92.23%, respectively.

The accuracy of the Linear SVM algorithm in our proposed model is 96.54 %, but in paper [23] the accuracy is 75.48%. The reason this algorithm has achieved lower results than other machine learning algorithms is that this algorithm works more efficiently with two classes, but if it works with a dataset that has multiple classes may achieve low accuracy, and it isn't suitable for large datasets.

LSTM also has archived low accuracy as it doesn't work efficiently with a large number of samples. As the number of samples increases, the accuracy of LSTM tends to decrease. Figure 5 shows a graphical representation of the accuracy of algorithms applied to features selected from the windows 10 dataset. This comparison shows that our proposed model has achieved the best results than the results of [23].

Fig. 6 and 7 show the graphical representation of performance metrics of our proposed model and of [23], respectively.

Fig. 6 and 7 show that our proposed model has achieved high values of performance metrics against another work [23].

All results of our proposed model are better than the results of [23]. It indicates that the ReliefF algorithm selects the most important features and it can work with multiple classes.

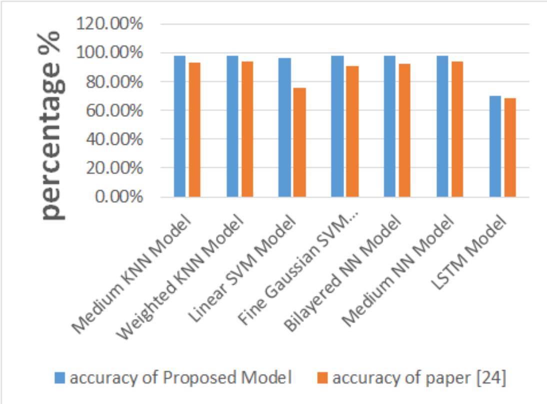


Fig. 5. Accuracy Comparison between Paper [23] and our Proposed Model.

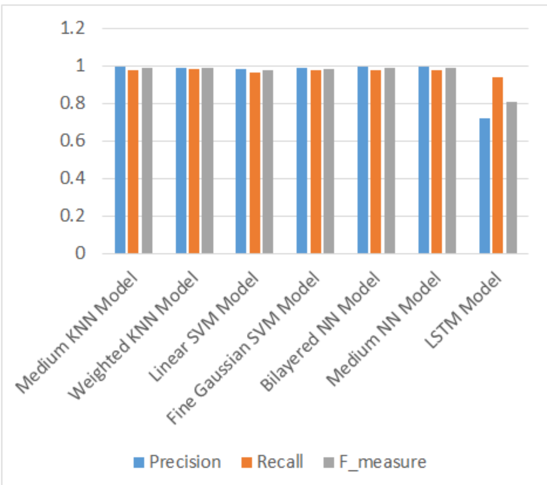


Fig. 6. Performance Metrics of our Proposed Model.

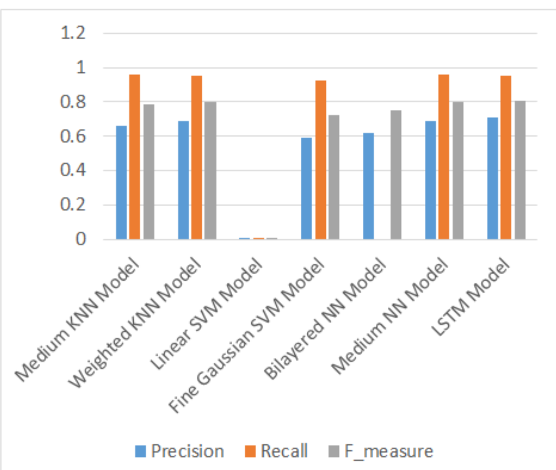


Fig. 7. Performance Metrics of Paper [23].

VI. Conclusion

This paper has introduced an efficient IDS for the IoT environment. The paper uses the benchmark dataset ToN_IoT which includes IoT telemetry data. ReliefF algorithm is proposed to efficiently select the most vital features to get more enhanced performance and computations reduction. A comparison is carried out with the correlation function by applying machine learning and deep learning algorithms to the selected features from the windows 10 dataset. The proposed model has been applied to the windows 10 subset. The experimental results reveal that our proposed model has achieved better results than theirs. Our proposed model has high accuracy such as using Medium Neural Network, Weighted KNN, and Fine Gaussian SVM has 98.39 %, 98.22 %, and 97.97 % respectively. In the future, we will use Transfer Learning to enhance the accuracy of the Long Short Term Memory (LSTM) model.

References

[1] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," Computer Networks 54, no. 15 (2010): 2787-2805.

[2] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, "Survey of intrusion detection systems: techniques, datasets and challenges," Cybersecurity (2019).

[3] J. Surana, J. Sharma, I. Saraf, N. Puri, and B. Navin “A Survey On Intrusion Detection System,” International Journal of Engineering Development and Research (2017).

[4] A.J. Deepa and V. Kavitha, “A Comprehensive Survey on Approaches to Intrusion Detection System,” Procedia Engineering (2012).

[5] J. Miao and L. Niu, “A Survey on Feature Selection,” Information Technology and Quantitative Management (ITQM 2016).

[6] M. ROBNIK and I. KONONENKO, “Theoretical and Empirical. Analysis of ReliefF and RReliefF,” Machine Learning (2003).

[7] Y. Chen, Y. Li, X. Cheng, and L. Guo,” Survey and Taxonomy of Feature Selection Algorithms in Intrusion Detection System,” Information Security and Cryptology, Lecture Notes in Computer science (2006).

[8] M. A. Hall, “Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning,” In Proc. of the 17th Int. Conf. on Machine Learning. Morgan Kaufmann Publishers Inc (2000).

[9] M. Trabelsia, N. Meddouria, and M. Maddouri, “A New Feature Selection Method for Nominal Classifier based on Formal Concept Analysis,” International Conference on Knowledge Based and Intelligent Information and Engineering Systems (September 2017).

[10] M.sujatha, and Dr. G. Lavanya Devi, “Feature Selection Techniques using for High Dimensional Data in Machine Learning,” International Journal of Engineering Research & Technology (IJERT) (2013).

[11] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," IEEE Symposium on Computational Intelligence for Security and Defense Applications (2009).

[12] N. Moustafa, and J. Slay, "UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems (UNSW-NB15 Network Dataset," Military Communications and Information Systems Conference (MilCIS). IEEE (2015).

[13] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization,” ICISSP (2018).

[14] M. Gamal, H. Abbas, and R. A. Sadek, “Hybrid Approach for Improving Intrusion Detection Based on Deep Learning and Machine Learning Techniques,” Springer (2020).

[15] O. Muhammad Altoumi Alsyaibani, E. Utami, and A. Dwi Hartanto, “Survey on Deep Learning Based Intrusion Detection System,” Telematika (2021).

[16] A. Mehnas Muthalib, S. Theres Mathew, S. Tom Mathew, R. Issac, and D. Sunny, “A SURVEY ON DEEP LEARNING ALGORITHMS,” International Research Journal of Engineering and Technology (IRJET, 2020).

[17] A. Aldweesh, A. Derhab, and A. Z. Emam, “Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues,” Journal Pre-proof (2019).

[18] S. Balakrishnan, V. K, and K. A, "Intrusion Detection System Using Feature Selection and Classification Technique," International Journal of Computer Science and Application (IJCSA) Volume 3 Issue 4, November 2014.

[19] S. Ramakrishnan1 and S. Devaraju, "Attack's Feature Selection-Based Network Intrusion Detection System Using Fuzzy Control Language," Taiwan Fuzzy Systems Association and Springer-Verlag Berlin Heidelberg (2016).

[20] P. Wu, N. Moustafa, S. Yang, and H. Guo, "Densely Connected Residual Network for Attack Recognition," IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom) (2020).

[21] S. Latif, Z. Idrees, Z. Zou, and J. Ahmad. Drann, "A deep random neural network model for intrusion detection in industrial iot," International Conference on UK-China Emerging Technologies (UCET) IEEE (2020).

[22] P. Kumar, G. P. Gupta, and R. Tripathi, "An ensemble learning and fog-cloud architecture-driven cyber-attack detection framework for IoMT networks," Computer Communications (2021).

[23] N. Moustafa, M. Keshky, E. Debiez and H. Janicke, "Federated TON IoT Windows Datasets for Evaluating AI-based Security Applications," IEEE 19th International Conference on Trust, Security, and Privacy in Computing and Communications (2020).

[24] Z. Huang, C. Yang, X. Zhou, and T. Huang, "A hybrid feature selection method based on binary state transition algorithm and ReliefF," IEEE Journal of Biomedical and Health Informatics (2018).

Augmented System for Food Crops Production in Agricultural Supply Chain using Blockchain Technology

Dayana D. S[1]
Research Scholar, Department of Computer Science
SRM Institute of Science and Technology, Chennai, India

Kalpana G[2]
Associate Professor, Department of Computer Science
SRM Institute of Science and Technology, Chennai, India

***Abstract*—The elevated version of the Agricultural Traceability System dealing with food production holds utmost significance in not only assuring food insecure, smart contracts and agri-insurance to farmers but also guaranteeing insurance to them during natural disaster. The proposed and improved system for food cultivation traceability deals with agri-crops and farmers that hires the Blockchain technology guarantee at par safety, agreement, distributed ledger, immediate payment, de-centralization and thereby achieving the goal of minimizing the cost incurred in the food processing system and building trust. Smart contracts play a pivot role in the field of agricultural insurance. Agricultural insurance based upon Blockchain comprises of major weather incidents and associated payouts enlisted on a smart contract, connected to the mobile wallets with timely weather updates notified by the field sensors and interrelated with data from proximity weather stations would enable prompt payout during any natural calamity such as flood or drought. A panel of advisers in the decentralized system which is professionally governed and managed by certain retired officers makes the traceability system more trustworthy. These professionals can offer wise suggestions to the planters aiding them to acquire productive outcome.**

Keywords*—Blockchain; distributed ledger; consensus; decentralized; stakeholders; agricultural insurance; payouts; trust-based farming system; food safety; panel of advisers; agri-supply chain

I. Introduction

Food safety tends to be an over concerning matter which is seeking attention worldwide. Both the agricultural sector and food quality are of utmost necessity without doubt and it is crucial that the end-user receives a quality product. But without any centralized framework, this is infeasible to achieve. Usually, the food supply chain is quiet long as it involves various sub departments such as farming, producing, transporting and much more. There are many restrictions and formalities which becomes hindrances in addressing the challenges faced and fulfilling the needs of the end customer. Not only the end-users, farmers and transporters are of major concern but there exist other challenges too that can’t be neglected. Factors or entities such as the stake holders, agricultural insurance, factory staff, wholesales shops as well as the retail shopkeepers are of significant importance in the process of food supply chain and involves many complications which must be closely observed as pointed out by [1].

The initial stage of food supply chain is Farming. If Unscientific farming methods and inefficient agricultural process are being practiced, it will hamper both the production and the quality. Also, the fertilized agricultural area will be affected and neglected by the farmers for carrying out further farming. Another underlying factor is the seed’s quality that will be sowed in the field. In case of small-scale fields, the loss incurred due to poor quality of seed can still be handled. But fields that are in huge acres and caters to voluminous production cannot take any chance with the quality of the seed. Adopting unscientific means in treating and handling animal-farms can lead to severe diseases in animals and can transmit to humans like the avian flu. Also, if the feed and food are contaminated with chemicals, it can be risky and unsafe for the animals and any other products according to [2].

According to experts, the cost of production and transportation has an impact on the market value of the finished product [3]. Once the products are collected from the producers, the overall expenses and transportation overhead is computed and evaluated by the stakeholders which may not be so trustworthy. Such computations are bound to be altered and misleading because of varying fuel cost, bribe and corruption involved thus causing inflation. Due to such deeds, economy of the entire nation is at stake. After the stakeholders, the wholesale and retail keepers come into the picture where it is very likely that the market value of the product can increase further. Inflation affects the trading of food retailers leading to shutdown of stores, rise of independent retails power, polarization of store sizes and the targeting of market struggles as put forth by [4]. Apart from the cost, the food safety too gets hampered with such inflation. In other words, inflation is directly proportional to the loss incurred, higher the loss, higher will be the corruption and higher will be the inflation. Both the quality and quantity get affected right from the stockholders to retail shopkeepers as emphasized by [5].

The reason behind all the above prominent issues is the lack of a well-designed traceability framework comprising of all the departments. According to the concept, all individuals who are a part of the food supply chain, from the seed seller to the end user, must be catered to and accommodated. [6]. Missing upon any one entity or not abiding by the said criteria, development of any system will contract black holes thus becoming a failure and inefficient structure in the market. Hence, it is essential that every single department is taken into consideration along with effective planning. The proposed system includes all the important departments that are essential in the formation of a complete food supply chain. That is starting from the Seed Seller → Farmer → Insurance Company → Government Panel → Producers -> Distributors -> Retailers and eventually the end-user.

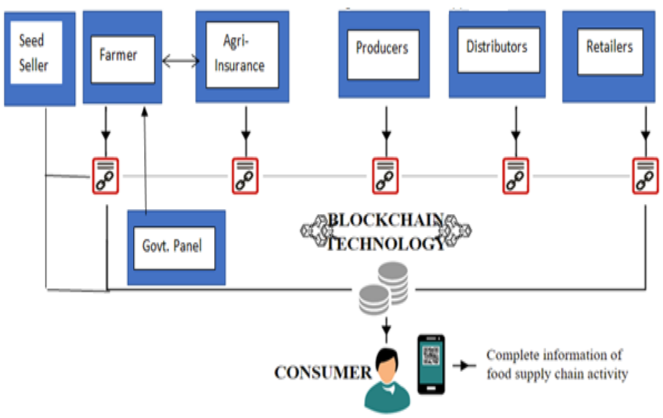


Fig. 1. Production Flow in Food Crops Supply Chain.

Fig. 1 shows the proposed system's production flow in food crops supply chain. The main goal of this article is to improve the agricultural traceability system for food production using Blockchain technology. The Blockchain technique is adopted to frame the above system as per [7] a Blockchain resembles a digital transaction ledger that is controlled using a network of various computing machines without involving the interference of any third party. Management of blocks in a Blockchain which represents individual transaction data files is accomplished using certain software platforms that enable data transmission, processing, storing and its representation in human understandable form. This technique works as per certain standard set of rules according to the consensus mechanisms (run by algorithms). In this form of a distributed consensus, a single authority is not designated as trustworthy but the responsibility is distributed across the entire network. There are various consensus mechanisms prevailing. Hash functions, or hashes, produces a digital fingerprint from the input data or referring an entire document. Similar to the human fingerprint that matches a single person, a digital fingerprint uniquely identifies a single data unit. Any minute alteration in the transaction data leads to totally dissimilar fingerprint. This makes the ledger highly secure against any tampering.

Every process that takes place along the food chain is there and then recorded to the Blockchain. This recorded information is indisputable and is willingly accepted by all the participants. Also, the recorded information is thoroughly verified by the stake holders of the agri-supply chain thus building an agreement amidst among the partakers. Every Block that is validated is then appended to the transaction chain to form a permanent record of the Blockchain. There are certain low-cost agricultural insurance schemes that offer social security to numerous farmers whose livelihood are severely impacted due to the natural calamities. Though such insurance schemes are beneficial, they are rarely availed by the poor farmers or rural people. The potential benefits of the proposed system include profit to all the stakeholders involved in the agri-supply chain. When a transaction is created by a stakeholder, that gets stored in pending transactions and it gets added to the Blockchain once the miner validates the block.

In an agricultural supply chain, traceability of products involves collection, exchange and handling of critical information by determining the source and all the information communication taking place in food supply chain. This information is dynamic in nature as the crops are produced and directed through various middle man; hence following and tracing it becomes infeasible. Spoiled food and its consequences to community health enforce the essential policy tool of traceability for precise monitoring of food standard and protection. The prevailing tool employed in agri-supply chain for traceability confronts data fragmentation and centralized controls that degrade data modification as well as its management. For overcoming the above situation, improvised agricultural traceability system has been recommended that benefits food production and farmers using the effective means of Blockchain technology. Moreover, the control panel's timely and consistent advise pertaining to the crop's status proves to be utmost helpful to the farmers.

The present research aims towards revealing the working of Blockchain in efficiently tracing the agricultural products, make the transparency between all the participants higher and increase food safety in the food processing system for achieving at par efficiency and information sharing. Parameters such as weather, crops, irrigation and fertilizers for agricultural inputs form the dataset. The suggested method creates a Blockchain-based food production system that reduces transmission costs, improves efficiency, and increases security throughout the food supply chain. Security, distributed ledger, consensus, speedy settlement, and decentralization are just a few of the characteristics that have been improved. The SHA-256 (Secure Hash Algorithm) hashing technique is used to determine the hash for agricultural inputs. Agricultural insurance based on Blockchain, which includes major weather incidents and associated payouts enlisted on a smart contract, and which is connected to mobile wallets for timely weather updates notified by field sensors and interrelated with data from nearby weather stations, would be possible. When combined with data from nearby meteorological sensors, it would be possible to pay out quickly in the event of a natural disaster such as a flood or drought.

Compared to the other existing systems, the proposed work has the following advancements:

- The novelty of the proposed system has a module called Agri-Insurance, which is responsible for providing Insurance to farmers for buying seeds and timely payout is given to the farmers during natural calamities.

- The proposed system deals about the agricultural crops and farmers which uses Blockchain technology, that guarantees security, agreement, shared record, fast payment and decentralized network, that achieves the goal of minimizing cost incurred in the food processing system and building trust.

- All transactions are recorded in distributed shared ledger which associates to IPFS that ensures the transparent and traceable in supply chain.
- Blockchain based proof of delivery with automated payments is made to all parties using cryptocurrency by the smart contracts when crops and products are delivered.
- Government Panel gives wise suggestion to Farmers enabling them to increase the crop yield and to take fruitful decisions.

The entire work is classified into the following: Section 2 highlights the literature survey on Blockchain technique implemented in agricultural domain, Section 3 discusses the recommended technique by elaborating the performance of Blockchain in agricultural processing and uplifting the overall trust and security in the agri-food supply chain, Results and discussions are put forth in Section 4 and the conclusion is stated in Section 5 along with the future works.

II. Related Work

It is very well-known that the present agriculture system experiences many setbacks and the technique of Blockchain is the beacon of hope. Though implementing and making best use of this technique still remains ignorant to many. Whereas on the other hand the so called 'Blockchain onlookers' are vigilant against the restrictions and inadequacies of the Blockchain and accordingly apply it in real sense. Hence the need of the hour is to determine the relevant situations in businesses where the implementation of BCT can be revealed in its true sense.

Numerous organizations are seeking towards implying the transparency and fault tolerance of Blockchain technique for handling situations that involves multiple unreliable entities for the distribution of certain resources. In addition, the ceaseless advancement and performance of the Blockchain in various crypto-currencies has won trust of everyone. The agriculture and food supply chain depicts the highly significant domains as indicated by [2].

It has been more than ten years since the whitepaper "Bitcoin was launched. Introduced by a pseudonymous author, this Peer-to-Peer Electronic Cash System" became the very first crypto-currency that enabled trusted financial transactions without the involvement of any banks or financial institutions. The present research lays foundation for the advancement of Bitcoin according to [3]. This technology in collaboration with Blockchain cracked the issue related to double-spending wherein the digital tokens had drawbacks because of duplicate or erroneous computer files.

BCT employs a set of techniques holding a sound year of experience in information technology as well as in commercial domain. Some of them are: private and public key, hash functions, databases particularly the shared databases, agreement algorithms and decentralized systems as put forth by [4]. The prime aim in implementing these techniques is to gain database consistency and integrity with respect to a distributed decentralized database with either controlled or uncontrolled database nodes (such as Bitcoin).

Distributive storage is the best part of this technique. In other words, the information residing in the database resides not just on a single server but rather on innumerable computers spanned across any geographic location. Users that are a part of this network are permitted to access the current registry. Moreover, participants can independently carry out any transactions without the interference of any intermediaries such as the banks. According to the Blockchain system, all digital documents are combined into blocks that are encrypted before computer miners organize them in chronological order by mining - solving relevant mathematical techniques [5].

Blockchain has gained significant popularity in the agricultural system as it immensely benefits the market agents as put forth by [6] Manufacturers are not in favor of employing state-of-the-art or highly cost techniques since the end user just receives the final product and has to do nothing with the overall food supply chain or the cost incurred in every process. Also, this value does not become monetized. The aforementioned system assures to handle this issue effectively. The Blockchain technique attempts to minimize the intermediaries involved from producer to end user through the translucent fixation of all the data in the supply chain. Moreover, the technique helps in determining the buyers who are ready to pay additional amount for those products which are superior in quality and requires expensive techniques in their production as observed by [7].

Food processors confront lack of adequate information for validating the actual source of products (like in case of hormone-friendly chickens). The processor's products gain trust only if there is precise information related to its processing as well as its source. And, in order to do so, precise data of the complete cultivation procedure has to be obtained from every producers/farmers, according to [8]. In fact the manufacturers must be inspired by the Recyclers to brief about all the processing information willingly, albeit they do not share it. Herein lies the importance of the "Blockchain" which helps the manufacturers to share such information in a secure manner that has to be further precisely validated.

Agricultural supply chain must be equipped with proficient logistics management with consistent supervision of development of agricultural products for achieving the safety of the product. Issues pertaining to food security and also the spoiled food calls out the necessary of improvised traceability system for the entire food supply chain as ascertained by [8], [9]. Also, with the trading of agricultural products to different geographical locations, there must be keen tracking and conformance as per the country specific guidelines [10] [11]. In an agricultural supply chain, product's traceability comprises of information collection and its communication and management by determining its source accurately.

Mao et al. [12] has recommended the approach of consortium Blockchain for improvisation of food trading system. Towards this, a Practical Byzantine Fault Tolerance (PBFT) algorithm has been suggested for the optimization of buyers trading portfolio within the food supply chain. The proposed system is verified by Mao et al. by employing the consortium Blockchain.

The traceability system applied in the food and agricultural supply chain enables tracking and tracing of the products through their origin, mentions [13]. Previously, the conventional traceability system just needed the database technology for allowing the user to store entire information in it and the MySQL database was one such popular means. Though the database technique helps in tracking information but it suffers from human error while performing the data collection.

The research has put forth a popular Blockchain technique referred to as the Ethereum Blockchain, for implementing the traceability system. Using the web3 API, data from Ethereum Blockchain can be revealed. Unlike the critical data, data in the ledger can be viewed by anyone in the Ethereum network. Bigger the network size, higher will be the expense and time involved in generating the blocks. Another disadvantage of using a Blockchain database is the requirement of a specific set of commands for each Blockchain in order to query the information, as mentioned by [14]. That is, searching information in a conventional Blockchain database can be difficult.

For building the prevailing food traceability systems (or models), following were taken into consideration: (i) information and communication techniques along with RFID (Radio Frequency Identification) and NFC (Near Field Communication); (ii) DNA bar-coding as highlighted by [15]. These tools were responsible towards food tracing and tracking along with its identification and monitoring so as to preserve its quality and safety across the entire supply chain.

The food industry Blockchain needs a data platform with intense investment in generous IT elevation and infrastructure, covering the essentials like API's for encrypting and recovering data, creating and maintaining ledger, storage, communications and much more. The block must be automatically generated by the ledgers via cultivation procedures. Governing and paying for such an IT framework is the point of concern as portrayed by [16].

There has been evaluation of various cases pertaining to the supply chain research literature, aiming towards comprehending the Blockchain impact on major supply chain parameters namely: transparency, cost, risk reduction, quality, trust, sustainability and flexibility. Blockchain significantly contributes in elevating transparency and accountability thus raising trust and cooperation amidst the supply chain members as surveyed by [17] for reviewing the implementation traits at an individual level, the UTAUT model can be employed across USA and India.

Allocating Blockchain distinct digital hashes for the products help to trace through the supply chain and also to determine their growth status, batch numbers and expiry dates. In addition, it helps in avoiding food wastage, enabling consumers to understand the environmental foot printing for the foods and guiding towards the distribution process of excess foods. This accurate food and transactions records helps in preventing any sort of fraud as well as determining the source for food borne illness. Moreover, the digital traits of these techniques can enable on-farm data sharing as reported by [18].

By incorporating the GPS directs every locality with respect to the distribution flow that enables traceability of products when encountered with an accident. The system is beneficial in the cold chain distribution, wherein if the temperature and humidity surpasses the given limit controlled on the smart contract, an alert is notified to the concerned personnel for regulating the delivery condition. Wu et al. offers the stakeholders with authenticated visibility of physical delivery process that draws recognition of goods transportation from the supplier to the customer as put forth by [19].

Generally, the stakeholders linked this system which exhibits high rating process to the customers. This could be smoothly achieved by the Blockchain technique by offering transparent, efficacy, confidentiality and also privacy amidst all the contestants. There lies immense scope and advancement for the organizations employing the Blockchain technique and the associated service applications. Irrespective of the overall Blockchain experience, the developers must inculcate at par measures against security risks right from designing to development according to [20].

Its quiet apparent that the agri-food system comprises of multiple stakeholders that desires seeks for superior and secured goods with maximum possible information, discussed by [21] There are chances that the issue of information asymmetry surfaces. According to the discussion, asymmetric information surfaces when the parties are partially intimated during the economic transaction thus preventing the distributions of resources that leads to failure in the marketplace.

The proposed traceability system ascertains to be more farmer friendly and gains popularity in contrast to the prevailing traceability systems. Farmers represent the backbone of the food supply chain and hence their presence is supreme most. The system targets towards offering unconditional and timely support to the farmers in a friendly manner so that maximum of them can benefit from it. Farming has been a keen interest and necessity of numerous people worldwide but the risks involved and lack of knowledge regarding the technique's add process involved, restrain them from doing so as elucidated by [22]. And hence the introduction of a rich experienced panel can uplift the vision and actions of farmers as well as of the entire market. In addition, the farmers can get immensely inspired to independently build their own business. The system would also guarantee better food quality and economical support from insurance agencies in case of any natural calamity, states [23]. The information can be encrypted using Advanced Encryption Algorithm [24].

III. Proposed System

A. Overview

There is proposal of an agriculture traceability system built for various stakeholders that are an integral part of a food supply chain. Towards this, a suitable trust-worthy framework has been built which emphasizes the food security and generating the revenue for every department which are involved. The system caters to and includes the primary entity of seed sellers to the final entity i.e., the end-user of the food supply chain. The centralized feature of the proposed system ascertains the systematic flow of the chain and makes sure that the processes involved are completely scientific oriented. Upon the willingness of the farmer to take up agriculture and get connected to the system, an agreement is made with the insurance company. Various agriculture insurance companies precisely specify their insurance criteria and qualification to the concerned farmers based upon which the farmer can make the decision of selecting one specific company for the policy. This agreement forms the first stage in the system. By examining the consistent information uploaded by the farmer, the insurance company can keep a check on the farmer's land or crops. This information can be accessed by all the departments in the chain for checking the growth status. Next, comes the process of good breed seed selection. This being an essential and critical step since the further results highly depends on this. As the plant grows, the Blockchain technique is employed for managing the further stages. The smart contracts in the Blockchain are responsible for triggering alerts and events upon receiving any function calls as a transaction. This helps the relevant entities to consistently supervise, track and receive appropriate alerts in case of any disaster. Resultant, any natural calamity can be timely and effectively restored within the food supply chain. Fig. 2 clearly depicts the Blockchain based Agricultural Traceability System. The seeds sold to the farmers are allocated standardized identifiers using which digital connectivity and tracking of any transaction between the participating entities can be achieved. Algorithm 1 elaborates the completer registration process of a new farmer. The IPFS helps in recording the consistent and timely growth of the crop. All the captured images related to the crop's growth are time stamped and the smart contract is utilized for storing the IPFS hash of the file. The hash values are determined through the SHA-256. Cereals grown by the farmer are stored in the elevator only after verifying the right humidity, temperature, moisture, heat etc. The processor then buys the cereal for giving the final finish to the product. Thereafter, the finished product is purchased by the distributor for shipping it further to the buyers.

B. Methodology

1) Process of registration: As aforesaid, the achievement of this system hinges around to how many farmers or agri-oriented people have been succesfully asociated with the propoed system. The suggested traceability system must inculclate and manage full fledge database involving all the participants. Hence, formal and precise registration with authentic information is necessary. As soon as a memeber request for joining, the database will be updated by the same. The credentials such as username and password are verified and based on its authenticity, a Blockchain is generated which is then allocated to the relevant member. It is assured that the farmer's information is kept confidential and just the relevant information about a specific farmer is made available to others. Only for the identification of the member's identity, personal information is made use of and its ascertained that there is no misuse or priavcy breach of this information. For logging into the system, the farmer must use the valid credentials such as the userid and password and upon exiting, its enusred that the page or acount is logged out. This is made compulsory so that there is no misuse of the account. Algorithm 1 depicts the registration process of a new member.

2) Agreementation: The proposed system enables the farmers and insurance companies to carry out the business practices in coordination and a smooth manner. It cut downs the farmer's effort and time in identifying the apt insurance schemes and neither the insurance companies require to spend theie energy and time in selling their policies. Based on the detail briefing given by the famer of his plan, the insurance company put forth their norms and condition once the farmer's plan is verified. The farmer is free to make the policy selection according to his contentment. The centralized system acts as an evidence for the entire process taking place between the famer and insurance company. Every single step gets recorded in the database as an evidence. This can be later useful, in case any conflicts or arguments emerges between the parties. The system is upartial against any member and it ensures that no conflicts takes place between them. That is, whether it be farmers, insurance company, processors or any department, all will be treated equally and can benefit from the sytem in a trasparent manner. This helps in building a united, trustworthy and sturdy food supply chain. The system ascertains that, as and when the farmer intimates regarding a natural calamity, a notification or alert is trigerred and transmitted to the relevant company. There is a consistent follow up by the system until the farmer receives the insured amount as per the terms and conditions. Hence, with this fail proof system, there is trusted and transparent relationship is etsbalished between all the participants. There may be cases or situations which may not comply with the given norms or conditions, in such scenarios the system automatically agrees or support the company policies rather than the client's plea. Hence, the system presents the whole agreement in a precise and transparent manner so that both the parties remain vigilant about the same.

3) Seed selection: Once the registration is completed and the contract is signed, the farmer needs to buy large bulk of good breed seed from the seed seller. This tends to be the most important decision since the entire crop production depends on the quality and fertility of the seed. The farmer gets to choose from the enlisted and registered seed companies present in the system. Algorithm 2 elaborates the process of seed selling by the relevant seed company. On signing up the initial contract, the smart contract verifies the farmer's registration and subsequently the payment for the seeds bought is done via Agri Insurance. That is the required money for purchasing the seeds is provided by the Agri Insurance to the farmers. If all goes well and successful, then the contract state becomes as Seed Request Submitted, the farmer state becomes Wait for Seeds and seed company state becomes Agree to Sell. This change of 'state' is notified and updated to all the participants in the supply chain. If not updated, then the state of contract reverts back to initial state thus terminating the transaction.

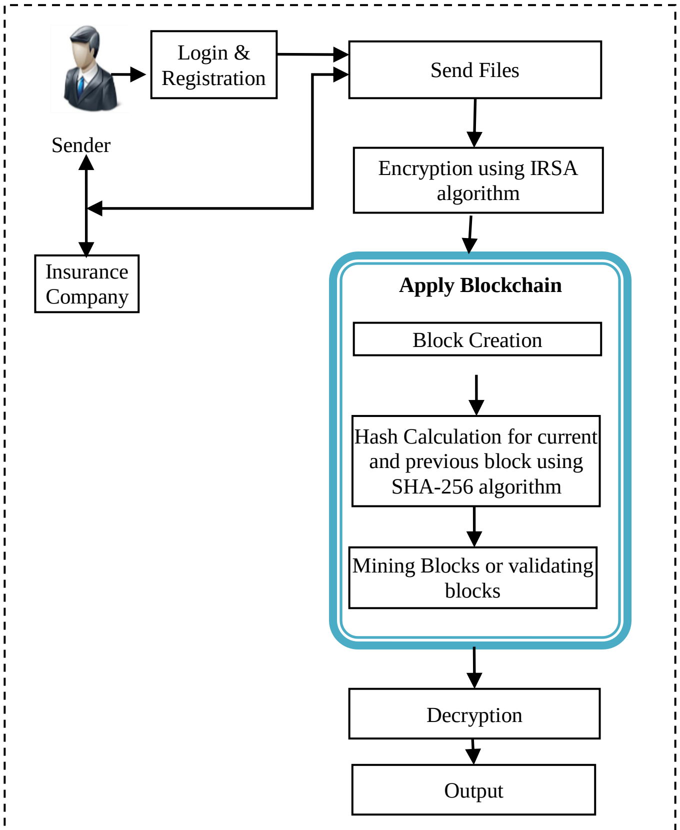


Fig. 2. Architecture Diagram for Blockchain based Agricultural Traceability System.

Algorithm 1 New Farmer Registration

Inputs : Request for New Registration

Nodes of the current Blockchain network(N)

Outputs: Newly entered farmer f

Step 1: (Kpub,Kpr) = generatekeys()

Step 2: uID→create userID();BlockchainAddress() + (Kpub,Kpr)

Step 3: Addr →create BlockchainAddress() + (Kpub,Kpr)

Step 4: (Uid,kpr) →safelystore(uID,Kpr)

Step 5: Walt→creates BlockchainWallet() + (Kpub, Kpr)

Step 6: for every node, n ϵ N loop

Step 7: distributeWallet(n,Walt)

Step 8: end for

Step 9: u←verifiedNewfarmer()

Algorithm 2: Selling Seed to Farmers

Step 1: Let G1, G2....Gn be the registered farmers,

Let t1,t2 be the token of farmer and Seed Distributor respectively

Consider Quantity, SType, SBrand,SPrice

Step 5: Contract State is created

Step 6: Farmer State→SeedsReq

Step 7: Seed Distributor State→Ready

Step 8: Restricting access to only registered farmers i.e. g ϵG

Step 9: If farmers=G and SPrice=paid then

Step 10: Contract state→SReqSubmitted

Step 11: Farmer State→WaitforSeeds

Step 12: SDistributor State→AgrreToSell

Step 13: Notification msg to state sale of seeds

Step 14: end

Step 15: Else

Step 16: Revert contract state and display an error msg

Step 17: End

4) Blockchain technology: Post seed trading process, the farmer enters the Blockchain or IPFS (Interplanetary File System) system. IPFS refers to a protocol within the Blockchain and depicts a P2P networks that facilitates file storage and transfer for a shared file structure. It grants users to accept host content and is based upon a decentralized system of user-operators, holding a segment of overall data. The IRSA algorithm is employed for encrypting the information sent by the farmer before it is stored. The information comprises of images pertaining to growth measurement, weather and soil reports which gets digitally recorded in the database.

In the Blockchain and IPFS system, every information whether it be data, reports or images, that are signed digital and attributes in a specific manner. The fields have installed cameras for capturing the images automatically which are then transmitted to the Blockchain where they get recorded. These cameras are specially crafted and provided to the farmers such that they cannot be tampered. Hence the images captured by them remain unaltered and authentic which can be trusted by all the entities in the Blockchain. Moreover, any illegitimate user is strictly prohibited from accessing this information.

In fact, the authorized user can also access just the permitted information. Data residing in the Blockchain is utilized for performing certain computations such as generating values for the observation including the hash value. These computed values are of utmost significance and directs if the farmer is following the correct method or not. Basically, there are two prime reasons for computing the hash value. First, for determining the growth rate, and second to track if the farmer is proceeding in the suggested and agreed manner. In case of any forgery or misconduct, penalties are imposed on the farmer which is being programmed automatically in the Blockchain. Hash value is computed using the SHA-256.

The food supply traceability system that incorporates smart contracts greatly aids in transmitting unimpeded information to all organisations in the supply chain without the involvement or control of any central authority. Each single transaction, right from seed selling, to the amount of crop produced and sold is recorded precisely which can be validated thoroughly. Like for instance, set quantity of cereal sold amidst the stakeholders with respect to the specified norms is not subject to change. Moreover, mixing of cereals with different quality is not allowed for selling. Since it is difficult to monitor the status of the field and crop growth, the IPFS helps in uploading the crop and land images periodically which can be verified and accepted by all.

For assuring elevated Quality Compliance, all the corresponding transactions amidst all the entities is traced. The sensors are highly equipped with the feature of sending consistent alerts pertaining to the cop and land status. The Blockchain technique prohibits modification of any info or alerts and permits the accessibility of information to all the authentic entities in a secure and decentralized way without the involvement of any central figure. Apparently, there can be cases of fraud by reporting misleading information by the stakeholders. Such data is automatically verified by the Blockchain whether it appears to be fraudulent or not. For guarding and preventing against such frauds, Blockchain can be programmed to nullify the supply chain process thereby imposing penalties on such stakeholders. By performing this precise and undoubted traceability s presented to the supply chain stakeholders.

5) Support to the farmer: While carrying out farming, the farmers need wise and effective suggestions from the system. Towards this, it is essential that the farmers upload timely and precise information and images pertaining to the crops. Certain essential factors that must be reported includes: growth of the crop, temperature, average rain measurement and soil components along with its percentage. As and when the values are fed by the farmer, the database must be updated. Using Algorithm 2, the transactions are mined and hash values are created. The database has a collection of information related to the crop. In addition, it has the fertilization requirement for the relevant crops at every level and in varying weather conditions. As the farmer updates regarding his crop, there are suggestions provided regarding the further stages. This enables the farmer to nourish its crop with the right amount of manure thus making the entire farming economical, Fig. 3.

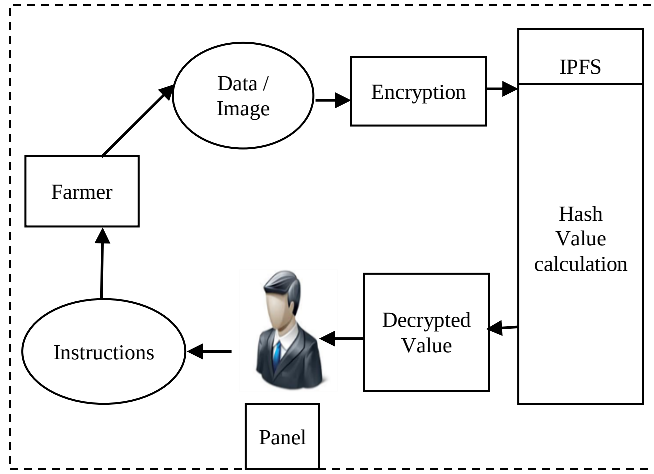


Fig. 3. Instruction Forming to the Farmer.

6) Trade in the end: After the farmer produces quality crop and there is completion of all the processes, the crop or the end product is ready for selling. Based upon the market value there can be negotiation in the price with the distributor and vice-versa. Each transaction taking place is recorded correspondingly in the system for performing subsequent processes. Upon completing the entire process, the famer is eligible for selling their products to the distributors. The system even records the overall interaction and trading of the farmer with the distributor. Above all, the system empowers and attempts to fulfil the farmer's effort without being partial or biased in its protocol.

Farmer Satisfaction Index $= p(pr, pg) * (100 - \sum(pd)/n)$

Here p(pr,pg) depicts the relation of pr (pay-outs requisite), pg (pay-outs given) and also pd resembles its pay-out variations (that is the variation among the required pay-out with respect to crop loss and pay-out with respect to the design agreements), n shows the total years. Present key factor equivalently caters to under and over-payments that is inconsiderable for the development of key insurance.

Lately, the retailers have flourished their own business with in depth understanding of stock holders and the product they comply with. Only the registered retailers are granted accessibility with respect to these essential attributes provided namely: Manufacturing date of the goods, purchase date, and the number of products Sold. Verification is done against the fulfilment of accepted sales agreement and product payment. If verified and true then the transaction is executed by the contract and the state becomes.

The state of the retailer turns to ProdDelSuccessful. For ensuring the successful delivery of the product, the contract notifies the successful delivery to the retailer. In case of any failure, the contract state turns into SaleRequestDenied. Eventually, the customer-retailer purchase or transaction is traced. The end user or the customer depicts the final entity of the entire network. At the beginning, the art of customer is ReadyToBuy. The smart contract allows those Customers who have the essential product information such as the Customer token, Retailer token, Purchased date, Sales Id and Product Id.

Once the payment is done successfully, the contract state becomes ProdSoldToCust, and the customer state depicts SuccessfulPurchase. In case any sales failure occurs, the contract intimates everyone in the network.

IV. Result and Discussion

The proposed system is tested in Blockchain network. The entire system is written in Java. The test system depicts briefly in Table I. Two major indices of system performance are: information uploading time and information response time. Fig. 4 to 6 exhibits the system test results of the mentioned indices.

Fig. 4 clearly reveals that the information upload time gets impacted according to the interval of upload request interval. This upload request interval varies between 100 - 900 times per second and based upon that the upload information time witnesses a rise from nearly 7s to 47s. A genuine rise of 600 to 700 is observed with the upload period rising around 25s to 42s. This reason may be attributed to the limitation of consensus agreement in Blockchain, whereas the single block which possess the restriction of transactions.

Fig. 5 and 6 exhibits that outcome of response period impacted at regular interval of transaction request. Fig. 4 presents the information response period exceeding from 2ms to 5ms with respect to the rise in capacity of on-chain transaction data ranges 1G - 9G. According to Fig. 5, the information response period increases a bit from 2.2ms - 3.2ms, with respect to the transaction request rising from 1,000 to 9,000 per second with the on-chain transaction data fixed to 1G. This is due to the mode of data query and P2P networking.

TABLE I. SOFTWARE ENVIRONMENT

Operating System	Windows 7
Development Platform	Eclipse Luna
Blockchain Entity	Ethereum
Running Platform	Java 8.0

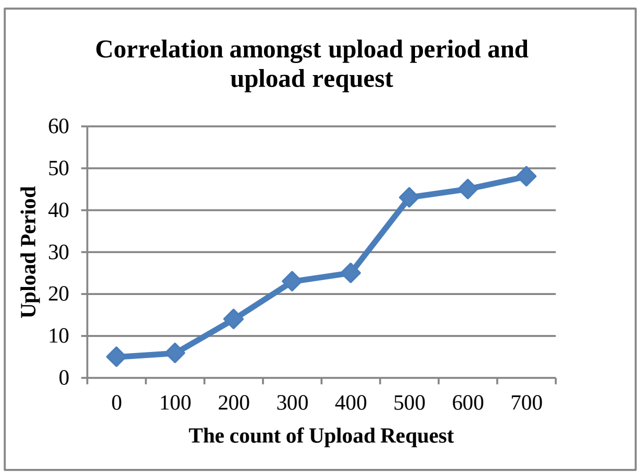


Fig. 4. Correlation amongst Upload Period and Upload Request.

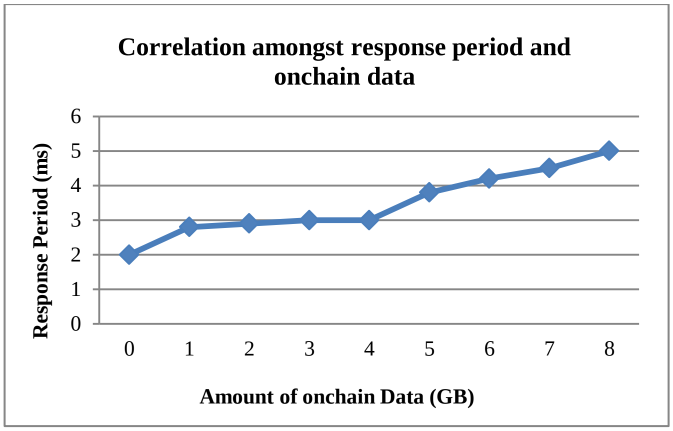


Fig. 5. Correlation amongst Response Period and Onchain Data.

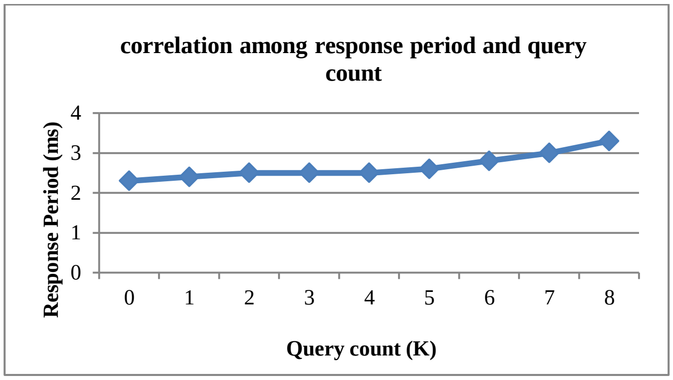


Fig. 6. The Correlation among Response Period and Query Count.

V. Simulation Result

Online dataset are being utilized for the collection of data and JAVA environment is used for the development purpose. Also entire dataset along with the results generated is managed and stored with the help of MySQL. Attributes such as crop, temperature, wind speed, irrigation, fertilizers and rainfall are main attributes. Other attributes used are seeds area etc.

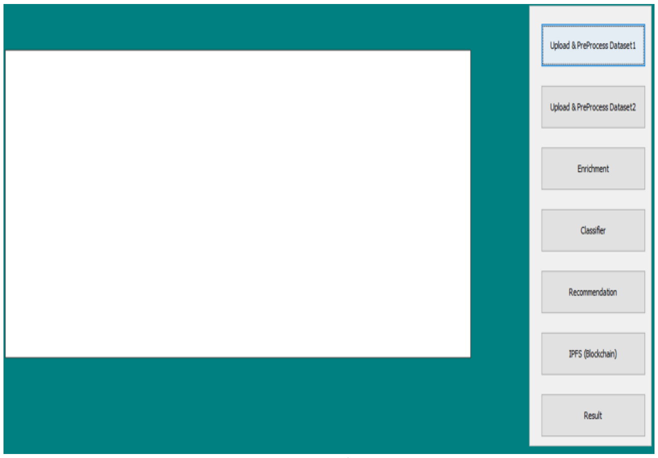
Fig. 7b depicts collection of dataset information taken from Internet dataset. The Dataset are assembled been pre-processed to get the desired output. The dataset uploaded is of weather information.

Fig. 7c Pre-processing of dataset, which removes the unwanted data to get the desired output that are been uploaded from the Internet dataset.

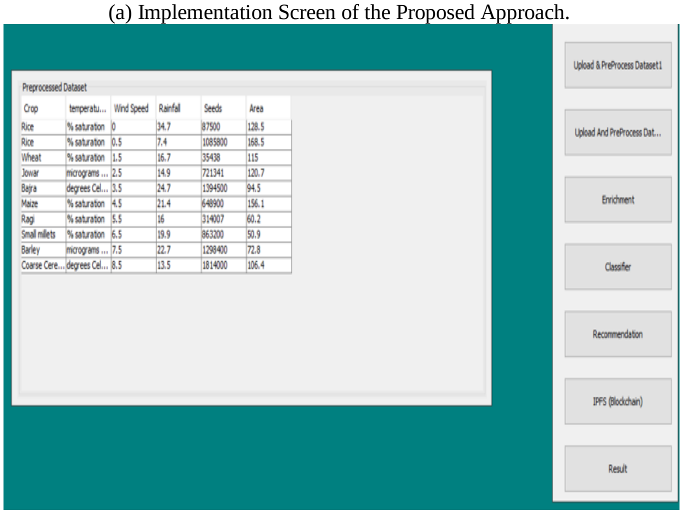
Fig. 7d depicts for clustering new Crop dataset, 1. Group classify the data depends on same type of soil, crops and the rainfall 2. Group classify data depends on type of soil, crop, soil, and the temperature 3. Then Label the crop as Crop Name, soil as Soil Type, Water as ‘Required water’, Rainfall as Average Rainfall, Temperature as Required Temp.

Fig. 7e Recommends crop based on selected soil type, rainfall and temperature to get the desired output.

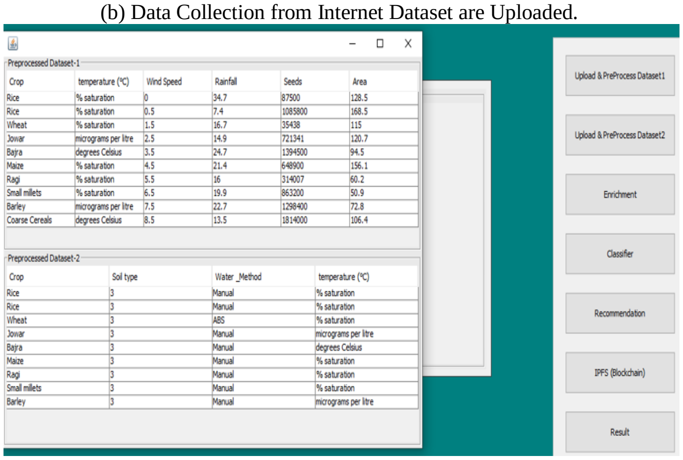
Fig. 7f The Blockchain IPFS system is applied in every information which are digitally signed Data residing in the Blockchain and is utilized in performing certain computations such as generating values for the observation including the hash value.



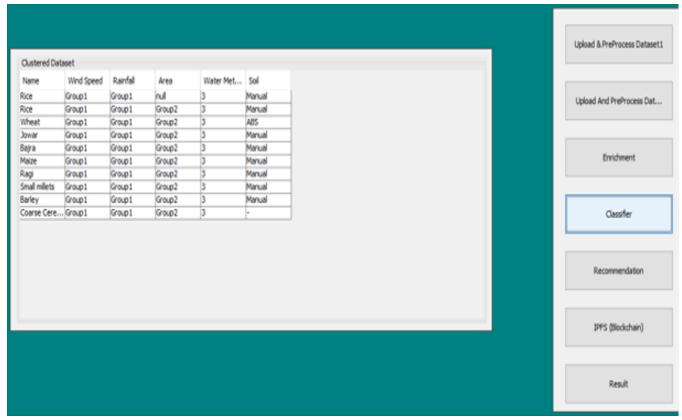
(a) Implementation Screen of the Proposed Approach.



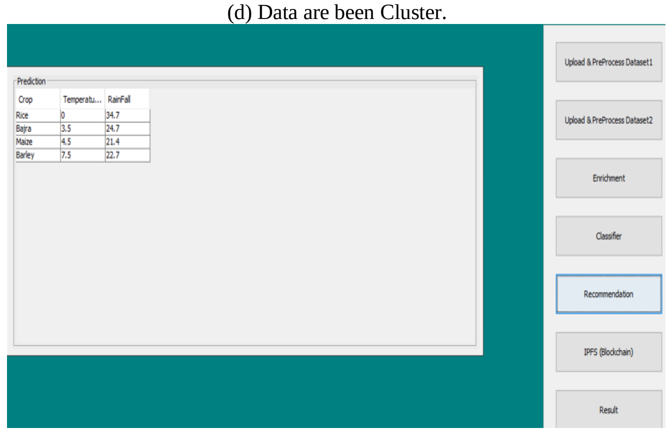
(b) Data Collection from Internet Dataset are Uploaded.



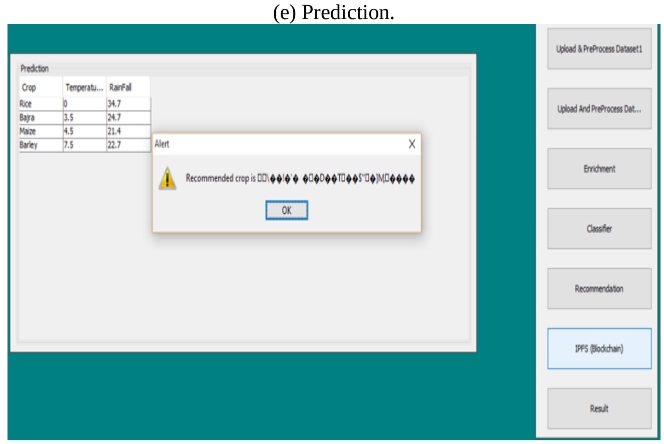
(c) Pre-processed Dataset.



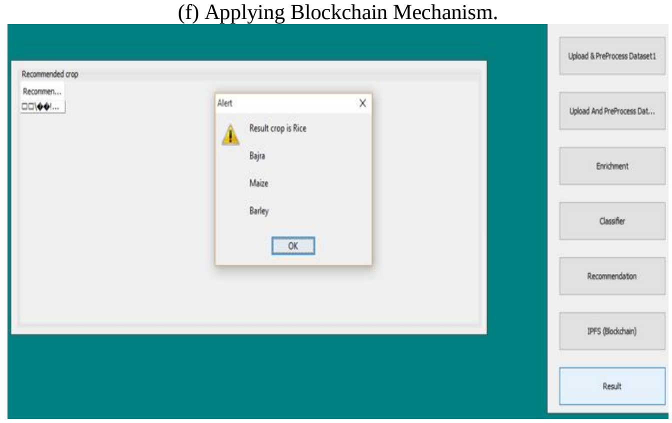
(d) Data are been Cluster.



(e) Prediction.



(f) Applying Blockchain Mechanism.



(g) Result Extraction.

Fig. 7. Decrypt the IPFS Folder to Get the Desired Result.

TABLE II. COMPARISON OF PROPOSED SYSTEM WITH EXISTING SYSTEMS

Attributes	Fine grained access [27]	Food Supply Chain [28]	Deep Reinforcement Learning [26]	Soybean Traceability [25]	Yield Estimation System [29]	Fruit and Vegetable Agricultural products [30]	Proposed System
Agri-Insurance	no	no	no	no	no	no	yes
Government Panel	no	no	no	no	no	no	yes
Product Traceability	yes	yes	yes	yes	no	yes	yes
Timely Payouts	yes	no	no	no	no	no	yes
Responsibility	yes	yes	yes	yes	yes	yes	yes
Integrity	yes	yes	yes	yes	yes	yes	yes
Recommendation	no	no	no	no	no	no	yes

The predominant attributes of the proposed agricultural traceability system for food production shown in Table II is expounded below:

1) Agri-Insurance: Insurance module is implemented in the agricultural traceability system for food production to provide insurance to the farmers to buy seeds as well as during natural calamities. Upon the willingness of the farmer to take up agriculture and get connected to the system, an agreement is made with Agri-Insurance. Natural calamities like flood, drought, earth quake, unpredictable weather or disaster, loss due to harvest, cyclone, the farmer can claim for the insurance. The system checks, when the farmer intimates regarding a natural calamity, a notification or alert is triggered and transmitted to the Agri-Insurance company. Agri-Insurance checks the timely weather incidents like rainfall, temperature, humidity, wind speed etc and associated payout enlisted on smart contract and enable prompt payment to the farmers.

2) Government panel: The farmers can get wise suggestion from the advisory board members or government panel. Which crop should be cultivated in a particular season to increase the crop yield and maximize the profit, type of fertilizers must be used, soil type, irrigation method. The farmers will be updated to adapt new methods and planting techniques. This panel members helps farmers understand their crops better and about soil and quality of seed, organic farming, crop rotation, the potentiality to examine soil nutrients to enhance better outcome.

3) Product traceability: Product Traceability helps the stake holders and farmers to track and trace the product in the supply chain management. When the end user scanning QR code, complete information about the product can be traced.

4) Timely payouts: The farmers can get financial assistance during disasters occurs to food crops. The insurance amount is processed and settled on time.

5) Responsibility: The proposed agri-supply system's stakeholders are integrated and de-centralized. All the transactions are transparent and secured so that the intruders are denied to hack the data from the supply chain system. Farmers registered in the Blockchain network, also accepted agreement can buy seeds, cultivate, sell and trade at the end to customers. The proposed system guarantee profit to all departments involved in the supply chain network minimizing the cost required in processing the food and build trust among the stakeholders.

6) Integrity: The transaction data which is stored in Blockchain and the broadcasted to all the stakeholders are immutable and persistent and it cannot be changed, tampered or deleted. The transactions taking place are transparent that have facility to track and trace the food crops from the beginning of the supply chain till the consumer purchase the finished product.

7) Recommendation: The farmers are recommended to acquire better crop yield such as the appropriate irrigation method, utilization of organic manure and fertilizers, method for crop rotation, high yield variety seeds, temperature, water, rainfall, soil type, humidity, weather, wind.

The proposed agri-supply chain using Blockchain, helps the farmers to buy seeds from authorized seed seller. The farmers can use agri-insurance to buy seeds and also during natural calamities. The finished goods can be send to the distributors and retailers and finally to the end consumers. The government panel members are the advisors who give suggestions to the farmers to get better yield.

VI. CONCLUSION AND FUTURE WORK

The Blockchain method has seen explosive growth in recent years. The Blockchain approach has seen a clear and considerable increase in the agriculture sphere. This technique makes strong promises about food safety since it stresses the traceability of an agricultural product's origins as well as the validation of agricultural inputs. There is a safeguard in place to prevent the source of contamination from being traced. The Blockchain based agri-supply chain system is designed to get profit for all entities involved in the network.Though the Blockchain professes to be unbiased, it assures that low-income farmers benefit from small payments through Agri-Insurance. In future, when the end-user scan the QR code, the farming operations, information about the product, traceability information and security information can be traced. Also, the price of Bitcoin can be forecasted using intelligence techniques.

REFERENCES

[1] Catherine A. Nguegan Nguegan and, Supply chain management problems in the food processing industry: Implications for business performance, January 2017.

[2] Chengedzai Mafini, Traceability In Food And agricultural Products, international trade cantre, 2017.

[3] Brigadier j matta, doctor of philosophy in management to jiwaji university, Gwalior, A study of supply chain management in food industry, 2016.

[4] Richard J. Currie, Inflation: Its Impact On Retail Trading, Retail and Distribution Management, ISSN: 0307-2363, 2018.

[5] Gabriel A. Huppé, Sabrina Shaw, Jason Dion, Vivek Voora,Food Price Inflation and FoodSecurity: A case study, International Institute for Sustainable Development.

[6] D. S. Dayana and G. Kalpana, "A Public-Blockchain based De-Centralized Application Framework for Agri Supply Chain Management System," *2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2022, pp. 267-275.

[7] Using Blockchain Technology to Manage Clinical Trials Data: A Proof-of-Concept Study,2018, JMR Informatics-PMCID: PMC6320404, PMID: 30578196.

[8] Iansiti, M. and K. R. Lachlan, "The Truth About Blockchain" Havard Business Review,pp. 118-127, 2017.

[9] Manski, S. "Building the blockchain world: Technological commonwealth or just more of the same?" Strategic Chan, Vol. 26, No. 5, pp. 511-522, 2017.

[10] Dayana.D.S and Kalpana.G, "Survey on Agri-Food Supply Chain Using Blockchain," *2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 2021, pp. 1619-1626.

[11] Vukolić, M. "The quest for scalable blockchain fabric: Proof-of-work vs. BFT replication,"International Workshop on Open Problems in Network Security, Springer, 2015.

[12] Leng, K. Bi, Y. Jing, L. Fu, H-C. and Van Nieuwenhuyse, I. 2018 "Research on agricultural supply chain system with double chain architecture based on blockchain technology," Future Generation Computer Systems,Vol. 86, pp. 641-9.

[13] Bermeo-Almeida, O. Cardenas-Rodriguez, M. Samaniego-Cobo, T. Ferruzola-Gómez, E. Cabezas-Cabezas, R. and Bazán-Vera, W. 2018 "Blockchain in agriculture: A systematic literature review," Communications in Computer and Information Science, Vol.883, pp. 44-56.

[14] Patil, A S. Tama, B A. Park, Y. and Rhee, K. 2018. "A framework for blockchain based secure smart green house farming," Lecture Notes in Electrical Engineering, Vol.474, pp. 1162-7.

[15] Bosona, T. and Gebresenbet, G. "Food traceability as an integral part of logistics management in food and agricultural supply chain," Food Control, Vol. 33, No. 2, pp. 32-48, 2013.

[16] Badia-Melis, R. Mishra, P. and Ruiz-García, L. "Food traceability: New trends and recent advances.A review," Food Control, Vol. 57, pp. 393-401, 2015.

[17] Kamilaris, A. Fonts, A and Prenafeta-Bold, F.X. "The Rise of Blockchain Technology in Agriculture and Food Supply Chain", Elsevier Trends in Food Science & Technology, Vol. 91, 2019.

[18] Zhang, Y. and Zou, T. "A Review of Food Traceability in Food Supply Chain", IMECS, 2017.

[19] Kumar, M.V. and Iyengar, N. "A framework for blockchain technology in rice supply chain management," Adv. Sci. Technol. Lett.,Vol. 146, pp. 125–130, 2017.

[20] Dabbene, F. Gay, P. and Tortia, C. "Traceability issues in food supply chain management: a review," Biosyst. Eng., Vol. 120, pp. 65–80, 2014.

[21] Christidis, K. and Devetsikiotis, M. "Blockchains and smart contracts for the Internet of Things,"IEEE Access, Vol. 4, pp. 2292-2303, 2016.

[22] Foteinis, S. 2018. "Bitcoin's alarming carbon footprint,"Nature, Vol. 554, pp.169–169, 2018.

[23] Kewell, B. Adams, R. and Parry, G. 2017. "Blockchain for good?"Strat. Change, Vol. 26, No. 5, pp. 429–437.

[24] Dayana, D. S. "An Efficient Approach for Network Mobility Based on AES Algorithm." Advanced Materials Research, vol. 984–985, Trans Tech Publications, Ltd., July 2014, pp. 1269–1275. doi:10.4028/www.scientific.net/amr.984-985.1269.

[25] Khaled Salah, Nishara Nizamuddin, Raja Jayaraman, And Mohammad Omar "Blockchain-Based Soybean Traceability in Agricultural Supply Chain", IEEE Access, vol. 7, pp. 73295-73305, 2019.

[26] Huilin Chen, Zheyi Chen, Feiting Lin, And Peifen Zhuang "Effective Management for Blockchain-Based Agri-Food Supply Chains Using Deep Reinforcement Learning", IEEE Access, vol. 9, 2021, pp. 36008-36018.

[27] S. Wang, Y. Zhang, and Y. Zhang, ''A blockchain-based framework for data sharing with fine-grained access control in decentralized storage systems,'' IEEE Access, vol. 6, pp. 38437–38450, 2018.

[28] Mohan M, Shanjay KM , Subashchandrabose M , and Saravanakumar C "Food Supply Chain Using Blockchain.

[29] Murat Osmanoglu, Bulent Tugrul, Tuncay Dogantuna, and Erkan Bostanc, "An Effective Yield Estimation System Based on Blockchain Technology", IEEE Transactions On Engineering Management, VOL. 67, pp. 1157-1168.

[30] Xinting Yang, Mengqi Li, Huajing Yu, Mingting Wang, Daming Xu, And Chuanheng Sun, "A Trusted Blockchain-Based Traceability System for Fruit and Vegetable Agricultural Products", IEEE Access, vol. 9, 2021, pp. 36282- 36293.

A Comprehensive Analysis of Blockchain-based Cryptocurrency Mining Impact on Energy Consumption

Md Rafiqul Islam[1]*, Muhammad Mahbubur Rashid[2]*
Mohammed Ataur Rahman[3], Muslim Har Sani Bin Mohamad[4], Abd Halim Bin Embong[5]
Mechatronics Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia[1, 2, 5]
Mechanical Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia[3]
Accounting, International Islamic University Malaysia, Kuala Lumpur, Malaysia[4]

***Abstract*—Blockchain already has gained popularity due to its highly secured network and same time enormous computational power consumption has become an undifferentiated debate among the users. A Blockchain network is reliable, secure, transparent, and immutable where the transactions cannot be reversed between sender and receiver. Blockchain technology is not only used for mining cryptocurrency, it has other applications in different sectors like agriculture, education, insurance, etc., but the noticeable concern is still energy consumption. On the other hand, there is a significant impact on the environment due to the use of excessive energy for mining cryptocurrency which releases more carbon dioxide (CO_2) in nature. The Proof-of-Work (PoW) algorithm is used for mining 'Bitcoin' which is consumed enormous computational power. However, an alternative solution like Proof-of-Stake (PoS) consensus protocol has been proposed to use instead of the Proof-of-Work algorithm for mining cryptocurrencies which is capable to reduce the significant amount of energy consumption. Not only that, but the use of renewable energy can also be an alternate option to use the Proof-of-Work algorithm for mining cryptocurrencies which is environment friendly. This paper aims to highlight blockchain technology, the energy consumption and impact on the environment, energy reducing method by using PoS consensus protocol instead of using the PoW algorithm, and discussion with some recommendations.**

Keywords*—Blockchain; cryptocurrency; bitcoin; Proof-of-Work (PoW); Proof-of-Stake (PoS)

I. Introduction

Blockchain already has gained popularity due to its highly secured network. Bitcoin is the first cryptocurrency that was introduced by the pseudonymous name Satoshi Nakamoto [1]. Blockchain technology was used to develop this cryptocurrency. Bitcoin is also called digital currency which is used for transacting the currencies between peer-to-peer (P2P) into the blockchain secured network without the intervention of a third part intermediator [2-3].

A Blockchain network is reliable, secure, transparent, and immutable where the transactions cannot be reversed between sender and receiver [4]. Another strong security feature cryptographic hashing algorithm is used to maintain the data integrity by applying the encryption and decryption method this algorithm is also used for joining the blocks in chronological order and forming a chain [5-6]. Decentralization of data and storing the same in all existing nodes in the network is an important feature to provide data backup [7]. The distributed ledger is used for keeping all the information in different geographical locations where every user is capable to view the data [8]. Proof-of-Work (PoW) algorithm is a very strong, secured, and powerful mechanism that is used for mining a new block into the blockchain network and keeping the transaction records in this block permanently [9-10]. In the last couple of years, a few numbers of a consortium like r3 Corda, Hyperledger Fabric, Ethereum, etc. have been formed in the banking sectors where some of them are already doing the transactions for cross-border payment among themselves on a pilot basis [11-13]. The not only banking sector, other sectors like healthcare, education, insurance, agriculture, etc. are highly considered for keeping their information in this blockchain network with the consideration of security and transparency [14].

Though blockchain technology has several advantages, some significant challenges need to be addressed by the implementation. Power consumption is one of the major concerns for people across the globe [15]. Proof-of-Work is the powerful algorithm of blockchain technology to mine a new block into the blockchain network which consumes the maximum amount of power [16]. Due to the enormous consumption of power, there is a huge impact on the environment due to the emission of carbon dioxide. Maximum power generates from fossil fuels around the world which help to produce carbon dioxide and pollute the environment simultaneously [17-18]. This paper will briefly discuss the background and architecture of blockchain in section II, cryptocurrency mining procedure by using the PoW mechanism in section III, methods of reducing the impact of producing cryptocurrency mining power and other impacts as well in section IV, and finally, the paper will be concluded by section 5.

II. Blockchain Background and Architecture

First, the blockchain concept was introduced in 1990. In 1992 the trusted data tempering protocol was ensured in the format of a chain that guarantees the privacy of data in the form of the integrity of the records [19]. Fig. 1 shows the history of blockchain technology.

*Corresponding Author

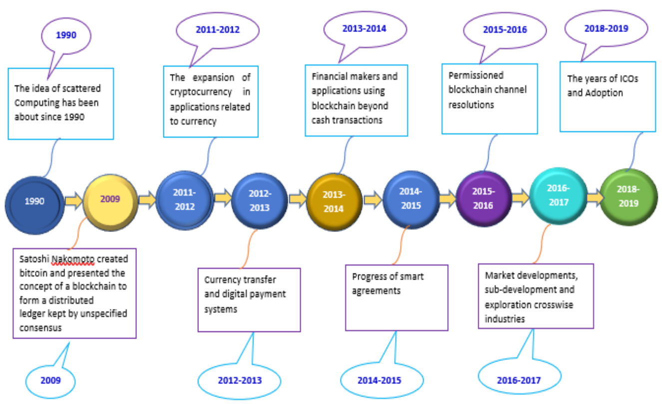


Fig. 1. A Brief Blockchain History [20].

The pseudonymous name Satoshi Nakamoto first introduced a new cryptocurrency called Bitcoin in 2009 for the peer-to-peer transaction into the blockchain network [1]. Gradually blockchain technology was becoming popular and more cryptocurrencies like Litecoin, Namecoin, and PPCoin were founded from 2011 to 2012 [21]. Some other remarkable cryptocurrencies Ripple, Dogecoin, and Stellar were introduced in the year 2012-2014 [22-23]. Another emerging cryptocurrency called Ethereum was came-up in the market by using a famous Ethereum blockchain network which was launched in 2015 [13]. Nowadays, there is a significant number of cryptocurrencies already have been introduced with their network infrastructures which are shown in Fig. 2. The maximum use of this blockchain technology and cryptocurrencies in the financial sectors. With consideration of the security, transparency, and reliability of this network, some other sectors like health care, insurance, education, energy, and agriculture have already started to implement blockchain-based applications in their companies in different countries around the world.

A. Blockchain Architecture

The architecture of blockchain is completely different from a traditional centralized database management system where the information is stored digitally in all the connected nodes in a decentralized manner rather than a centralized version [25]. All the blocks in the blockchain network are connected through a chain that a cryptographic algorithm uses to maintain security [26]. However, the block maintains three different parts data section, a hash of the present block, and the hash of the previous block. The useability of the blockchain network depends on the nature of the business aspect. The cryptocurrency blockchain keeps the transactional records in the nodes. Every time hash value will be changed if any new data will add or amended to the data in the blocks of the network [27]. The hash value is generated by using the hash function and it converts the data in a fixed format like 32 bits, 64 bits, 128 bits, 256 bits whatever may be the size of the input data [28]. It depends on the nature of the hash function. The previous block hash connects to the next block to form a blockchain network. The initial block is called the genesis block where the hash value is 0 [29]. For example, in Fig. 3, a simple blockchain has been presented where the genesis block contains the 0 (zero) hash value for the initial, 7A4RFGY8 is the present block hash for block 1, 8GJI7RGY represents the block 2 hash, and so on where all the blocks have connected each other cryptographically. Due to this complex format, manipulation of records is almost impossible. To compromise any bit of information in the network, all the consequence hash values need to be changed in the entire network, and at least 51% of nodes to be agreed to make any amendment [30]. The distributed ledger is used to make transaction into the blockchain network which is an immutable ledger and help to resist the temperament of the data in the block [31].

B. Data Signing and Verification Process

The data signing and verification process is shown in Fig. 4. The plain text format data pass into the hash function and convert the data as the fixed hash format. The hash values are encrypted by using the private key of the sender and converted as a digital signature [32]. This digital signature is attached to the data and creates digitally signed data which send to the receiver.

On the other hand, after receiving the digitally signed data by the receiver, the receiver verifies the originality of the data. First, the receiver decrypts the encrypted data by using the signatory's public key which helps to find the hash value [33]. Second, the receiver then passes the plain text data sent by the sender will pass through the hash function and get the hash value. The receiver will then check the integrity and originality of the data by matching both hash values [34].

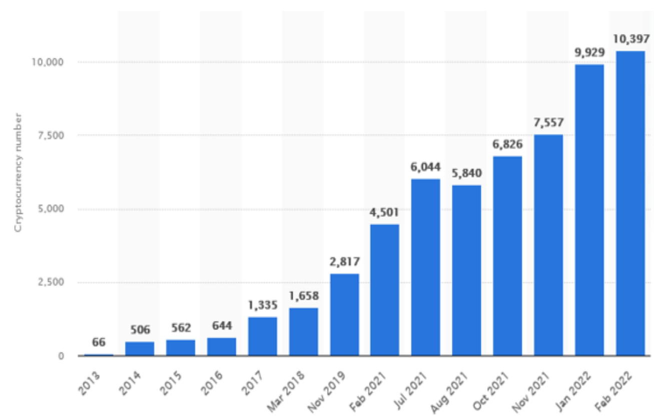


Fig. 2. Cryptocurrency Statistics [24].

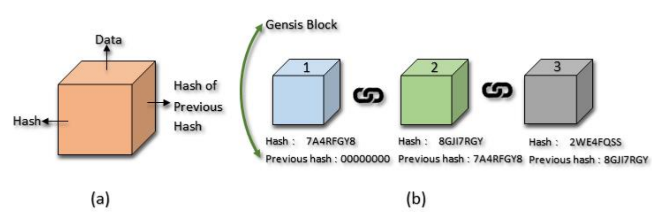


Fig. 3. Blockchain Structure.

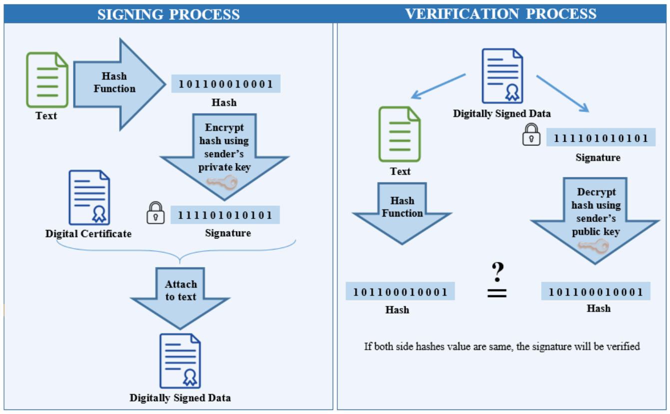


Fig. 4. Digital Data Certification Process in the Blockchain.

C. Transaction Process into the Blockchain Network

There are a few steps that must follow for adding a transaction into the blockchain network which are shown in Fig. 5. Before transacting into the network, the authentication and authorization process needs to be done by using cryptographic keys and proof of work simultaneously [35]. As shown in Fig. 5, the transaction process describes the popular bitcoin network which runs in a public network where anyone can participate. Initially, the transaction initiator sends the transaction request into the public bitcoin network for authentication and adding the data into the network. The transaction requests all the nodes into the network for verification for the same [36]. Once the transaction is validated by the nodes, the block is then added to the blockchain network, and all the participants' nodes are rewarded by bitcoin for participating for proof of work [37]. After that, the transaction will be completed by updating the blockchain network. Private and public keys are used to validate the transactions where the private key is used for digitally signing the data and the public key is used for decrypting the data at the receiver end [38-39]. On the other hand, authorization is required to add the data into the block through a consensus mechanism which means the majority nodes need to be agreed to complete the transaction [40].

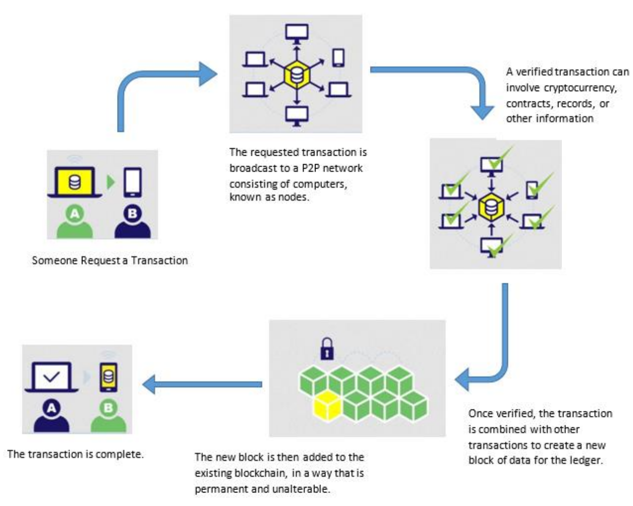


Fig. 5. Transaction Process in the Blockchain Network.

III. Energy Consumption and Impact of Blockchain

Though blockchain is a highly secured platform, energy consumption becomes very high especially when the mining process has occurred which is a big concern in the industry [41-42]. The validation for proof of work algorithm is a complex as well as trial and error process for matching with hash value [43]. For example, an average of ten minutes is required to mine a new block for the bitcoin miners. In terms of data security and integrity, it is highly recommended for using this technology where energy consumption is a big concern for different entities [44].

In an article has been published by Raynor de Best in February 2022, the total number of cryptocurrencies more than 10,000 shown in Fig. 2 which have been drastically increased from 2013 [45]. However, all these digital currencies are not active in the market. Different algorithms are used for mining these cryptocurrencies. Among the plenty of digital currencies, bitcoin is the most popular and useable cryptocurrency across the globe which consume an exorbitant amount of energy for mining bitcoin. There is an article written by Eugene Kim in September 2021 in business insider mentioned that approximately 0.5% electricity of total global electricity used for mining the bitcoin which is approximately more than 7 (seven) times the energy consumption by google per year [46]. The energy consumption for mining the bitcoin from 2017 to 2021 is presented in Table I, the electricity consumption rate has been significantly increased every year, and this consumption in the year 2021 is nearly 6 times that of 2017. From Table I, the number of bitcoins mining gradually decreased from 2017 to 2022. On the other hand, the usage of electricity for mining bitcoin gradually increased over the same period, and finally, bitcoin mining costs were increased.

On the other hand, the Ethereum blockchain uses much less electricity than the bitcoin blockchain for mining Ether tokens which is shown in Table II [47]. As per Table II, the total electricity of 145.39 TWh has been used to mine approximately 59.5 million ethers whereas 480.82 TWh electricity has been used to mine 2.83 million bitcoins only which is shown in Table I. The average mining cost for one bitcoin is USD 19,229, whereas the average mining cost for one ether is USD 217.52 only which is shown in Table III.

An article was published in "The New York Times" in September 2021, approximately 91-terawatt hours of electricity is consumed annually to mine bitcoin which is more than the used by Finland where the population is about 5.5 million [52]. The bitcoin mining electricity consumption was higher than the electricity consumption by Ireland in 2017 [53]. Bitcoin uses more electricity annually than Argentina [54]. on the other hand, the Cambridge University researchers said the bitcoin uses approximately 121.6 terawatt-hours (TWh) which is more than the electricity usage by Argentina (121TWh), 108.8 TWh used by Netherland, and the United Arab Emirates uses 113.20 TWh annually [55] that represent in Fig. 6. China is the highest electricity consumer for mining bitcoin in the world which is 71.70% and the remaining balance is approximately 28.3% uses by the rest of the world whereas USA and Russia are second and third which is shown in Fig. 7 [56]. David Vetter published an article in Forbes where he predicted that by 2024

China alone will use more power for mining cryptocurrency than Italy uses electricity annually [57].

TABLE I. ELECTRICITY CONSUMPTION FOR MINING BITCOIN FROM JAN 2017 TO OCT 2021

Year	Electricity uses by Bitcoin mining (TWh) [47]	Electricity uses growth (%)	No of Bitcoin mining ('000) [48]	Per Bitcoin electricity consumption (KWh)	Avg. retail electricity cost/KWh in USA (USD) [49]	Per Bitcoin mining cost (USD)
2017	36.68		700	52,400	0.1048	5,492
2018	45.81	24.89	680	67,367	0.1053	7,094
2019	73.12	59.62	680	107,529	0.1054	11,334
2020	77.78	6.37	460	114,382	0.1059	12,113
2021	177.43	128.12	330	537,667	0.1118	60,111
Total	410.82		2,830			

TABLE II. ELECTRICITY CONSUMPTION FOR MINING ETHER FROM MAY 2017 TO DEC 2021

Year	Electricity uses by Ethereum (TWh) [50]	Electricity uses growth (%)	Approx. no of Ether mining ('000) [51]	Per Ether electricity consumption (KWh)	Avg. retail electricity cost/KWh in USA (USD) [49]	Per Ethereum mining cost (USD)
2017	9.56		9,000	1062	0.1048	111.30
2018	9.63	0.73	8,000	1204	0.1053	126.78
2019	8.14	15.47	10,000	814	0.1054	85.80
2020	14.64	79.85	15,000	976	0.1059	103.36
2021	103.42	606.42	17,500	5910	0.1118	660.34
Total	145.39		59,500			

TABLE III. MINING COST COMPARISON BETWEEN BITCOIN AND ETHER

Year	Per unit of Bitcoin mining cost (USD)	Ave. Bitcoin mining cost (USD)	Per unit of Ether mining cost (USD)	Ave. Ether mining cost (USD)
2017	5,492		111.30	
2018	7,094		126.78	
2019	11,334	19,229	85.80	217.52
2020	12,113		103.36	
2021	60,111		660.34	

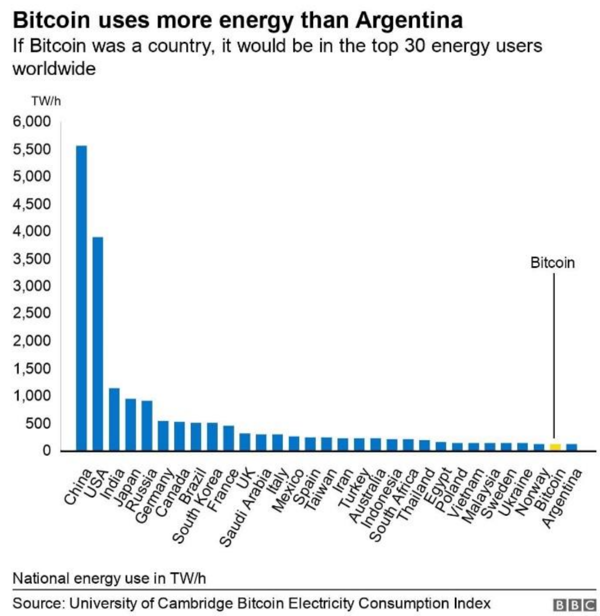


Fig. 6. Bitcoin Mining Power Consumption Globally.

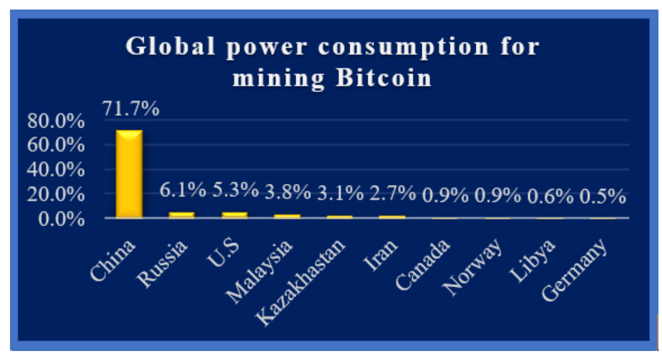


Fig. 7. Country-Wise Monthly Average Bitcoin Mining Computing Power.

However, the electricity consumption has been abruptly increased for mining bitcoin which is 177.43 TWh in 2021 as per Table I. Bitcoin mining needs 17MJ (megajoules) of power to generate an equivalent $1 amount of bitcoin whereas gold mining takes 5MJ to produce an equivalent $1 amount of gold [58]. On the other hand, there is a huge impact on the environment due to excessive use of electricity for mining bitcoin which releases more carbon dioxide (CO_2) into nature. Since the long increase of CO_2 in the environment is one of the big concerns in the world, cryptocurrency mining is increasing day by day which might be a severe disaster for the environment [59-60]. It is not only helping to increase CO_2 in nature, and it helps for rising temperature in the environment [61]. Table IV represents the fossil CO_2 emission from 2017 to 2022 and it shows that the CO_2 emission was 0.16% in the year 2017 and it sharply increased to 0.80% in 2021 which is alarming for the environment in the future.

On the other hand, CO_2 emission by Ether is approximately half of the bitcoin mining, which is shown in Table V, where the CO_2 emission is 0.46% in 2021. The combined impact of emission CO_2 has become noticeable for the environment due to generate bitcoin and ether where 0.99 Giga Tone (Gt) extra

CO_2 has been added in the nature shown in Table V. Bitcoin and Ether are not only the cryptocurrencies in the world, there are more than 10 thousand cryptocurrencies have been created by different companies though all of them are not active commercially, it will be extremely alarming for the nature in future if all the cryptocurrencies will go for commercial production.

TABLE IV. CARBON DIOXIDE EMISSION FOR MINING OF BITCOIN

Year	Fossil CO_2 emission in nature (Gt) [62]	Global electricity consumption (TWh)	Electricity consumption by bitcoin mining (TWh)	Fossil CO2 emission by bitcoin mining (Gt)	Fossil CO_2 emission by bitcoin mining (%)
2017	36.40	23,150	36.68	0.06	0.16
2018	34.80	22,153	45.81	0.07	0.21
2019	36.70	23,432	73.12	0.11	0.31
2020	36.60	23,176	77.78	0.12	0.34
2021	35.90	22,270	177.43	0.29	0.80
Total	180.40	114,181	410.82	0.65	

TABLE V. CARBON DIOXIDE EMISSION FOR MINING OF ETHER

Year	Fossil CO_2 emission in nature (Gt) [59]	Global electricity consumption (TWh)	Electricity consumption by Ether mining (TWh)	Fossil CO_2 emission by Ether mining (Gt)	Fossil CO_2 emission by Ether mining (%)	Fossil CO_2 emission by Bitcoin and Ether mining (Gt)
2017	36.40	23,150	9.56	0.02	0.04	0.08
2018	34.80	22,153	9.63	0.02	0.04	0.09
2019	36.70	23,432	8.14	0.01	0.03	0.12
2020	36.60	23,176	14.64	0.12	0.06	0.24
2021	35.90	22,270	103.42	0.17	0.46	0.46
Total	180.40	114,181	145.39	0.34		0.99

Fig. 8 represents the future prediction of electricity need for mining the bitcoin and Ether only which is drawn based on the bitcoin and Ether mining data from 2017 to 2021 where the electricity consumption will be increased by approximately more than 300 TWh in the year 2023 which can be higher than the annual consumption of many countries in the world.

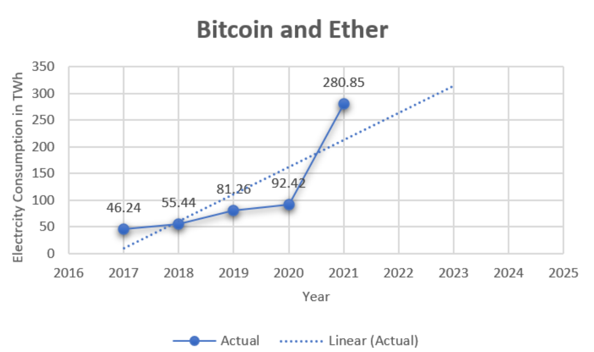


Fig. 8. Prediction of Electricity Consumption for Mining Bitcoin.

However, the proof-of-work algorithm is used to solve a complex mathematical problem for mining the bitcoin is the main concern for consumption of high energy [63-64]. It is an open public platform, and any participant can join the network and can be awarded by solving the mathematical puzzle. The PoW consensus algorithm is used to validate the transaction as well as generate the new blocks into the blockchain network [65-66]. The participants in the blockchain network send the transaction and all the transactions are gathered using the distributed ledger in the blocks through the validation process [67]. Though millions of participants compete to solve the puzzle, only a single participant will be declared a winner who can solve the cryptographic puzzle first, and this message will convey to all the users in the network [68]. So, the bitcoin mining power consumption is not so high concerning the winner [69]. But the participants those are not succeeded to solve the puzzle, their computers also use their power for the same which ultimately enormous amount of energy. The bitcoin mining process is shown in Fig. 9 where all the participants try to solve the complex mathematical puzzle to find out the nonce that can be matched with the target value [70]. To solve the mathematical puzzle, a large amount of computational power is required which depends on the issues are i) Hash function, ii) Integer factorization, and iii) Guided Tour Puzzle protocol.

- Hash Function: SHA-256 cryptographic hashing function is used for mining bitcoin where chunk amount data is used as input and bring it down the value in fixed format 256 bits [71]. Whatever the input, the out-hash value will be the same which fixes 256 bits or 64 bits hexadecimal value [72]. There is no simple way to get the hash value which depends on the lots of input data to needs to be used as the trial basis and it comes consumes huge computational power.
- Integer Factorial: It is used to secure the public key encryption process [73]. It is the way to represent the present whole number of the multiplication of two other numbers [74].
- Guided Tour Puzzle protocol: It works to protect the Denial of Service (DoS) attach to the blockchain network [75]. It also insists to focus on the nodes to compute the memory-bound puzzle which helps the users to use the abandoned computational power [76].

Presently, the blockchain network for bitcoin is growing faster, the users are facing more difficulties to solve the cryptographic puzzle and the algorithm needs more power to solve the same which tends to use more power for mining bitcoin.

Earlier, only a CPU was used for mining the bitcoin which is very slow and consumed more power to generate the bitcoin. To minimize the mining cost, special hardware called GPU is used for the same purpose. GPU is around 100 times faster than traditional CPU. Application Specific Integrated Circuit (ASIC) is another option that is also father than GPU, CPU, and FPGA which is shown in Fig. 10.

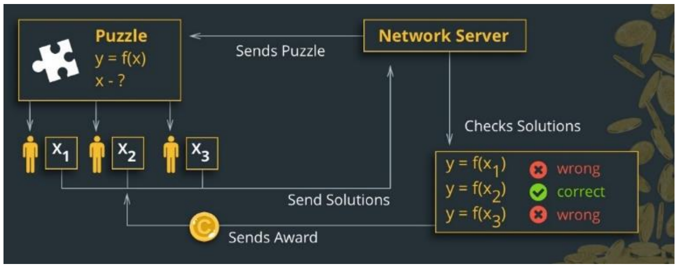


Fig. 9. Bitcoin Mining Process [77].

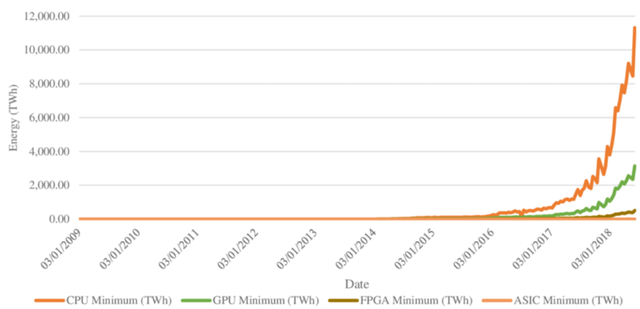


Fig. 10. Power Consumption by CPU, GPU, FPGA, ASIC [78].

IV. Energy Reducing Methods in Blockchain

Several layers consume the power for mining cryptocurrency, but the enormous power is used by the Proof-of-Work hashing algorithm. Here we have discussed the three methods A. Proof-of-Stake (Pos), B. Delegated Proof-of-Stake (DPoS), and C. Use of renewable energy. Method A and B will contribute to saving more energy for mining the cryptocurrency especially bitcoin which is used as an example. On the other hand, if the cryptocurrency network owners do not want to change the PoW algorithm, method iii) can be the alternative solution that is not capable to save energy for mining bitcoin, but it is environment friendly.

A. Proof-of-Stake (PoS)

PoS is a consensus algorithm that is the improved version of the PoW algorithm that may reduce the 99.5% power consumption for mining the cryptocurrency [79]. This algorithm does not depend on competition to generate a suitable hash among the users. The protocol determines the selector based on the ownership of the coin supply which replace the computational power and the transaction validation process will be done by this stakeholder. In Fig. 11, there is no miner is needed to validate the transaction. The participant's 33% stakeholder will be the validator of the transaction due to holding the maximum amount of stake in the network. The validators will not be rewarded for block validation and get the transaction fees only. Due to the single validator concept, the energy consumption is very low and there is no wastage of power. On the other hand, a successful miner for mining the bitcoin will be the gainer, the use of power by other miners should be considered a wastage of power. This wastage of power generates an excessive amount of carbon dioxide in the earth. If the PoS consensus algorithm was used instead of PoW for mining the bitcoin, a significant amount of power could be saved, and the CO_2 emission cloud is insignificant. As per Tables VI and VII, the carbon dioxide emission could be minimized by 179.75 Giga tones and 180.06 Giga tones for mining bitcoin and Ether respectively from the year 2017 to 2021 by using the PoS algorithm instead of the PoW algorithm. So, the PoS consensus algorithm is much more efficient and cost-effective for mining bitcoin and ether, and at the same time, it is environmentally friendly as well [80].

TABLE VI. Impact of PoS on Electricity and CO_2 to Mine Bitcoin

Year	**Electricity consumption for mining Bitcoin (TWh) using PoW**	**Electricity Requirement for mining Bitcoin (TWh) using PoS**	**Electricity saves (99.5%) by using PoS (TWh)**	**CO_2 emission in nature (Gt) using PoW**	**CO_2 emission for mining Bitcoin (Gt) using PoS**	**CO_2 emission saving by using PoS (Gt)**	**CO_2 emission saving (%) by using PoS**
2017	36.68	0.18	36.50	36.40	0.06	36.34	99.84
2018	45.81	0.23	45.58	34.80	0.07	34.73	99.80
2019	73.12	0.37	72.75	36.70	0.11	36.59	99.70
2020	77.78	0.39	77.39	36.60	0.12	36.48	99.67
2021	177.43	0.89	176.54	35.90	0.29	35.61	99.19
Total	410.82	2.05	408.77	180.4	0.65	179.75	99.64

TABLE VII. Impact of PoS on Electricity and CO_2 to Mine Ether

Year	**Electricity consumption for mining Ether (TWh) using PoW**	**Electricity Requirement for mining Ether (TWh) using PoS**	**Electricity saves (99.5%) by using PoS (TWh)**	**CO_2 emission in nature (Gt) using PoW**	**CO_2 emission for mining Ether (Gt) using PoS**	**CO_2 emission saves by using PoS (Gt)**	**CO_2 emission saving (%) by using PoS**
2017	9.56	0.05	9.51	36.40	0.02	36.37	99.95
2018	9.63	0.05	9.58	34.80	0.02	34.73	99.94
2019	8.14	0.04	8.10	36.70	0.01	36.62	99.97
2020	14.64	0.07	14.57	36.60	0.12	36.50	99.67
2021	103.42	0.52	102.90	35.90	0.17	35.73	99.53
Total	145.39	0.73	144.66	180.4	0.34	180.06	99.81

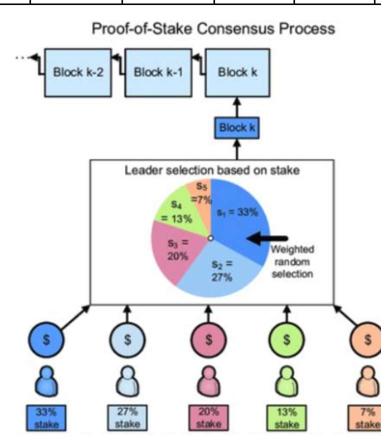


Fig. 11. Proof-of-Stake Consensus [80].

B. Delegated Proof-of-Stake (DPoS)

There is another protocol called Delegated Proof-of Work (DPoS) is the evolution of the PoS algorithm where the users are required to keep in stake their coins for getting the validation power for the transaction. The users of the blockchain network need to provide the votes and elect the delegate for validating the next available block. A maximum of 20 to 100 delegates can be chosen to validate the new block and the delegates for one block cannot be the delegate for the next block where the delegates will be eligible to receive the transaction fees from that validated block [81]. In terms of power consumption, DPoS is almost the same as the PoS algorithm which can be a good choice to use this protocol instead of the PoW algorithm for mining bitcoin or other cryptocurrencies. A comparison among PoW, PoS, and DPoS has been provided in Table VIII.

TABLE VIII. COMPARISON OF POW, POS, AND DPOS

PoW	PoS	DPoS
i. Number of participants becomes very high for mining cryptocurrency. ii. Energy consumption volume is very high. iii. Transaction speed is slower than PoS and DPoS. iv. Transaction validation method is very comparative for receiving the reward	i. Number of the participant is one and depends on the stake of the coin. ii. No extra power is required except the single computer power. iii. Transaction speed is very higher than PoW. iv. Competition for validation of transaction is not required due to a single selected validator.	i. Number of participants is limited to between 10 to 100. ii. Less energy consumption than PoW iii. Transaction speed is very higher than PoW and PoS iv. Competition for validation of transaction is not required due to selected validators.

C. Use of Renewable Energy

On the other hand, renewable energy may be the alternative option for mining cryptocurrency by using the same PoW algorithm. Though it will not save the electricity consumption rate, it will help to reduce the environmental pollution where the carbon dioxide emission will be zero. The largest bitcoin mining location in Dalian, China where the hash rate is 360,000 TH is used for mining 750 bitcoin every month and the average monthly cost is $1,170,000, and the second-highest mining location in Moscow, Russia where the hash rate is used 38PH for mining 600 bitcoins in every month and the monthly cost is $120,000 [82].

Renewable energy sources are preferable rather than non-renewable energy sources for mining cryptocurrency whereas IBM and Intel also prefer to use green energy for blockchain-based cryptocurrency transactions. Renewable power can be generated from different sources like solar, wind, water, etc. in different communities and the same can be distributed through blockchain applications by establishing a microgrid system. Power can be transmitted between peer-to-peer consumers into the blockchain network.

However, one of the main challenges to implementing blockchain technology is power consumption. So, the use of the PoS consensus algorithm can be a better option instead of using the PoW algorithm to reduce the power consumption whereas the use of renewable energy for the PoW algorithm may be the alternative option.

V. DISCUSSION AND RECOMMENDATION

The background and architecture of the blockchain technology, the energy consumption, and the impact of blockchain, energy reducing methods by applying blockchain applications have been discussed in detail in this review paper. To overcome the enormous energy consumption issues for mining cryptocurrencies and reduce the carbon dioxide emission in the environment, the following important recommendations can be addressed in the future.

- To minimize the consumption of computing power, the Proof-of-Stake (PoS) consensus algorithm is highly recommended to use instead of use Proof-of-Work (PoW) algorithm for mining cryptocurrencies.
- Electricity consumption can be minimized by approximately more than 99% by using the PoS consensus algorithm which will lead to saving the computation power cost of the end-users.
- A significant amount of CO_2 emission in the environment can be minimized by using the PoS algorithm instead of using PoW algorithm.
- Renewable energy sources are preferable rather than non-renewable energy sources for mining cryptocurrencies by using the PoW algorithm, though it will not help to reduce the computational power, it will help to minimize the CO_2 emission in nature.

VI. CONCLUSION

Blockchain already has become one of the leading-edge technologies that provide the highest level of data security through using a cryptographic hashing algorithm where data tempering is almost impossible. Nowadays, the maximum application of this technology uses for mining cryptocurrency where bitcoin is the leader in the market. But the consumption of computing power is the major challenge to adopting this technology for mining cryptocurrency and other sectors. Due to the consumption of enormous energy, carbon dioxide emission becomes very high which ultimately pollutes the environment. An alternative solution like Proof-of-Stake consensus protocols has been proposed to use instead of the Proof-of-Work algorithm for mining cryptocurrencies. Not only that, but the use of renewable energy can also be an alternate option to use the Proof-of-Work algorithm for mining cryptocurrencies which is environment friendly. Blockchain technology might be more useful in different sectors if the high energy consumption is addressed properly which will be able to build a secure blockchain network and save the transactional cost as well.

REFERENCES

[1] S. Nakamoto, “Bitcoin: (2008), A Peer-to-Peer Electronic Cash System”, 2008, Available: https://bitcoin.org/bitcoin.pdf, 2008
[2] B. A. Tama, B. J. Kweka, Y. Park, and K. H. Rhee, “A critical review of blockchain and its current applications”, In 2017 International

Conference on Electrical Engineering and Computer Science (ICECOS), pp. 109-113, 2017.

[3] A. H. Dyhrberg, S. Foley, and J. Svec, “How investible is Bitcoin? Analyzing the liquidity and transaction costs of Bitcoin markets”, Economics Letters, 171, pp.140-143, 2018.

[4] H. Watanabe, S. Fujimura, A. Nakadaira, Y. Miyazaki, A. Akutsu, and J. Juni , “Blockchain contract: A complete consensus using blockchian,” In 2015 IEEE 4th global conference on consumer electronics (GCCE), p. 577–8. doi:http://dx.doi.org/10.1109/GCCE.2015.7398721, 2015.

[5] A. Kosba, A. Miller, E. Shi, Z. Wen, and C. Papamanthou "Hawk: The Blockchain Model of Cryptography and Privacy-Preserving Smart Contracts", 2016 IEEE Symposium on Security and Privacy (SP), San Jose, CA, USA, pp. 839-858, doi: 10.1109/SP.2016.55, 2016.

[6] C. Vijai, S. M. Suriyalakshmi, and D. Joice., “The Blockchain Technology and Modern Ledgers Through Blockchain Accounting,” Adalya Journal, vol. 8(12), 2019.

[7] T. T. A. Dinh, R. Lui, M. Zhang, G. Cheng, B. C. Ooi, and J. Wang, “Untangling Blockchain: A data processing view of blockchain systems,” IEEE Transactions on Knowledge and Data Engineering, Vol. 30, Issue 7, pp. 1366–1385, DOI: 10.1109/TKDE.2017.2781227.

[8] M. Haferkorn and J. M. Q Diaz, “Seasonality and interconnectivity within cryptocurrencies-an analysis on the basis of bitcoin, litecoinand namecoin,” In International Workshop on Enterprise Applications and Services in the Finance Industry, Springer, Cham, pp. 106-120, 2014.

[9] S. McLean and S. Deane-Johns, “Demystifying Blockchain and distributed ledger technology–hype or hero,” Computer Law Review International, vol. 17(4), pp.97-102, 2016.

[10] RJ Yang, R. Wakefield, S. Lyu, S. Jayasuriya, F. Hang, X. Yi, T. Yang, G. Amarasinghe, and S. Chen, “Public and private blockchain in construction business process and information integration,” Automation in Construction, 118, pp.103276, 2020.

[11] M. R. King, “R3 Corda: A Distributed Ledger Technology for Financial Services,” Harvrad Business Publishing Education, [Available]: https://hbsp.harvard.edu/product/W18592-PDF-ENG, September 2018.

[12] M. S. Krstic, L.J. Krstic, “HYPERLEDGER FRAMEWORKS WITH A SPECIAL FOCUS ON HYPERLEDGER FABRIC,” https://doi.org/10.5937/vojtehg68-26206, DOI: 10.5937/vojtehg68-26206.

[13] J. Frankenfield, “Ethereum,” Investopedia, [Available]: https://www.investopedia.com/terms/e/ethereum.asp, 2022.

[14] A. Tandon, A. Dhir, A.K.M. N. Islam, M. Mantymaki, “Blockchain in healthcare: A systematic literature review, synthesizing framework and future research agenda,” Computers in Industry, https://doi.org/10.1016/j.compind.2020.103290.

[15] C.F. Calvillon, A. Sánchez-Miralles, and J. Villar, “Energy management and planning in smart cities,” Institute for Research in Technology (IIT) ICAI School of Engineering, Comillas Pontifical University, Santa Cruzde Marcenado26, Madrid, Spain, 55 (2016) 273-287.

[16] M. Joung and M, J. Kim, “Assessing demand response and smart metering impacts on long-term electricity market prices and system reliability,” Appl Energy 2013;101:441–8. http://dx.doi.org/10.1016/j.apenergy.2012.05.009.

[17] R.J. Green, “Electricity wholesale markets: designs now and in a low-carbon future,” Energy J 2008;0:95–124., https://ideas.repec.org/a/aen/journl/dn-sea06.html.

[18] C. Burger, A. Kuhlmann, P. Richard, and J. Weinmann, “Blockchain in the energy transition: A survey among decision-makers in the Germany energy industry,” (2016), [Available]:https://www.dena.de/fileadmin/dena/Dokumente/Meldungen/dena_ESMT_Studie_blockchain_englisch.pdf.

[19] S. Haber, and W. S. Stornetta, “How to Time-Stamp a Digital Document,” Advantages in Cryptography-CROPTO’ 90 pp 437-455, Conference on the Therory and Apllication of Cryptography.

[20] M. Davids, “Blockchain History: From Cryptology To The Wanted Technology,” https://nominex.io/blog/trading/blockchain-history-from-cryptology-to-the-most-wanted-technology/, 2020.

[21] P. D. DeVries, “An Analysis of Cryptocurrency, Bition, and the Future,” International Journal of Business Management and Commerce, Vol 1, No. 2, 2016.

[22] S. Jani, “An Overview of Ripple Technology & Its Comparision with Bitcoin Technology,” [Available]: https://www.researchgate.net/publication/322436263_An_Overview_of_Ripple_Technology_its_Comparison_with_Bitcoin_Technology, 2018.

[23] G. Mcfarlane, “What is Litecoin?,” Investopedia, [Avaiable]: https://www.investopedia.com/articles/investing/040515/what-litecoin-and-how-does-it-work.asp, 2021.

[24] R. de Best, “Bitcoin (BTC, BTH) energy consumption relative to selected countries worldwide 2021,” [Available] : https://www.statista.com/statistics/881522/bitcoin-energy-consumption-relative-to-select-countries/, 2021.

[25] M. Haferkorn and J. M. Q Diaz, “Seasonality and interconnectivity within cryptocurrencies-an analysis on the basis of bitcoin, Litecoin and Namecoin,” In International Workshop on Enterprise Applications and Services in the Finance Industry, Springer, Cham, pp. 106-120, 2014.

[26] L. C. Schaupp and M. Festa, “Cryptocurrency adoption and the road to regulation,” In Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age, pp. 1-9, 2018.

[27] H. F. Atlam, A. Alenezi, M. O. Alassafi, and G. Wills, “Blockchain with internet of things: Benefits, challenges, and future directions,” International Journal of Intelligent Systems and Applications, vol. 10(6), pp. 40-48, 2018.

[28] A. Dorri, S. S. Kanhere, and R. Jurdak, “Towards an optimized blockchain for iot,” in Proceedings of the Second International Conference on Internet-of-Things Design and Implementation, ser. IoTDI ’17. New York, NY, USA: ACM, 2017, pp. 173–178. [Available]:http://doi.acm.org/10.1145/3054977.3055003.

[29] H. Kopp, C. Bosch, and F. Kargl,“Koppercoin – a distributed file storage with financial incentives,” 12th International Conference on Information Security Practice and Experience, Zhangjiajie, China, pp. 79–93, 2016.

[30] T. Wu and X. Liang, “Exploration and Practice of Inter-Bank Application Based on Blockchain,” ICCSE 2017 - 12th Int. Conf. Comput. Sci. Educ., no. Iccse, pp. 219–224, 2017.

[31] P. Daian, I. Eyal, A. Juels, and E. G. Sirer, “(Short Paper) PieceWork: Generalized Outsourcing Control for Proofs ofWork,” Financial Cryptography and Data Security Lecture Notes in Computer Science, pp. 182–190, 2017.

[32] I. Bentov, C. Lee, A. Mizrahi, and M. Rosenfeld, “Proof of activity: Extending bitcoin’s proof of work via proof of stake (extended abstract),” ACM SIGMETRICS Performance Evaluation Review, vol. 42, no. 3, pp. 34–37, 2014.

[33] R. L. Rivest, A. Shamir, and L. Adleman, “A method for obtaining digital signatures and public-key cryptosystems,” Commun. ACM, vol. 21, no. 2, pp. 120–126, [Available]: http://doi.acm.org/10.1145/359340.359342, 1978.

[34] A. Dorri, S. S. Kanhere, R. Jurdak, and P. Gauravaram, “Lsb: A lightweight scalable blockchain for iot security and privacy,” arXiv preprint rXiv:1712.02969, 2017.

[35] A. Kiayias, A. Russell, B. David, and R. Oliynykov, “Ouroboros: A Provably Secure Proof-of-Stake Blockchain Protocol,” Advances in Cryptology – CRYPTO 2017 Lecture Notes in Computer Science, pp. 357–388, 2017.

[36] M. R. Islam, M. M. Rashid, M. Mahmud, M. A. Rahman, M. H. S. B. Mohamad, and A. H. B. Embong, “A Review of Blockchain Security Issues and Challenges”, 2021 IEEE 12th Control and System Graduate Research Colloquium (ICSGRC), 2021, pp. (227-232), DOI: 10.1109/ICSGRC53186.2021.9515276, 2021.

[37] G. Zyskind, O. Nathan, and A. S. Pentland, “Decentralizing privacy: Using blockchain to protect personal data,” in Security and Privacy Workshops (SPW), 2015 IEEE. IEEE, pp. 180–184, 2015.

[38] M. Li, L. Zhu, and X. Lin, “Efficient and Privacy-preserving Carpooling using Blockchain-assisted Vehicular Fog Computing,” IEEE Internet of Things Journal, pp. 1–1, 2018.

[39] K. Jonathan and A. K. Sari, “Security Issues and Vulnerabilities On A Blockchain System: A Review,” In 2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI) , pp. 228-232, 2019.

[40] S. McLean and S. Deane-Johns, “Demystifying Blockchain and distributed ledger technology–hype or hero,” Computer Law Review International, vol. 17(4), pp.97-102, 2016.

[41] D. Mingxiao, M. Xiaofeng, Z. Zhe, W. Xiangwei, and C. Qijun, “A Review on Consensus Algorithm of Blockchain,” 2017 IEEE Int. Conf. Syst. Man, Cybern., pp. 2567–2572, 2017.

[42] V. Dostov and P. Shust, “Cryptocurrencies: an Unconventional Challenge to the AML/CFT Regulators?” J. Financ. Crime, vol. 21, no. 3, pp. 249–263, 2014.

[43] E. L. Greebel, K. Moriarty, C. Callaway, and G. Xethalis, “Recent key Bitcoin and Virtual Currency Regulatory and Law Enforcement Developments,” J. Invest. Compliance, vol. 16, no. 1, pp. 13–18, 2015.

[44] R. Khalid, N. Javaid, S. Javaid, M. Imran, and N. Naseer. “A Blockchain-Based Decentralized Energy Management in a P2P Trading System,” 2020 IEEE International Conference on Communications (ICC), Dublin, Ireland, pp. 1-6.

[45] R. de Best, “Number of cryptocurrencies worldwide from 2013 to February 2022,” [Available]: https://www.statista.com/statistics/863917/number-crypto-coins-tokens/, 2022.

[46] E. Kim, “Bitcoin mining consumes 0.5% of all electricity used globally and 7 times Google’s total usage,” [Available]: https://www.businessinsider.com/bitcoin-mining-electricity-usage-more-than-google-2021-9, 2021.

[47] R. de Best, “Global Bitcoin (BTC, BTH) energy consumption up until October 19, 2021”, Available: https://www.statista.com/statistics/881472/worldwide-bitcoin-energy-consumption/, 2021.

[48] J. Frankenfield, “Bitcoin Mining, Investopedia,” [Available]: https://www.investopedia.com/terms/b/bitcoin-mining.asp, 2022.

[49] B. Alves, “Average retail electricity prices in the United States from 1990 to 2021 (in U.S. cents per kilowatt hour)”, Available: https://www.statista.com/statistics/183700/us-average-retail-electricity-price-since-1990/, 2022.

[50] R. de Best, “Global Ethereum (ETH) energy consumption up until January 10, Available: https://www.statista.com/statistics/1265897/worldwide-ethereum-energy-consumption/, 2022.

[51] N. Maddrey, “Coin Metrics' State of the Network: Issue 63”, Available: https://coinmetrics.substack.com/p/coin-metrics-state-of-the-network-65f?s=r, 2020.

[52] R. de Best, “Number of Bitcoins in circulation in worldwide from October 2009 to January 2022”, [Available]: https://www.statista.com/statistics/247280/number-of-bitcoins-in-circulation/, 2022.

[53] eia, “Electricity Power Monthly”, https://www.eia.gov/electricity/monthly/epm_table_grapher.php?t=epmt_5_03.

[54] J. Huang, C. O’Neill, and H. Tabuchi, “Bitcoin Usages More Electricity Than Many Countries. How is that Possible?,” The New York Times, 2021, https://www.nytimes.com/interactive/2021/09/03/climate/bitcoin-carbon-footprint-electricity.html, 2021.

[55] A. Hern, “Bitcoin mining consumes more electricity a year than Ireland,” https://www.theguardian.com/technology/2017/nov/27/bitcoin-mining- consumes- electricity-ireland, 2017.

[56] M. Amar, “Bitcoin (BTC) consumes more electricity than the entire country of Argentina. And WHY,” [Avaiable]: https://www.linkedin.com/pulse/price-bitcoin-increases-so-does-currencys-energy-consumption-amar, 2021.

[57] C. Criddle, “Bitcoins Consumes More Electricity than Argentina,” [Avaiavle]: https://www.bbc.com/news/technology-56012952, 2021.

[58] C. Yurong, “What makes China a world cryptocurrency mining hub?,” [Available]: https://news.cgtn.com/news/2021-03-29/What-makes-China-a-world-cryptocurrency-mining-hub--Z1CL94J3Ik/index.html, 2021.

[59] D. Vetter, “Forbes: Bitcoin Could Churn Out 130 Million Tons Of Carbon, Undermining Climate Action. Here’s One Way To Tackle That,” https://www.forbes.com/sites/davidrvetter/2021/04/06/bitcoin-could-churn-out-130-million-tons-of-carbon-undermining-climate-action-heres-one-way-to-tackle-that/?sh=1aa370b65a8f, 2021.

[60] L. Ante, “Smart Contracts on the Blockchain – A Bibliometric Analysis and Review,” Telemat, Informatics 57, 101519. [Available]: https://doi.org/10.1016/j.tele.2020.101519, 2020.

[61] L. Ante, F. Steinmetz, and I. Fiedler, “Blockchain and energy: A bibliometric analysis and review”, Renew. Sustain. Energy Rev. 137, 110597. https://doi.org/10.1016/j.rser.2020.110597, 2021.

[62] CO2.com, “Global Carbon Emissions,” [Available]: https://www.co2.earth/global-co2-emissions.

[63] U. Gallersdörfer, L. Klaaßen, and C. Stoll, C., “Energy Consumption of Cryptocurrencies Beyond Bitcoin”, Joule 4, 1843–1846. https://doi.org/10.1016/j.joule.2020.07.013, 2020.

[64] iea.org, “Global Energy Review 2021 (Electricity),” [Available]: https://www.iea.org/reports/global-energy-review-2021/electricity.

[65] Enerdata, “Electricity domestic consumption,” [Available]: https://yearbook.enerdata.net/electricity/electricity-domestic-consumption-data.html.

[66] Digiconomist, “Bitcoin Energy Consumption Index,” 2021, [Avaiable]: https://digiconomist.net/bitcoin-energy-consumption/ (accessed 3.5.21).

[67] Z.B. Zheng, S. A. Xie, H. N. Dai, X. P. Chen, and H. M. Wang, H.M. “An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends” , In Proceedings of the 2017 IEEE International Congress on Big Data (Big Data Congress), Honolulu, HI, USA, 25–30 June 2017; pp. 557–564. 2017.

[68] IBM Inegration Bus, “Public key encryption”, [Avaiable]: https://www.n-able.com/blog/sha-256-encryption, 2021.

[69] R. Duffy, “Proof-of-Stake could make Ethereum 99.95% more energy-efficient, How does it works?”, [Available]: https://www.morningbrew.com/emerging-tech/stories/2021/05/19/proofofstake-make-ethereum-9995-energyefficient-work], 2021.

[70] T. Aura, P. Nikander, and J. Leiwo, “DoS-resistant authentication with client puzzle,” in 8th International Workshop on Security Protocols, vol. 2133, 2000, pp. 170-181

[71] Alexanderia, “What is SHA-256,” [Available]: https://coinmarketcap.com/alexandria/glossary/sha-256.

[72] N-Able, “SHA-256 Algorith Overview”, [Avaiable]: https://www.n-able.com/blog/sha-256-encryption, 2019.

[73] S. Albrecht, S. Reichert, J. Schmid, J. Strüker, D. Neumann, and D. Fridgen, “Dynamics of Blockchain Implementation - A Case Study from the Energy Sector,” Proceedings of the 51st Hawaii International Conference on System Sciences, 2018.

[74] O. Avan-Nomayo, “Bitcoin miner beats 1 in 1.3 million odds to mine a BTC block”, [Available]: https://www.theblockcrypto.com/post/129894/bitcoin-miner-beats-1-in-1-3-million-odds-to-mine-a-btc-block, 2022.

[75] J. Sedlmeir, H. U. Buhl, G. Fridgen, and R. Keller, “The Energy Consumption of Blockchain Technology: Beyond Myth. Bus,” Inf. Syst. Eng. 62, 599–608. https://doi.org/10.1007/s12599-020-00656-x, 2020.

[76] A. Tar, “Proof-of_Work, Explaned”, [Available]: https://cointelegraph.com/explained/proof-of-work-explained, 2018.

[77] S. Kufeoglu, M. A Ozkuran, “Energy Consumption of Bitcoin Mining,” Available: https://www.researchgate.net/publication/337886683, 2019.

[78] Lisk Academy, “Proof of Stake. Lisk”, [Available]: https://lisk.io/academy/blockchain-basics/how-does-blockchain-work/proof-of- stake, 2018.

[79] C. T. Nguyen, D. T. Hong, D. N. Nguyen, D. Niyato, H. T. Nguyen, and E. Dutkiewicz, “Proof-of-Stake Consensus Mechanisms for Future Blockchain Networks: Fundamentals, Applications and Opportunities”, DOI: 10.1109/ACCESS.2019.2925010, IEEE Access.

[80] Gemini, “What are Proof of Stake and Delegated Proof of Stake”, [Available]: https://www.gemini.com/cryptopedia/proof-of-stake-delegated-pos-dpos, 2021.

[81] Sunbird, “Largest Bitcoin Mining Farms in the World,” [Available]: https://www.sunbirddcim.com/infographic/largest-bitcoin-mining-farms-world.

[82] H. Zhao, “Bitcoin and Blockchain consume an exorbitant amount of energy,” [Available]: https://www.cnbc.com/2018/02/23/bitcoin-blockchain- consumes-a-lot-of-energy-engineers-changing-that.html, 2018.

Bridge Pillar Defect Detection using Close Range Thermography Imagery

Abd Wahid Rasib[1]*, Muhammad Latifi Mohd Yaacob[2], Nurul Hawani Idris[3]
Programme Geoinformation, Faculty of Built Environment and Survey
University Teknologi Malaysia
Johor, Malaysia

Khairulazhar Zainuddin[4]
Centre of Studies for Surveying Science and Geomatics
Universiti Teknologi MARA
Perlis, Malaysia

Rozilawati Dollah[5]
School of Computing, Faculty of Engineering
Universiti Teknologi Malaysia
Johor, Malaysia

Norbazlan Mohd Yusof[6], Norisam Abd Rahaman[7]
Centre of Excellence, PLUS Berhad
Selangor, Malaysia

Shahrin Ahmad[8], Norhadi A. Hamid[9], Abdul Manaf Mhapo[10]
Geolatitude Technology Sdn Bhd
Johor, Malaysia

***Abstract*—Currently, radiometric thermography image has been explored adequately as alternative advance Non-Destructive Testing (NDT) especially for early detection analysis in various applications. Systematic image calibration, higher spatial resolution and high degree order image processing, thermography imagery potential to be used in concrete structure defect detection. Therefore, this study is carried out to examine the defect on bridge pillar surface concrete using drone-based thermography sensor (7–13 µm). Close range remote sensing NDT based on drone platform and imagery segmentation analysis have been applied to interpret the crack line on two pillars at North-South Expressway Central Link (ELITE) Highway. As a result, thermography imagery segmentation and support by multispectral radiometric imagery (RGB) successfully to delineate the micro crack line on the bridge pillar concrete using K-means clustering method. Overall, this study successfully shows the higher order optional platform using drone and thermography sensor that potentially to be applied in forensic concrete structure defect detection for tall structure building.**

***Keywords*—Defect; detection; bridge pillar; drone; thermography; close range remote sensing**

I. Introduction

Thermography that uses a mid-infrared band is the part of remote sensing that manages the acquisition, processing and interpretation of information procured essentially in the thermal infrared (TIR) area of the electromagnetic (EM) range. In thermography remote sensing, the radiations emitted from the target surface will be measured which differ from optical remote sensing where the reflected radiation from target surface is the one that will be measured and these measures will determine a body's radiant temperature that were determined by two variables, kinetic temperature and emissivity [1]. Current technology has allowed the concept of thermography remote sensing being adapted by the use of drone or Unmanned Aerial Vehicle (UAV) as many compact thermal sensors were available. Nowadays. the thermal sensors are extensively utilized in different surface temperature and thermal emission measurements. The old-style issue on kinetic temperature and emissivity assurance through the force and its circulation over the wavelength region of close-range remote sensing can be explained not similar concept as the airborne or spaceborne thermal sensors. Concerning low altitude or drone-based sensors, the atmospheric impacts are insignificant, lab-level adjustment are progressively available, and theoretically the measurement of temperature are more accurate [2]. Close-range remote sensing imagery technology detection are believed to be a much more practical and more relevance approach to the work of detecting a surface that are not accessible by the remote sensing sensor.

Concrete structure surface defect come in quite few types with cracking and scaling being one of the most common defects. In order to detect defect of a certain structure or infrastructure, the methods that are considered as the most suitable are usually a Non-Destructive Testing (NDT) method because this method allows to avoid interfering with the asset's current state. The majority of NDT for structural elements, such as ultrasonic, magnetic field, and eddy current methods, are best suited for detecting flaws at depths of 5 to 100 cm and have two major drawbacks: they require physical contact with the tested object and scan images slowly [3]. Other method that would be very useful to the concrete structure defect detection is a LiDAR measurement method. LiDAR is a method for determining ranges by targeting an object with a laser and measuring the time for the reflected light that return to the receiver. As a result, LiDAR provides point cloud data, which is basically 3D geometric information that may be used to measure surface anomalies, by populating a surface with hundreds of laser points [4]. As it was mention

*Corresponding Author

earlier, this study will incorporate a thermography imaging method which is also consider as part of NDT method. Thermography, commonly known as thermal scanning or infrared imaging, is a rapid and remote technology that is increasingly being utilized in conjunction with NDT methods [5,3]. Usually, defects produce an excessive heat during operation due to the excess current flow at the fault point results in power dissipation [6].

Nowadays the concept of infrared thermography (IRT) imagery was being utilized in many ways. Firefighters use them to see through smoke, find people and localize hotspots of fires. Law enforcement uses the technology to manage surveillance activities, locate and apprehend suspects, investigate crime scenes and conduct search and rescue operations. Power line maintenance technicians locate overheating joints and parts eliminate potential failures. Where thermal insulation becomes faulty, building construction technicians can see heat leaks to improve the efficiencies of cooling or heating. Physiological activities, such as fever, in human beings and other warm-blooded animals can also be monitored with thermographic imaging. IRT imagery have great thermal resolution capabilities, but they have a basic spatial resolution constraint [6]. This indicating that thermography imaging concept has proven its usefulness and other ways that this concept can be used is for the detection of defect on a concrete surface such as bridge pillar.

Surface temperature data can be collected using IRT method on which its data can detect the internal faults in concrete [7]. The work defect detection on a concrete surface can be carried out done by differentiating the temperature on its surface as thermography create an illusion of differences temperature for human to detect the difference in thermography imagery or in this case a defect on a concrete surface. The concept of thermography imagery in detecting the defect structure is when the thermographic camera detects the radiation emitted by an object as it was heated by sun or any sourced of heat. The thermography NDT image is mainly judged by analyzing the abnormal conditions of the temperature change law at the crack manifested that when the ambient temperature is lower than the object temperature, the cracks are shown as hot spots or regions in the thermal camera. While cracks would manifest themselves as cold spots when the region of interest is colder than the surrounding environment [8].

This project utilized a drone-based platform in which the thermal image acquired vertically to reconstruct a complete pillar structure. A thermal sensor mounted to the drone to capture the thermography image of the pillar surface. The acquired imagery then underwent in a vertical flight plan as in order to completely cover all the area of the pillar. Therefore, when this concept was applied on the maintenance work of the bridge pillar, a thermal sensor will be attached on a drone and thermography image of the pillar surface will be collected. From those images, the temperature of the pillar surface will be identified and the differences in temperature will produced an assessment that indicate there is something thermal hot spot with the concrete surface of the pillar hence, a defect is detected. This defect might result as the presence of cracked on the surface of the concrete which needed to be inspected and immediate action can be performed to avoid any unwanted event.

II. Study Area

The research areas that explored are located at part of North-South Expressway Central Link (ELITE) Highways' bridge pillar, which is also known as Asian Highway 2 (AH2), situated near to Management and Science University (MSU) Shah Alam, Selangor, Malaysia (03°04'41.3"N,101°33'06"E) as shown in Fig. 1. This study focused on the crack detection of two pillars or piers of the highway bridge. The data acquisition that has been collected then processed and analyzed using the specific software to detect crack on bridge surface.

(a)

(b)

Fig. 1. The Study Area Located at Part of ELITE Highways Bridge near to MSU Shah Alam, Selangor. (a). Location Map (Source: Google Earth), (b). Concrete Pillar.

III. Aim and Objective

The aim of this research is to study the performance of thermal imaging technique for detecting the defect of vertical concrete surface. In pursuit of the aim of this research, this study specifically addresses two objectives as follows:

- To demonstrate the vertical concrete surface defect detection technique using drone-based close-range thermography imagery.
- To detect the defect on bridge surface using segmentation method by utilizing thermal group of UAV remote sensing.

IV. Methodology

A. Close Range Thermography Imagery

Close range photogrammetry also known as terrestrial photogrammetry is a procedure in which photos are collected with relatively high convergent camera orientations around an entity, often facing towards the entity's center [9]. The concept of it refers to the imagery collected in a range less than 300 meters away from a study object. In this project, the close-range drone approach was utilized where the thermography imagery was acquired using a sensor carried by DJI Inspire 1. The drone is a commercial off-the-shelf (COTS) photography platform that is ready to maneuver right out the box. The drone is featured with an onboard color camera equipped with a lens and a stabilized gimbal, also a retractable landing gear that can be pulled up out of the view, giving the camera 360 degrees of view towards the world below without obstruction.

The reason to acquire and process the thermography imagery photogrammetrically was to produce thermal orthomosaic with ground resolution distance (GSD) below 10mm. Meanwhile, thermography lacks the image's features. The recorded details are based on temperature, limiting the success of image matching among overlapped photographs for orientation in photogrammetric software. For that reason, the normal color image was also acquired to assist the photogrammetric image orientation process.

The DJI FC350 camera and the FLIR Zenmuse XT (Table I) were used to capture the color and thermographic images, respectively. The former is an onboard camera for the DJI Inspire 1 drone system. It features a 12.4-megapixel effective pixel and a 20-millimeter lens. Additionally, the camera is detachable from the drone, allowing for the attachment of the former camera for thermographic imaging. The FLIR Zenmuse XT is designed to be mounted on the DJI Inspire 1, capturing images between 7.5 and 13.5 meters and storing them in 640 x 512 pixels image.

In a typical photogrammetric mapping project, the flight mission configures autonomous data collection in the open space area prior to the flight mission. The benefits of an autonomous flight mission include the drone maneuvering autonomously as a result of the GPS satellite signal assisting with flight direction and hovering. Additionally, the image can be captured automatically based on the coordinates of the configured camera station. The coordinates are then geotagged to the corresponding captured photo station to facilitate the point cloud construction process in photogrammetric software, as the exterior orientation parameter is available.

However, in this project, the drone was piloted manually because the data acquisition took place the beneath of bridge, where a tick concrete structure interfered with the GPS signal. Manually controlling the drone without locking the GPS signal was difficult in areas where the drone was uncontrollable due to turbulence. Additionally, the image was captured manually, as it was not necessary to preconfigure the flight planning. The image captured from a photogrammetric image block is made up of individual strips. The overlap of images was set to 80%. The camera-to-object distance was set to 1.5m in order to obtain thermography images with a 10mm GSD.

B. Establishment of Control Point

Photogrammetric processing enables the estimation of an object's scale and orientation based on the image's exterior orientation parameter. However, because the work is being conducted in a tick structure environment, accurate geotagging of GPS signals on the image is compromised. As a result, the image's geotagged parameter did not accurately reflect the image's actual location on the ground.

Alternatively, the use of GCP can be used to resolve the determination of the constructed object's coordinates and scale. Additionally, GCPs are used to determine the exterior orientation parameter of an image by resecting the coordinate to the image point via a mathematical equation. With a precise exterior orientation for the image, light rays are projected to precisely and accurately model space from images. As a result, the 3D model of the object constructed has an accurate scale and orientation similar to that of the ground.

On the other hand, the thermal image's nature is to capture the object's heat/temperature. Thus, the thermal image does not contain a common GCP marker used in photogrammetry. The use of a thermal patch may be appropriate. However, the patch's coordinates cannot be precisely determined. For this project, the standard photogrammetric control point marker was designed based on the focal lens, sensor resolution, and camera-to-object distance of the respective camera. Ten markers were printed on A4-size paper and adhered to the surface of each pillar.

TABLE I. Specification of the Instrument

Instrument	Specification
 DJI FC350	Total Pixels: 12.76M Effective Pixels: 12.4M Image Max Size: 4000x3000 ISO Range: 100-3200 (video) 100-1600 (photo) Electronic Shutter Speed: 8s – 1/8000s FOV: 94° Lens: 20mm
 FLIR Zenmuse XT	Spectral Band: 7.5 – 13.5 µm Thermal Resolution: 640 × 512 pixels Full frame rate: 30 Hz Field of view: 9mm lens (69°×56°)
 DJI Inspire 1	Flight autonomous: 20 minutes Payload: 1.7 kg

In this study, the location of markers on two pillars was set to the upper area as the focus point for thermography analysis and shown in Table II. Each marker's position was determined using the tacheometry technique using an arbitrary coordinate system consisting of a single baseline. The geometric control point coordinates of each pillar are shown in Table III and Table IV, respectively. The information from the GCPs was then used to create the thermography orthomosaic, as explained in the following section.

TABLE II. LOCATION OF CONTROL TARGETS

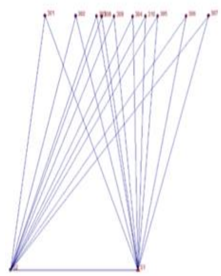
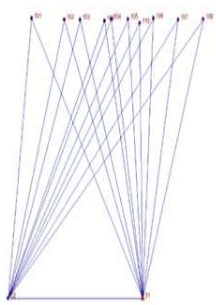
Pillar No	Location of target	Network of target
34A		
34B		

TABLE III. GEOMETRIC CONTROL COORDINATE FOR PILLAR 34A

Station	X	Y	Z
S1	20.000	20.000	10.000
S2	10.429	20.000	10.336
301	13.020	32.802	18.052
302	15.327	32.790	18.057
303	16.894	32.782	17.858
304	19.610	32.757	17.119
305	21.482	32.777	17.187
306	23.620	32.783	17.533
307	25.308	32.800	17.601
308	17.286	32.759	16.399
309	18.203	32.763	15.956
310	20.566	32.767	15.966

TABLE IV. GEOMETRIC CONTROL COORDINATE FOR PILLAR 34B

Station	X	Y	Z
S1	20.000	20.000	10.000
S2	10.315	20.000	9.860
101	12.008	32.790	17.529
102	14.361	32.718	17.697
103	15.499	32.718	17.639
104	17.740	32.749	16.925
105	18.949	32.754	16.938
106	20.770	32.789	16.986
107	22.548	32.730	17.083
108	24.363	32.746	17.333
109	17.258	32.678	15.218
110	19.784	32.560	15.114

C. Orthophoto Construction

After constructing the tie point cloud, a dense point cloud was generated. The distinction between a tie and a dense point cloud is that the former is used to determine the exterior orientation parameter by resecting the tie point. Meanwhile, the latter is used to transform each pixel in the image into three-dimensional space. From the dense point cloud, a variety of photogrammetric products such as a 3D model and orthomosaic can be generated. The thermographic orthomosaic in this work was created using the Pix4D mapper (academic version).

The process began with aligning the color images in Pix4D in order to create a tie point cloud and estimate the exterior orientation parameter for each image. The GCPs are then used to reorient the image via bundle adjustment and to scale the pillar's 3D point cloud for metric measurement. Following that, a dense 3D point cloud was generated and the coordinates of pillar features such as edges and contrast spots were extracted in order to align the thermography imagery. Due to high resolution of color image, the common features on overlapped images were seamlessly matched for tie point construction.

The thermal dataset was then sequentially inserted onto color photographs that had previously been aligned. The edges and contrast spots were marked as control points for the thermal point cloud. The coordinates for the control point were extracted from the object coordinates previously determined using the color image. This procedure enables the transformation of the constructed pillar's coordinates to the features visible on the thermographic image. When the thermal imagery was successfully aligned, the thermal image's dense point cloud was also generated. Following that, the thermal orthomosaic was created using the orthoplane function included with Pix4D Mapper. Fig. 2 to Fig. 5 has shown the process using Pix4D Mapper.

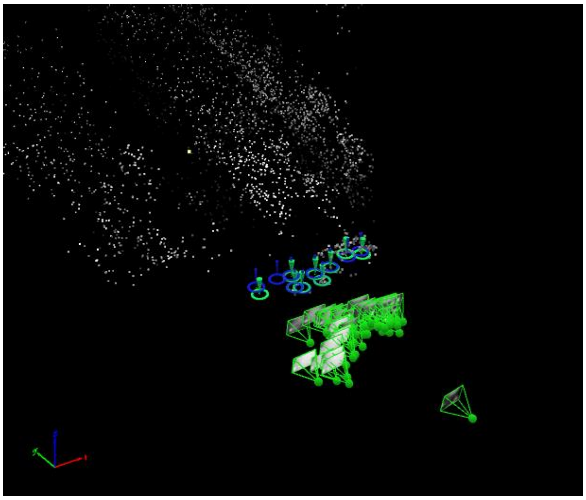


Fig. 2. Aligned Photographs in Pix4D Mapper.

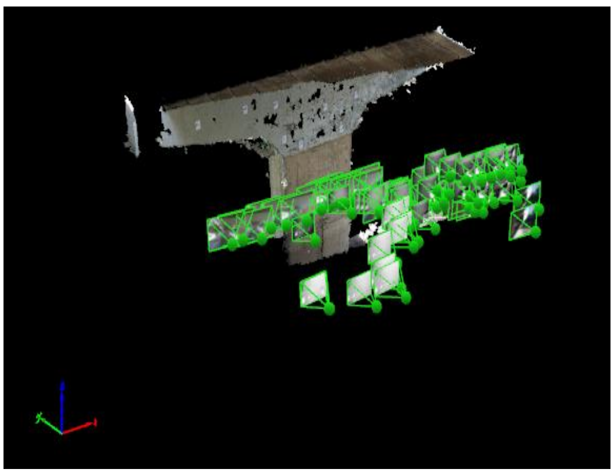


Fig. 3. Dense Point Cloud.

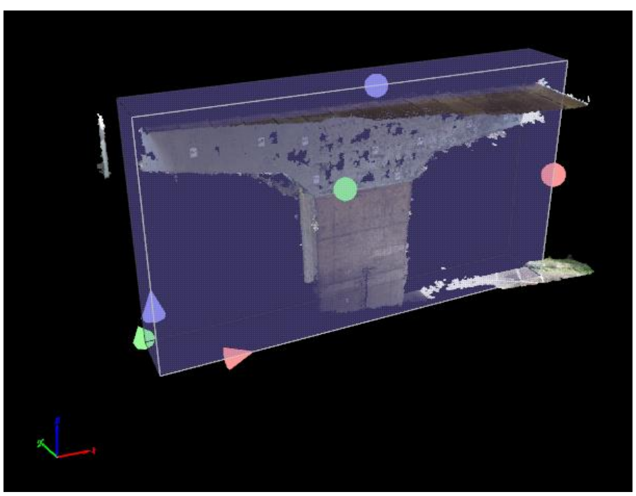
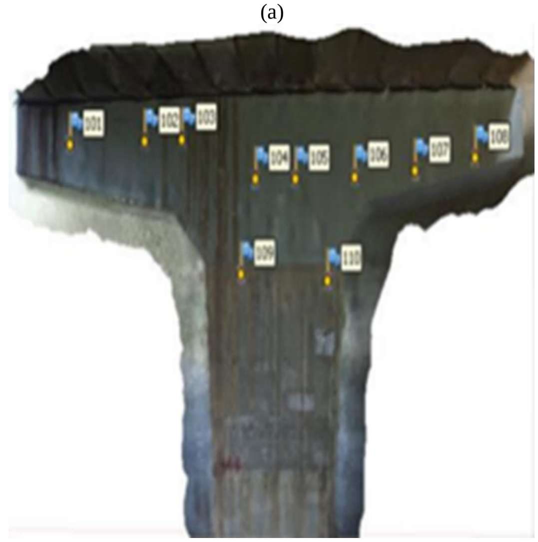


Fig. 4. Thermal Orthoplane for Mosaic Extraction Process.

(a)



(b)

Fig. 5. RGB Image Stitching. (a) 34A (b) 34B.

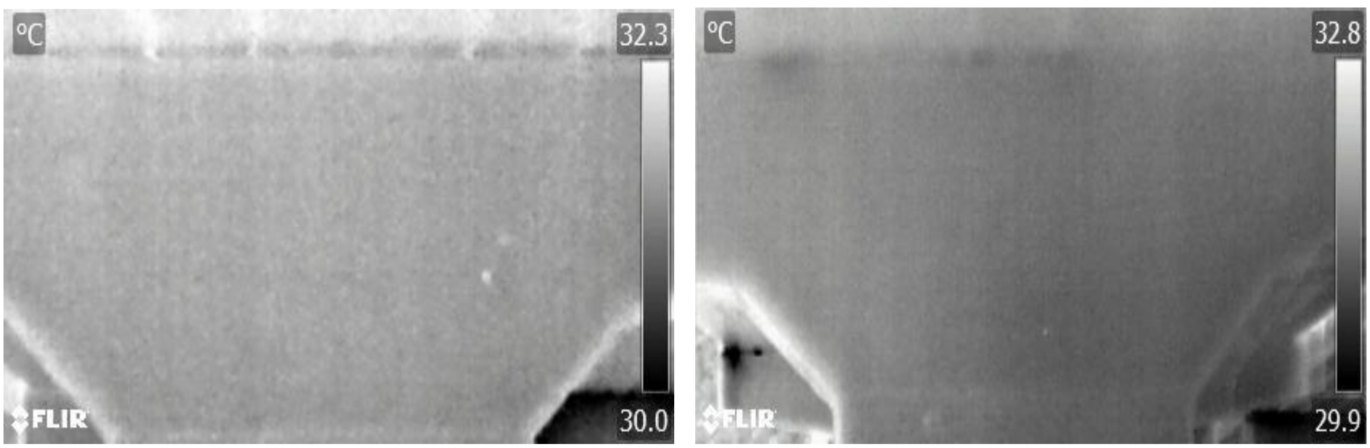
FLIR Tools+'s image stitching for thermal images is fairly simple, beginning with importing the images into the FLIR Tools+ workspace. FLIR Tool+ stitched together several smaller images into a larger one using panorama mode by analyzing each image to detect pixel patterns that match pixel patterns in other images. Fig. 6 shows the results of image stitching from a thermal image using the FLIR Tool+ software.

D. Image Enhancement and Segmentation

According to [10], there are many researchers has used the image processing technique to detect a crack or defect in a certain structure, for example [11] has proposed preprocessing, image segmentation and feature extraction, [12] have presented a detection method based on the use of acoustic emission and digital image correlation and [13] used Gabor filtering to presented a method for automatically detecting cracks in digital photographs. In this project the method used was the image enhancement and image segmentation.

(a) (b)

Fig. 6. Thermal Image Stitching. (a) 34A (b) 34B.



(a) (b)

Fig. 7. Thermal Image Enhancement Result. (a) 34A (b) 34B.

For Image enhancement is the process of adjusting digital images so that the results are more suitable for display or further image analysis. Furthermore, Image enhancement is used to improve the perception of visual quality in an image as well as to provide better input for other computerised image processing processes [14]. This process allowed the user to remove noise, sharpen, or brighten an image thus, making it easier to identify key features on the image and for this case, is to detect the defect area. Using FLIR Tool+ software, the image enhancement process can simply be achieved by adjusting the image contrast between temperature. Fig. 7 below shows the result of image enhancement.

Image segmentation is the process of dividing an image into several segments in order to transform the image's interpretation into something more meaningful and easier to work with or being analyzed [15]. As for this project, the segmentation was used to identify a weak region on the pillar. For image segmentation, the k-means clustering method was implemented in both ENVI and Matlab software. When working with small datasets, K-means clustering performs well because it can separate items in photographs and produce superior results. The ENVI software image segmentation process was carried out using the software's existing k-means clustering function, whereas the Matlab software required the implementation of a simple code, as shown below.

K-means clustering:

Matlab Programming Code:

To read an image into workspace:

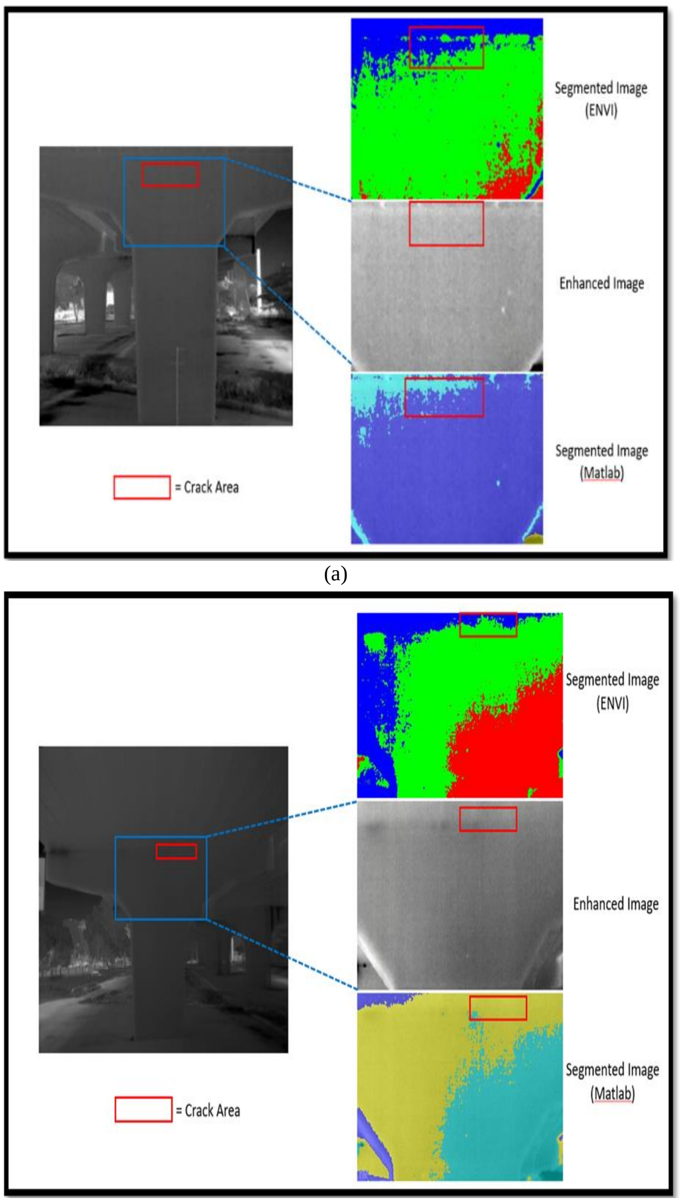
```
I = imread('34a.jpg');
imshow(I)
title('Original Image')
```

To segment the image into three regions using k-means clustering:

```
[L,Centers] = imsegkmeans(I,3);
B = labeloverlay(I,L);
imshow(B)
title('Segmented Image')
```

V. Result and Analysis

To begin with, the enhanced image in Fig. 8 below shows that the quality of the thermal image has been improved as it has become brighter. These findings will make the image segmentation process more visible. The K-means clustering method of image segmentation divides the image into several predefined clusters and groups the different clusters with the same color to create the segmentation region. The segmented image in Fig. 8 shows that the results are quite similar, but the ENVI result appears to produce a much more detailed result. The segmentation process produced three distinct clusters, with the hottest area assuming to be the most affected area of defect.



(b)

Fig. 8. Image Enhancement and K-Means Segmentation Results. (a) 34A (b) 34B.

Subsurface deficiencies in concrete heat up at faster rate than sound surface which means the hotter region on the concrete surface appear to be the weak region that could possibly be a crack. The algorithm cluster thermal image by extracting the hottest pixel is brightest blob for great contrast between defect and non-defect area. The output from the segmentation process on the enhance thermal image it can be seen that the hotter region falls in to the crack area but in the segmented image there are also other hotter region on the pillar that fall out of the crack area and this is indicating that area might also have defect or just another weaker spot.

VI. Discussion and Conclusion

This study was successful in identifying a new method for detecting high concrete structure defects that was integrated with advances in aerial data acquisition and imagery processing. As is well known, the drone platform is very useful for acquiring low altitude aerial information using a camera or sensor payload. While the image processing system can produce good segmentation for various types of surfaces. It is a difficult task to create the 3D thermal image enhancement orthoplane where there is difficulty in stitching the homogenous structure surface such as to produce 3D bridge pillar. Another critical step is to use image processing system to determine the micro level of structure cracks from thermography image. As a result of this breakthrough, another advanced NDT process method for structure inspection is now available.

Furthermore, the capability of thermography imagery in dealing with the problem of defect detection on a concrete surface can be seen. This new method is quite useful and effective in that thermal imagery can be interpreted into other forms of imagery through the process of segmentation, providing a better initial view to solve the problem. Thermography imagery will always be an interesting NDT method for monitoring the health or condition of structures, not just concrete structures, but also other types of structures in a cost-effective manner. Sirca Jr and Adeli (2018) [16] state that this is due to the unique environmental constraints and the massive amount of data that must be processed, applying the thermography method of defect detection to large scale three-dimensional concrete structures such as power plant cooling towers and hydroelectric dams will remain be a challenge. This statement has surely been a reason to keep on studying and explore the potential of the thermography imagery in solving the structure defect problem.

In conclusion, the aim of this study in which to study the performance of thermal imaging technique for detecting the defect of bridge surface has been achieved as the final result of the study has shown a positive outcome. Furthermore. The objective which is to demonstrate concrete surface defect detection technique using UAV thermography imagery and to detect the defect on bridge pillar surface using segmentation method by utilizing thermal group of UAV remote sensing has also been achieved. This study has profoundly used the method of close-range UAV in collecting thermal which then the data were processed into a better form of imagery in order to conduct the segmentation process that produce a clear vision to the defect on the concrete surface.

ACKNOWLEDGMENT

The authors would like to express their profound gratitude to Faculty of Built Environment and Survey (FABU), Universiti Teknologi Malaysia (UTM) for all the support that been provided. The authors would also like to express their gratitude to PLUS Malaysia Berhad (PMB) for data collection and excellent technical assistance in this study, which was greatly appreciated. A special thanks also goes to Geolatitude Technology Sdn. Bhd., Johor, for providing drone technical expert assistance during field activities. Last but not least, the authors would like to express their gratitude to Center of Excellence, PMB for the research collaboration grant awarded under UTM CR DTD cost center no. R.J130000.7652.4C585 and FRGS grant with cost center no. R.J130000.7852.5F478.

REFERENCES

[1] Prakash, A. (2000). Thermal Remote Sensing: Concepts, Issues And Applications. International Archives of Photogrammetry and Remote Sensing. Vol. XXXIII, Part B1. Amsterdam 2000.

[2] Yao, H., Qin, R., & Chen, X. (2019). Unmanned Aerial Vehicle for Remote Sensing Applications—A Review. Remote Sens. 2019, 11, 1443; doi:10.3390/rs11121443.

[3] Cannas, B., Carcangiu, S., Concu, G., & Trulli, N. (2012). Modeling of active infrared thermography for defect detection in concrete structures. In Proceedings of the 2012 COMSOL Conference.

[4] Bian, H., Bai, L., Chen, S. E., & Wang, S. G. (2012, April). Lidar based edge-detection for bridge defect identification. In Nondestructive Characterisation for Composite Materials, Aerospace Engineering, Civil Infrastructure, and Homeland Security 2012 (Vol. 8347, p. 83470X). International Society for Optics and Photonics.

[5] X. Maldague, Theory and Practice of Infrared Technology for Non destructive Testing. Wiley, New York (2001).

[6] Tangyunyong, P., & Schmidt, C. (2011). Thermal Defect Detection Techniques. *Revised and updated from “Thermal Defect Detection Techniques” by Dan Barton and Paiboon Tangyunyong, and “Thermal Failure Analysis by IR Lock-in Thermography” by O. Breitenstein, C. Schmidt, F. Altmann and D. Karg, Microelectronics Failure Analysis, Desk Reference, Oh Edition (2011).

[7] Zheng, D., Tan, S., Li, X., & Cai, H. (2021). Research on the Infrared Thermographic Detection of Concrete under Solar Heating. Advances in Civil Engineering, 2021.

[8] Jung Yang, Wei Wang, Guang Lin, Qing Li, Yeqing Sun and Yixuan Sun (2019) Infrared Thermal Imaging-Based Crack Detection Using Deep Learning. IEEE Access Vol 7. DOI 10.1109/ACCESS.2019.2958 264. 182060-182077.

[9] Jiang, R., Jáuregui, D. V., & White, K. R. (2008). Close-range photogrammetry applications in bridge measurement: Literature review. Measurement, 41(8), 823-834.

[10] Mohan, A., & Poobal, S. (2018). Crack detection using image processing: A critical review and analysis. Alexandria Engineering Journal, 57(2), 787-798.

[11] Zhang Yiyang. The design of glass crack detection system based on image pre-processing technology, in: Proceedings of Information Technology and Artificial Intelligence Conference, 2014, pp. 39–42.

[12] S.Y. Alam, A. Loukili, F. Grondin, E. Rozie` re, Use of the digital image correlation and acoustic emission technique to study the effect of structural size on cracking of reinforced concrete, Eng. Fract. Mech. 143 (2015) 17–31.

[13] M. Salman, S. Mathavan, K. Kamal, M. Rahman, Pavement crack detection using the gabor filter, in: Proceedings of 16th International IEEE Annual Conference on Intelligent Transportation Systems, 2013, pp. 2039–2044.

[14] Samanta, S., Mukherjee, A., Ashour, A. S., Dey, N., Tavares, J. M. R., Abdessalem Karâa, W. B., ... & Hassanien, A. E. (2018). Log transform based optimal image enhancement using firefly algorithm for autonomous mini unmanned aerial vehicle: an application of aerial photography. International Journal of Image and Graphics, 18(04), 1850019.

[15] Dass, R., & Devi, S. (2012). Image segmentation techniques 1.

[16] Sirca Jr, G. F., & Adeli, H. (2018). Infrared thermography for detecting defects in concrete structures. Journal of Civil Engineering and Management, 24(7), 508-515.

Learning Effectives of Virtual Land Surveying Simulator for Blended Open Distance Learning Amid Covid-19 Pandemic

Yeap Chu Im[1], Muhammad Norhadri Bin Mohd Hilmi[2], Tan Cheng Peng[3]
Azrina Jamal[4], Noor Halizah binti Abdullah[5], Suzi Iryanti Fadilah[6]
School of Science and Technology, Wawasan Open University, Penang, Malaysia[1,2, 3, 4]
School of Housing, Building and Planning, Universiti Sains Malaysia, Penang, Malaysia[5]
School of Computer Sciences, Universiti Sains Malaysia, Penang, Malaysia[6]

***Abstract*—Many universities worldwide were forced to physically close campuses due to lockdown and resumed the in-person classes compliant with a stringent set of standards of procedures (SOPs) as Covid cases drop. This has profoundly disrupted the hands-on lab face-to-face learning process that is harder to be moved online. Virtual simulation lab could be the answer and its use in many courses has been extensively studied. However, it is relatively little studied when it comes to land surveying courses. The purpose of the study is to explore the learning effectiveness of virtual surveying field lab for blended open distance learning (ODL) students at Wawasan Open University (WOU) in the time of Covid-19. This study used a mixed-method that combines qualitative and quantitative approaches to get a fuller picture and deeper meaning of learning behavior while using descriptive and inferential statistical methods in SPSS platform. Respondents were selected using the purposive sampling method. Survey questionnaires were designed and distributed to students before and after lab simulation class. Instructors were interviewed after the lab simulation class. Students' learning results for the surveying course were compared with the past-year examination results at pre-Covid-19 times before the virtual simulator was introduced. Both qualitative and quantitative data set were collected and analyzed. The findings revealed that the virtual simulator has enhanced students' learning interest and efficiency for surveying course in a ODL setting. Both students and instructors have responded positively towards the virtual simulator learning experiences. Students' achievement in the final examination amid Covid-19 was better than pre-Covid-19 performance. It is recommended that the virtual simulator shouldn't be a replacement to physical instrument but as a complement.**

***Keywords*—Covid-19; 3D simulator; virtual laboratory; land surveying; blended open distance learning**

I. Introduction

The abrupt outbreak of Covid-19 pandemic has forced many universities worldwide to temporarily close physical campuses due to lockdown and shifted the in-person classes to online mode. However, the practical lab class that requires hands-on and is more difficult to be moved online has to be halted and to be adjourned to a much later time. Even if the hands-on lab is forcibly moved online, the hands-on lab most likely has to go hands-off, leaving out the hands-on experience whereby students passively watching the demonstration videos from the instructors, that either live or pre-recorded. Like the other countries, Malaysia resorted to the extreme measures to curb the spread of the Covid-19, began to implement the unprecedented lockdown starting in March 2020 until May 2020 with the gradually eased lockdown restrictions. Universities in Malaysia, public and private, have not been spared from the impacts of the Covid-19, were ordered by the Higher Education Ministry to stay shut physically and to go fully online until end of 2020, with a few exceptions that let students who need to carry out practical laboratory work using specialized equipment to enter the campus, subjected to the stringent set of standard operating procedures (SOPs) [1]. In January 2021, following a second wave of Covid-19, much of the country again was into its second major lockdown. The whole nation was plunged back into a full lockdown in June 2021 as third wave hit. In virtue of the national vaccine rollout success, universities finally opened their campus doors to fully vaccinated students to return starting Oct 2021, but still with strict SOPs in place [2]. Admittedly, with many un-certainties and the ups and downs of repetitive lockdowns, the Covid-19 pandemic has indeed profoundly disrupted the whole educational system and in particular, the hands-on lab face-to-face learning process.

Nonetheless, the land surveying hands-on outdoor lab has faced its own unique challenges even before the Covid-19 pandemic [3-10]. The physical surveying instruments that are generally sensitive and delicate have to be handled with great care, but wear out despite years of good maintenance, and get expensive when replacing with more up-to-date or higher quality one. Insufficient quantity of the physical surveying instruments leads to the unequal access to the instruments when students are asked to practice in bigger groups. The practical session is further bounded by the location within the vicinity of the campus with the same terrain characteristics due to accessibility and safety issues; the unpleasant external environmental conditions like haze or rainy weather especially in this tropical climate zone; and the allotted time for the hands-on lab to cover fairly each surveying topics in the course syllabus. The demonstration on the operation of the surveying instruments from the instructors may not be seen clearly by the students in the field, particularly some detailed operation steps with tiny finger motion like setting up the instruments. After the demonstration, students are usually dispatched in groups to carry out the fieldwork, often with little supervision, in the end, are asked to report their fieldwork results to the instructors and have no way to trace back their multi-steps process. All these factors have thrown wrenches to the learning process in getting students to familiarize with the use of modern surveying instruments and preparing them adequately for surveying work in the real world.

With these challenges in mind, virtual land surveying simulator seems like the perfect solution. But then in-depth study on the learning effectiveness of virtual surveying simulator should be conducted before incorporating it into the land surveying course curriculum. Although the use of virtual simulator has been extensively studied in many fields, but there have been relatively much lesser studies of its development and implementation when it comes to architecture-engineering-construction education, especially for land surveying courses, in an open distance education setting, to say the least [11-14]. Recognizing the value that comes after it, this study kicked off way before Covid-19. While working on the solution, this study proved to be very timely during the nationwide shutdown of on-campus teaching due to the Covid-19 surge. Therefore, the main objective of the study is to explore the learning effectiveness of virtual surveying field lab simulation for blended open distance learning (ODL) students at Wawasan Open University (WOU) in the time of Covid-19.

II. Literature Review

Ever since the emerging of this new virtual reality technology, computer generated simulation has been exponentially increased its use in many education and training courses involving hands-on laboratories in wide spectrum of fields, in particular, dominating mainly in science, engineering and medical, leading the trend by renowned universities and institutions like Massachusetts Institute of Technology (MIT), Carnegie Mellon University, to name a few; and its application has been extensively studied in the literature review over the past decade, showing great deal of interest in this topic worldwide [15-19]. In architecture-engineering-construction (AEC) education, the underpinning VR technology has been adopted to support design, visualization, structural analysis, hazards recognition, safety training, decision-making, operational training of heavy equipment such as cranes; without facing the dire do-or-die consequences as in real construction project.

Several successful attempts to explore in-depth the virtual laboratory for construction surveying are deemed valuable and commendable. Using QuickTime VR technology, Ellis et al. (2006) implemented an interactive multi-media learning re-sources for leveling survey only and evaluated its application among undergraduate students in the School of the Built Environment at Leeds Metropolitan University [20]. Mills and Barber (2008) employed virtual interactive learning tool running on AJAX for traversing survey only at Newcastle University's School of Civil Engineering and Geosciences [4]. In addition to levelling and traversing, a virtual simulator for survey training, SimuSurvey and an improved version using XNA gaming platform, SimuSurvey X, were developed at the Department of Civil Engineering in National Taiwan University that cover other surveying topics as well [3, 6, 21]. In the College of Technology at Purdue University, a virtual learning environment was created and a pilot study with undergraduate students was conducted for one chaining module only in surveying fundamental course before extending to differential level surveying module. [7, 22] Gao and Wu (2017) designed an experimental virtual surveying system using VR modeling language for students in Henan University of Urban Construction. [8] León and Morales (2018) created interactive survey instructor (ISI) that was being used by the surveying engineering school from University of Costa Rica for teaching two modules only, differential leveling and traversing [23]. In response to the Covid-19 pan-demic, Tingerthal and Kaoni (2021) prepared a virtual laboratory centered on a series of interactive videos for students at Northern Arizona University, without the gamifying survey experience that allows students to operate virtual instruments [10]. Whereas Bolkas et al. (2021) presented surveying reality (SurReal) software solution that simulated the virtual reality environment and surveying level instrument in an immersive and interactive virtual laboratory based on levelling task only at Pennsylvania State University Wilkes-Barre campus [9].

Triumphantly, all the findings of these studies related to the virtual laboratory for construction surveying in recent decade bring forth positive results and reach a conclusion on the betterment of teaching and learning process with the use of virtual simulation laboratory for surveying when querying about participants' perception [3, 4, 7, 9, 10, 20, 22, 23]. The findings are equally exciting when measuring the performances of the students. Kuo et al. (2007) found that approximately two-third of the students were able to answer correctly the follow-up quizzes after simulator lab class [3]. Hazar and Nicoletta (2011) presented the students' average e-grade using with the simulator was 65% [7]. Hazar et al. (2014) saw an increase of declarative knowledge by 28% in a given written test and procedural knowledge by 30% in a field surveying exercise among students who used the simulator [22]. Gao and Wu (2017) acclaimed the success of the university in winning a total of 60 person-time awards of undergraduate surveying skill competitions at national and provincial level, owing to the great teaching effect of the simulator [8]. Tingerthal and Kaoni (2021) reported students who completed 19 or more out of the 23 practice quizzes that interleave in the virtual lab has achieved higher scores (average=93.3%) on the summative quiz at the end of the module than students who completed less (average=63.7%) [10].

However, it is relatively little studied on the development and implementation of the virtual simulator in the field of architecture-engineering-construction, particularly, the land surveying course, in comparison to the vast study of simulator in other fields. Most of the universities involved in the studies are conventional brick-and-mortar universities with surveying courses primarily taught in-person, while developing the virtual surveying simulator as a webpage that is hosted on the internet server [20, 4, 8, 10, 20] or downloadable on personal computers with low-end graphic cards [3, 7] or as a lightweight web application [23] or as a software installed on a stationary workstations in the campus that work faster than a personal computer [9]. It is therefore fewer studies when it comes to pedagogical implementation in an open distance education setting in the pre Covid-19 era and also during the outbreak of Covid-19. This study is to answer the research question at the onset of the study whether the virtual surveying simulator has any effect on the learning interest and efficiency for open distance learning students amid Covid-19, that will help to fill the research gaps.

III. Research Methodology

A. Research Setting

Basic surveying course is the basic major course within the undergraduate construction management program at Wawasan Open University (WOU). It is an introductory course to the fundamentals of construction surveying that comprises the basic principles and branches of surveying, the use of survey instruments such as auto level, theodolite, EDM, total station, etc. the various common methods of surveying procedures such as levelling, traversing, bearing etc.; and the applications of surveying in building construction. Students from main campus in Penang and four geographically apart study centers at different locations in Perak, Johor, Kuala Lumpur and Sarawak, are required to take this basic surveying course in the first year of their studies and to complete it in one semester. In year 2020, there were 208 active current undergraduate construction management program students and about 30 of them taking basic surveying course in each year cohort. These students were invited to participate in the virtual surveying field lab simulation class, along with 21 graduating students and 53 graduated students as well, so to compare between those with and without prior knowledge of surveying background.

B. Ecosystem Approach

Being an ODL university, WOU robust virtual learning spaces are naturally flexible and accessible to cater to students' online and remote learning experience as shown in Fig. 1, even despite the interruption brought about by Covid-19. Learning management system (LMS) at WOU uses MOODLE that delivers one stop service for students [24]. Through the LMS gateway, the surveying course page provides for students all the surveying materials at one place including surveying course materials, lab manual, lab procedures, demo videos and animations, quizzes, course assignments, past year exam papers, online forum, announcements, additional web resources, library learning resources etc. On the other hand, for instructors, it provides LMS analytic reports of users' access and views, customization, instructor supports, helpdesk etc. The online submission system (OAS) is where the students can submit, resubmit, apply submission extension, and view the grades and feedback of their written assignments whilst the instructors can assess the assignments, give comments, moderate and export the marks, approve students' request, set deadlines and mark allocation etc. However, there have been physical surveying field labs with practical training initially conducted on-campus that had to go online due to the nationwide shutdown of on-campus teaching. As such, the virtual surveying lab simulators that are available to students and instructors to install onto personal computers / laptops or to access it through local network remotely made it possible for online surveying lab classes to continue effectively.

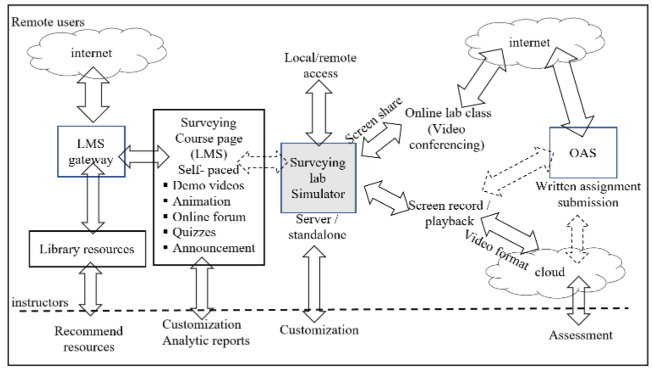


Fig. 1. Ecosystem of Virtual Learning Platforms for Surveying Courses.

C. Virtual Land Surveying Simulator

Amid the Covid-19 pandemic, virtual land surveying simulator was introduced for the first time to students before the physical surveying field lab class was conducted on WOU main campus and four regional learning centers, with the stringent SOPs in place such as masking, temperature screening, social distancing, MySejahtera scanning etc. After the first introduction, students would then access it remotely by downloading and installing the virtual simulator onto their own personal devices or to access it through local network, for their own practices. The virtual surveying simulator used in this study is SimuSurveyX, an improved version of a computer-based survey surveying simulator, SimuSurvey that was first developed by researchers from Department of Civil Engineering in National Taiwan University using OpenGL graphic library and the C# object-oriented programming (OOP) language [21]. A feasibility study on the original version of the virtual surveying simulator was conducted [3]. Following that the user interface of SimuSurvey was then redesigned by using user-centered design (UCD) approach [5]. The original version was further improvised by using XNA gaming platform. SimuSurveyX provides a virtual environment and simulates various surveying instrument and accessories for users' manipulation controlling like auto level, theodolite, total station, tripod, reflector prism, level staff and ranging rod. It covers five basic surveying training activities namely leveling, horizontal angle & vertical angle surveying, traversing and free mode surveying [6].

D. Sampling Method and Instrument

Prior to the virtual surveying field lab simulation classes with students, instructors attended the training workshop on the operation of the virtual surveying instrument through the scheduled meeting online. Pre and post survey questionnaires were de-signed and distributed to the students before and after the virtual surveying field lab classes. There were three sections in the pre-survey whereas there were two sections in the post-survey. The instructors who are also the subject experts were interviewed after the virtual lab classes as well. The total respondents who have responded voluntarily to participate in this study are 28 students and 7 instructors of WOU. Both students and instructors were identified and selected as the respondents using the purposive sampling method. During the virtual lab classes, instructors gave demonstration and guided the students for five virtual surveying activities namely distance measurement, angle & direction measurement, levelling, traversing and calculation related to surveying, in order to achieve the learning objectives whereby students should be able to set up the virtual surveying instruments, to perform measurements using different virtual surveying instruments and to carry out calculation related to surveying.

E. Data Collection and Analysis Procedures

At the end of the virtual lab classes, students were asked to answer the quizzes related to the surveying activities and their answers were graded. As the basic surveying course is offered at WOU every academic year, students' learning results for basic surveying course are tested again in the final examination at the end of the semester and these results were compared with the past year examination results at pre-Covid-19 times before the virtual surveying simulator was introduced. Mixed method combining qualitative and quantitative approaches is used in this study whereby the quantitative data set was collected and analyzed using descriptive and inferential statistical methods in Statistical Package for Social Sciences (SPSS v. 28) software platform while the qualitative data set was examined further to get a fuller picture and to perceive deeper meaning of students' learning behavior, with the primary objective to explore the learning effectiveness of virtual land surveying field lab simulation for blended open distance learning (ODL) students at WOU in the time of COVID-19 pandemic.

IV. Results

A. Demographic Characteristics

The virtual surveying field lab simulation class was planned for construction management program students in year 2020, there were 208 active current undergraduate, 21 graduating students and 53 graduated students invited to participate the study. A total of 28 students responded. Two anonymous pre and post survey questionnaires were designed in the form of binary yes or no questions, multiple choices, a 4-point Likert-type scale, as follows, 1=strongly disagree to 4=strongly agree, and open-ended questions whereby their other comments were also welcome. In the pre-survey questionnaires, there were three sections. The first section was to attain the background of students that include age, year of studies, prior knowledge in surveying courses, work and surveying experiences etc. Overall, we had a good balanced distribution among participants with diversified demographic and background characteristics (Table I).

B. Pre-Survey

The second section of the pre-survey questionnaires was to gauge students' attitude towards conventional surveying course without virtual surveying simulator that comprises of three parts: students' degree of interest in surveying course itself, their degree of interest in operation of physical surveying instrument and their challenges in handling the physical instrument. Students were asked to indicate their level of agreement with statements in this section. Using boxplot, Fig. 2 to 5 illustrates the students' responses to the statements. Fig. 2 shows that students generally were keen in surveying course itself, for example, statement 1: I am very interested in the surveying course (median=3); statement 9: the surveying course prepares me for the sub-sequent higher-level courses in my overall study (median=3); and statement 13: I think the surveying course is useful for my profession (median=3).

Fig. 3 shows that students mostly were keen in learning how to operate physical surveying instrument, for example, statement 1: I am very interested in the operation of physical surveying instrument (median=3); however, majority of the students disagree statement 14: operating physical surveying instrument is easy (median=2); and they are willing to spend more time to practice on how to operate physical surveying instrument after leaving lab class in statement 18 (median=3).

Fig. 4 shows that students faced many challenges in handling the physical surveying instrument, for example, statement 1: the physical surveying instrument is not enough for students to take turn to practice (median=3); statement 17: I want more interaction between instructors and students (median=3); statement 18: I need more guidance from instructor when I operate surveying instrument; and statement 20: I cannot understand without audio visual teaching aids used by instructor (median=3).

TABLE I. DEMOGRAPHIC CHARACTERISTICS OF PARTICIPANTS

Demographic characteristics		**%**
Age	20 – 34 years	50
	35 – 44 years	39
	45 – 54 years	11
Years of studies	Year 1 – 2	36
	Year 3 – 4	43
	Year 5 or more	14
	Graduated	7
Taken basic surveying course at WOU	Taking now	57
	Taken already	22
	Not taken	21
Taken land surveying course during diploma / technical / vocational studies	Yes	39
	No	61
Construction work experience	Less than 1 year	11
	1 – 5 years	36
	6 – 10 years	14
	More than 10 years	39
Surveying experience at workplace	Yes	43
	No	57

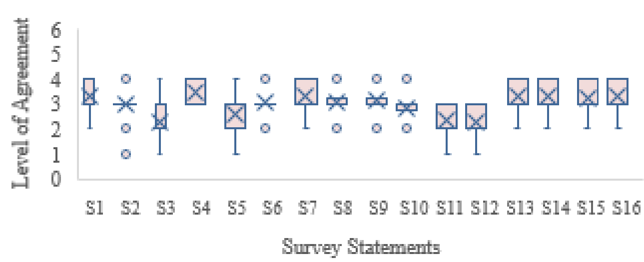


Fig. 2. Students' Responses to Statements Regarding their Degree of Interest in Sruveying Course.

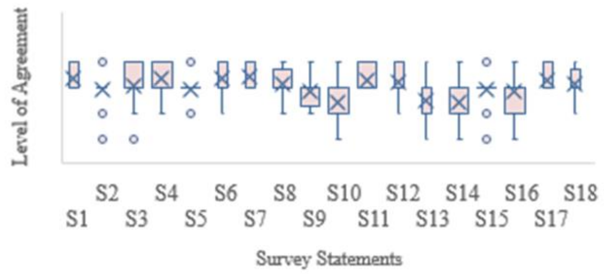


Fig. 3. Students' Responses to Statements Regarding their Degree of Interest in Operation of Physcial Surveying Instrument.

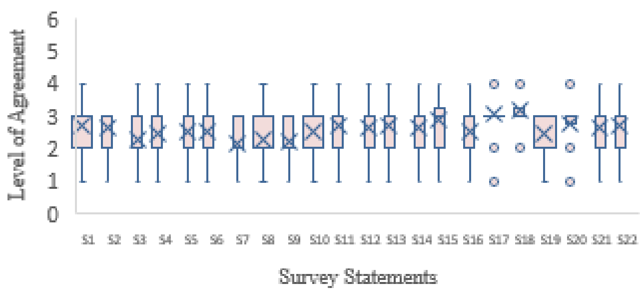


Fig. 4. Students' Responses to Statements Regarding their Challenges in Handling the Physcial Surveying Instrument.

The third section of the pre-survey questionnaires was to gauge students' attitude towards unconventional surveying course with virtual surveying simulator that comprises of two parts: students' degree of interest in learning how to operate virtual surveying instrument and their readiness to run virtual simulator on their personal devices such as their device types, window version, graphic card, internet connection, etc. Fig. 5 shows in general students were keen in learning how to operate virtual surveying instrument, for example, statement 1: I am very interested in the operation of virtual surveying instrument (median=3) and they are willing to spend more time to practice on how to operate virtual surveying instrument after leaving lab class in statement 16 (median =3). All the students possessed personal devices either laptop or desktop with Windows 7 and above. 61% of their devices have graphic card. 75% of the students had stable internet connection access and 86% of them were comfortable to download and install virtual simulator on their own.

C. Post Survey

Whereas in the post survey questionnaires, there were two sections in the post-survey. The first section was to find out if any changes of students' attitude to-wards unconventional surveying course with virtual surveying simulator after participating the virtual surveying field lab simulation classes. This section comprises of students' rating on self-efficacy, effectiveness of learning, instructors, virtual surveying activities, virtual surveying software, lab manual and exercise provided during the virtual lab simulation classes. Fig. 6 shows students' ratings and the students in general have responded positively towards the virtual surveying simulator learning experiences.

The second section of the post survey questionnaires was to determine students' preference between surveying course with and without virtual surveying simulator. Fig. 7 shows that majority of the students agreed the surveying course with virtual simulator is better, corresponding to the challenges with the physical instrument, for example, statement 1: virtual surveying instrument is better because every student can handle the surveying instrument at their desktop (median=3); statement 17: virtual simulator is better because there is more interaction between instructors and students (median=3); statement 18: virtual simulator is better because instructor could provide more guidance when I operate surveying instrument; and statement 20: virtual simulator is better because instructor could use it as audio visual teaching aids (median=3).

Fig. 5. Students' Responses to Statements Regarding their Degree of Interest in Operation of Virtual Surveying Instrument.

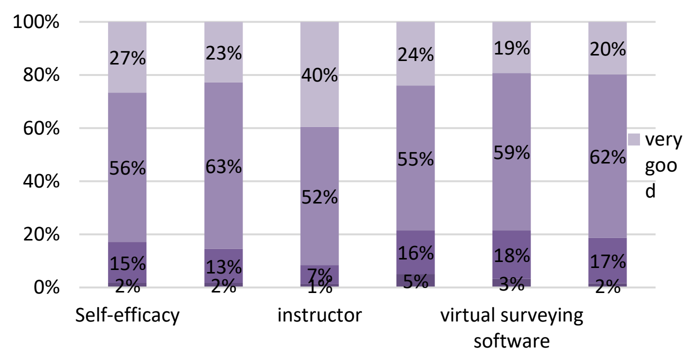


Fig. 6. Students' Rating on their Learning Experiences towards Surveying Simulator Lab Class.

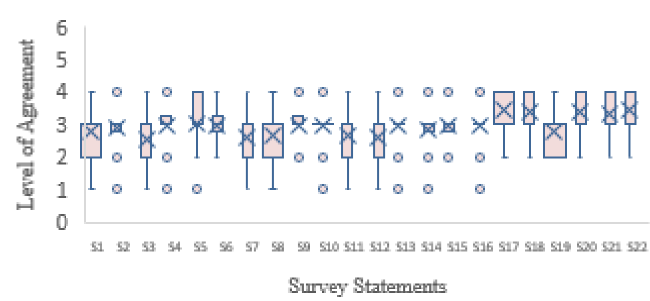


Fig. 7. Students' Response to their Preference between Surveying Course with and without Virtual Surveying Simulator.

D. Comparison between Pre and Post Survey

Students' level of agreement on 10 parallel survey statements each related to learning interest and learning efficiency in surveying respectively from both pre and post survey are compared. A parametric test i.e., student t test for dependent or paired t test was used to find out any statistically significant mean change in learning interest and learning efficiency of surveying following the virtual simulator lab class as sample size, $n < 30$. The distribution is normal after verifying with the Kolmogorov-Smirnov test as p-value > 0.05. The mean±SD of learning interest in surveying before surveying simulator lab class was 30.86±2.965, and mean±SD after simulator lab class was 32.21±3.985. There was significant mean change in learning interest of surveying after simulator class as p-value < 0.05 ($p = 0.022$) (Fig. 7). The mean±SD of learning efficiency in surveying before surveying simulator lab class was 26.86±3.608, and mean±SD after simulator lab class was 30.68±3.486. There was also significant mean change in learning efficiency of surveying after simulator class as p-value < 0.05 (p =<0.001) (Fig. 8).

E. Direct Method

Besides reporting on the indirect method that captures students' perceptions of their learning experiences through survey questionnaire, a direct method involving an evaluation of students' quizzes in answering four different surveying topics upon completion of the virtual lab class (Fig. 9) and also the comparison of the year end final examination results of surveying course with the past year results before the introduction of virtual simulator (Fig. 10) were also employed in this study.

Fig. 9 illustrates that more than half of the students, i.e., 61% of them, passed their quizzes (average=44%), the lowest score for students who failed was 13% and the highest score was full mark, 100%. Majority of the students agreed that they had no difficulties in understanding the questions provided in the quizzes (median=3) and better understanding of the questions after carrying out the virtual surveying activities (median=3), but think they would be able to solve the questions if they have more time (median=3).

Fig. 10 depicts that the mean±SD score for group in year 2020 is higher at 67.70±21.544 as compared to year 2019 (mean±SD = 41.81±21.360) and 2018 (mean±SD = 57.93±21.251). One-way ANOVA analysis confirms that there is a statistically significant difference between the mean score for each group as the significance value is 0.000 ($p = 0.000$), which is p-value < 0.05. The Tukey's HSD test i.e., the preferred test for conducting post hoc tests on a one-way ANOVA. The multiple (pair-wise) comparisons of scores between the groups shows that there is a statistically significant difference in the scores between year 2020 and 2019 ($p = 0.000$) and be-tween year 2019 and 2018 ($p = 0.029$). However, there is no significant difference of the scores for year 2020 and 2018 ($p = 0.198$). In short, this proves that the virtual simulator lab class introduced to the group in year 2020 helped the students scored better in their final examination as compared to the groups not introduced with the simulator lab class in year 2019 and 2018.

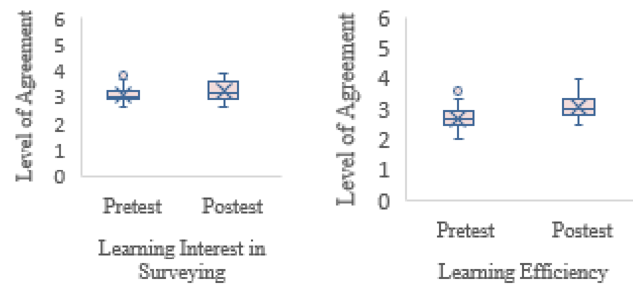


Fig. 8. Comparison between Student's Responses for Pre and Post Virtual Simulator Lab Class.

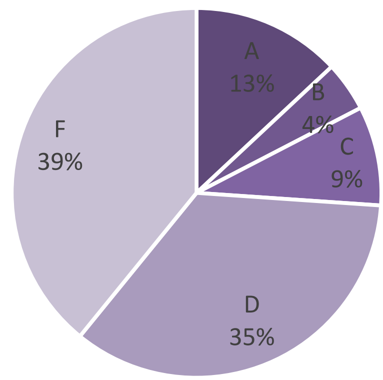


Fig. 9. Students' Quizzes Result at the End of the Simulator Class.

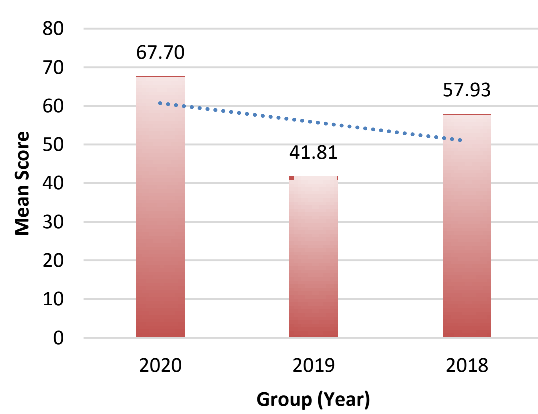


Fig. 10. Comparison of Mean Scores for Final Exam Results in Year 2020, 2019 and 2018.

F. Instructors' Rating

While the virtual simulator was received positively by the students, instructors were also asked to give their rating on students' learning experiences that comprises of learning interest, learning effectiveness, interaction between instructors and students, virtual surveying activities, virtual surveying software, lab manual and exercise pro-vided, based on their observation during the virtual lab class. (Fig. 11) The instructors in general have responded positively towards the virtual surveying simulator learning experiences.

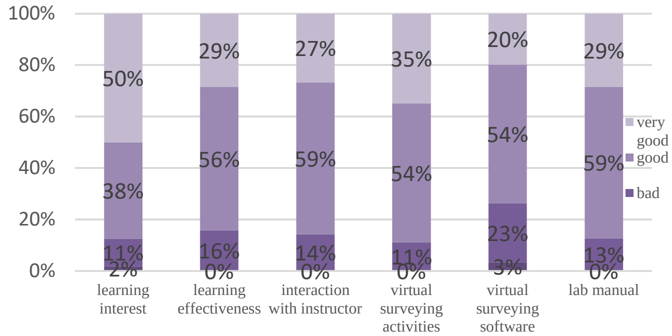


Fig. 11. Instructors' Rating on Students Learning Experiences towards Surveying Simulator Lab Class.

V. Discussion

The statistically significant and positive results from the study clearly answer the research questions posed at the onset of the study. It is hypothesized that the virtual land surveying simulator will enhance students' learning interest and learning efficiency for land surveying course in blended distance learning setting at Wawasan Open University (WOU) in the time of Covid-19 pandemic. The null hypothesis of this study, which is, virtual land surveying simulator will have no effect at all on the learning effectiveness has been rejected. This finding is established when comparisons are made in between pre and post survey statements related to learning interest and learning efficiency in surveying respectively; in between the indirect method reflecting students' perceptions of their learning experiences and the direct method evaluating students' performance through quizzes and year end final examination; and in between students' rating and instructors' rating towards surveying simulator lab class.

In order to explain the results obtained, the instructors' interview and observation gave some insights on the students' learning behavior. The instructors observed that during the simulator class students paid full attention to them, the interaction between them and students were good, eventually students could clearly understand the learning objective and were able to operate the virtual surveying instrument as instructed. They also commented that during the learning process, virtual surveying simulator helped students to understand some of the surveying theories and concepts that are abstract. The virtual surveying simulator allowed students to screen record the surveying activities history to the cloud / drive to share with the instructors in order to playback the recorded surveying activities to review students' learning problem and to provide constructive feedback.

It is found that the virtual surveying software is simple and easy to use, however, it is noted that there are some limitations with the software, for example some components of the virtual surveying instrument such as screws, knobs, buttons are not designed for students to manipulate, and some details of the virtual instrument operation procedure and interface are missing for students to learn. Students were not able to learn teamwork skills as they couldn't cooperate with the other course mate in a team to carry out the virtual surveying activities together. Despite those minor flaws, the instructors and students still perceived the virtual surveying simulator has effectively enhanced the learning interest and efficiency in both surveying course and in physical surveying activities.

In an open distance learning (ODL) setting, when linking the demographic characteristics of the participants, it was found that ODL students whom were largely made up of working adults, mature in age, more computer savvy with technology readiness, have certain measures of exposures to surveying experiences at their workplace or have some prior knowledge in surveying by taking land surveying related courses during pre-university study, they were more interested and had positive attitude in surveying, physical surveying instrument and virtual surveying simulators, as p-value < 0.05 ($p =< 0.001$) using paired T-test. Their interests to discover more aspects of surveying were further enhanced with the virtual surveying simulator. Their surveying skills and understanding were reinforced through repeated practices and exercises using surveying simulator after classes.

Differing from the previous studies conducted by the conventional universities [3, 4, 7, 8, 9, 10, 20, 22, 8, 23], as an ODL university that offers blended or fully open distance learning delivery without regular face-to-face (F2F) contact in the physical classroom, WOU in this study adopted an ecosystem approach whereby the virtual land surveying lab simulator is part of that learning support system on 24x7 basis that comprises of online learning management system (LMS) and online assignment submission system (OAS) that provide students access to self-directed learning materials, online counselling and tutoring, electronic library resources, forum discussions etc. from anywhere remotely and at any time with personalized and self-paced learning experience. In assuring the quality of online teaching and learning, the quality assurance (QA) components are being integrated within the LMS of WOU [25]. Even during the Covid-19 lockdown, nationwide, the regional learning centers and offices in the urban cities that are equipped with physical libraries, laboratories, computer terminals and administrative support were closed temporarily, the interruption in learning and teaching process was able to be kept to a minimal level.

In this study, it is also recognized that virtual surveying simulator comes in handy especially amid pandemic outbreak, so that students could still carry out the surveying activities virtually, as the university couldn't get enough physical surveying instruments for students to keep their social distancing since the purchasing and maintenance cost of physical instruments is often expensive. This same sentiment is also echoed by other affected universities as well. Bolkas et al. (2021) pointed out that the Covid-19 pandemic has pushed universities to remote learning and opened the door to implement the novel desktop-based VR platform for surveying labs. [9] Tingerthal and Kaoni (2021) also expressed that the Covid-19 pandemic has forced the universities to pivot in the modality of online teaching and proposed the virtual surveying laboratory to incorporate even into a post Covid-19 curriculum. [10].

VI. Conclusion

It is concluded that the virtual land surveying simulator has shown statistically significant and positive results in stimulating learning interests and enhancing effectiveness for land surveying courses among blended open distance learning (ODL) students at Wawasan Open University (WOU) in the time of Covid-19 pandemic, solving many bottleneck problems associated with conventional surveying course. Both students and instructors have responded positively on the learning experiences towards the virtual surveying simulator lab class. Students' quizzes result at the end of the simulator class overall was satisfactory. Moreover, after the introduction of virtual surveying simulator, students' achievement in the year end final examination amid Covid-19 was better than pre-Covid-19 performance.

It is strongly recommended that the virtual surveying simulator should be added to the surveying field lab along with the physical surveying instrument, but never as a substitution to do away with the physical surveying instrument. It is suggested that the use of virtual surveying simulator in designing teaching activities as an effective visual teaching aid. More activities related to field of surveying should be added and at the same time students should be given more time to complete those surveying activities, at least a couple hours more. It is preferred that students first learn the physical surveying instrument to carry out the physical activities before learning the operation of the virtual surveying simulator to carry out the virtual surveying activities. Smaller class size is ideal for both physical and virtual surveying lab classes.

It will be still a long time before we could return to pre-Covid era and in the meantime, we will be living together with coronavirus in the new normal. Even in a post-Covid world, the landscape of education will never look the same anymore. More future work will focus on the immersive lab learning experiences with the promising use of virtual reality (VR), augmented reality (AR) and artificial intelligence (AI). And the virtual lab simulator could model after unmanned aerial vehicle / unmanned aircraft system (UAV/UAS) or known as drone, which keeps rising as a viable alternative to traditional land surveying. Greater samples size could be collected and more academic years could be monitored to compare the examination scores in the future research.

ACKNOWLEDGMENT

This study is supported by the Centre for Research & Innovation (CeRI) grant that sponsored the visiting scholar program to the Department of Civil Engineering, at National Taiwan University for the research collaboration on development of SimuSurveyX. We wish to express gratitude to Professor Patrick Shang-Hsien Hsieh and Professor Jen-Yu Han of National Taiwan University.

REFERENCES

[1] Ministry of Higher Education, "Media statement: pengendalian aktiviti akademik di kampus institusi pendidikan tinggi semasa dan pasca perintah kawalan pergerakan" available at: https://www.mohe.gov.my/en/broadcast/media-statements/pengendalian-aktiviti-akademik-di-kampus-institusi-pendidikan-tinggi-semasa-dan-pasca-perintah-kawalan-pergerakan assessed 27 May 2020.

[2] Ministry of Higher Education, "Media coverage: breathing life back into higher education available at: MoHE - Breathing Life Back Into Higher Education" assessed 19 September 2021.

[3] Kuo, H.L, Kang, S.C., Lu, C.C., Hsieh, S.H. and Lin, Y.H., "Feasibility study: using a virtual surveying instrument in surveyor training," in: the proceedings of the International Conference on Engineering Education (ICEE, 2007) 3-7 September 2007, Coimbra, Portugal.

[4] Mills, H. and Barber, D., "A virtual surveying field course for traversing," in Sharing good practices: E-learning in surveying, geo-information sciences and land administration FIG International workshop 2008, 11-13 June 2008, Enschede, The Netherlands.

[5] Lu, C.C., Kang, S.C., Hsieh, S.H. and Shiu, R.S., "Improvement of a computer-based surveyor-training tool using a user-centered approach," Advanced Engineering Informatics, Vol. 23, 2008, pp. 81-92, doi: https://doi.org/10.1016/j.aei.2008.07.001.

[6] Kang, S.C., Chuang, S.K., Shiu, R.S., Chen, Y. and Hsieh, S.H., "SimuSurvey X: an improved virtual surveying instrument running off a game engine," in: proceedings of the International Conference on Computing in Civil and Building Engineering (ICCCBE, 2010) July 2010, Nottingham, UK.

[7] Hazar, D. and Nicoletta, A. V., "An innovative software application for surveying education, Computer application in engineering education," Vol. 22, No. 3, 2011, pp. 551-562, doi: https://doi.org/10.1002/cae.20580.

[8] Gao, C.Y. and Wu, Q. T., "Design and practice of surveying experiment system based on a virtual platform," International Journal of Emerging Technologies in Learning (iJET), Vol. 12, No. 4, 2017, pp. 53-61, doi: https://doi.org/10.3991/ijet.v12i04.6924.

[9] Bolkas, D., Chiampi, J., Chapman, J., Fioti, J., Gaffney, D., "Surveying Reality (SurReal): software to simulate surveying in virtual reality," ISPRS International Journal of Geo-Information 2021, Vol. 10, 2021, No. 5, doi: https://doi.org/10.3390/ijgi10050296.

[10] Tingerthal, J. and Kaoni, K., "Implementing a virtual surveying lab," in: 2021 ASEE Virtual Annual Conference, 26-29 July 2021, Virtual Meeting.

[11] Chi, H.L., Kang, S. C., Wang, X., "Research trends and opportunities of augmented reality applications in architecture, engineering, and construction," Automation in Construction, Vol. 33, 2013, pp. 116-122, doi: http://dx.doi.org/10.1016/j.autcon.2012.12.017.

[12] Rahimian, F. P., Arciszewski, T., Goulding, J. S., "Successful education for AEC professionals: case study of applying immersive game-like virtual reality interfaces," Visualization in Engineering 2014, Vol. 42 No. 4.

[13] Li, X., Yi, W., Chi, H. L., Wang, X., Chan, A. P. C. "A critical review of virtual and augmented reality (VR/AR) applications in construction safety," Automation in Construction, Vol. 86, 2018, pp. 150-162, doi: https://doi.org/10.1016/j.autcon.2017.11.003.

[14] Wang, P., Wu, P., Wang, J., Chi, H.L., Wang, X., "A critical review of the use of virtual reality in construction engineering education and training," International Journal of Environmental Research and Public Health, Vol. 15, No. 6, 2018, pp. 1204, doi: https://doi.org/10.3390/ijerph15061204.

[15] Freina, L. and Ott, M., "A literature review on immersive virtual reality in education: state of the art and perspectives," in: conference eLearning and software for education (eLSE, 2015) April 2015, Burcharest, Romania.

[16] Potkonjak, V., Gardner, M., Callaghan, V., Mattila, P., Gütl, C., Petrović, V. M., and Jovanović, K., "Virtual laboratories for education in science, technology, and engineering: A review, Computers & Education," 95(2016), 2016, pp. 309-327, doi: https://doi.org/10.1016/j.compedu.2016.02.002.

[17] Salmerón-Manzano, E. and Manzano-Agugliaro, F., "The Higher Education Sustainability through Virtual Laboratories: The Spanish University as Case of Study". Sustainability, Sustainability 2018, Vol. 10, No. 11(4040), 2018 doi: https://doi.org/10.3390/su10114040.

[18] Checa, D. and Bustillo, A., "A review of immersive virtual reality serious games to enhance learning and training." Multimedia Tools and Applications 2020, Vol. 79, No. 4, 2020, pp. 5501-5527, doi: https://doi.org/10.1007/s11042-019-08348-9.

[19] Shalaunda, M.R. and Kent, J. C., "Virtual laboratories in undergraduate science and engineering courses: a systematic review, 2009 – 2019," Journal of science education and technology 2021, Vol. 30, 2020, pp. 16-30, doi: https://doi.org/10.1007/s10956-020-09866-0.

[20] Ellis, R.C.T., Dickinson, I., Green, M., Smith, M., “The implementation and evaluation of an undergraduate virtual reality surveying application,” in: proceedings of the Built Environment Education Conference (BEECON, 2006), 12-13 September, 2006, London, UK.

[21] Lu, C.C., Kang, S.C. and Hsieh, S.H., “SimuSurvey. A computer-based simulator for survey training,” in: proceedings of the 2007 W78 Conference, 26-19 June 2007, Maribor, Slovenia.

[22] Hazar, D., Nicoletta, A. V., Garver, S., “An interactive virtual environment for learning differential levelling: development and initial findings,” Advanced in Engineering Education (ASEE), Vol. 4, No. 4. 2014.

[23] León, J. G. and Morales, G. L., “The interactive surveying instructor (ISI) for teaching optical-mechanical instruments readings in surveying engineering,” in: International Association for Development of the Information Society (IADIS) International Conference on e-Learning, 17-19 July 2018, Madrid, Spain.

[24] Chew, B. L., Rahim, M. A., Vighnarajah, “Integration of EBSCO discovery service widget into the learning spaces of LMS: A case study of Wawasan Open University,” Asian Association of Open Universities Journal, Vol. 12, No. 2, 2017, pp. 137-153, doi: https://doi.org/10.1108/AAOUJ-01-2017-0013.

[25] Liew, T.K. and Toeh, A.P. “Assuring the quality of online teaching and learning: the case of Wawasan Open University,” Asian Association of Open Universities Journal, Vol. 7, No. 1, 2012, pp. 13-33.

Fusion of Statistical Reasoning for Healing Highly Corrupted Image

Golam Moktader Daiyan[1]
School of Computer Science and Engineering
Digital Media Technology Key Laboratory of Sichuan Province
University of Electronic Science and Technology of China

Leiting Chen[2]
School of Computer Science and Engineering
Digital Media Technology Key Laboratory of Sichuan Province
University of Electronic Science and Technology of China
Institute of Electronics and Information Engineering in Guangdong, University of Electronic Science and Technology of China

Chuan Zho[3]*
School of Computer Science and Engineering
University of Electronic Science and Technology of China
Digital Media Technology Key Laboratory of Sichuan Province, University of Electronic Science and Technology of China
Institute of Electronics and Information Engineering in Guangdong, University of Electronic Science and Technology of China

Golam Moktader Nayeem[4]
School of Computer Science and Engineering
University of Electronic Science and Technology of China

Abstract*—The accurate approximation of pixel value for preserving image details at a high concentration of noise has led the researchers to improve filters performance. A few image restoration filters are effective at lower density noise. Filters are commonly deployed for cameras, image processing tasks, medical image analysis, guided media data transmission, and real-time machine learning. This article proposes a mathematical model for the exact pixel value estimation at a high noise density for RGB and Gray images. The mathematical model is implemented to fuse statistical reasoning on the optimized mask sizes while preserving image details. Different parameter returns from the median filter, the trimmed median filter, the trimmed mean filter, and mood analysis form a mathematical function. The filter iteratively selects different schemes to calculate pixel values at different noise densities with minimum image information. Different processing masks are analyzed to preserve local data at specific image locations correctly in high density. A robust estimator counts false approximation of pixel values as discontinued, identified, and removed. At the post smoothening process, the filter recovers the misclassification of noise-free pixel and blur effects in the image. The qualitative experiments show satisfactory results in storing the details of the image from any image. The performance of the fusion filter is verified with visual quality and performance analysis matrices such as the image enhancement factor, the similarity indicator and the noise ratio from the peak signal.

Keywords*—*Salt and pepper noise; median filter; statistical reasoning; performance analysis matrices; high-density noise; mood; trimmed median; trimmed mean; peak signal-to-noise ratio; image enhancement factor; structural similarity index

I. Introduction

Some notable developments have attracted attention from the image processing industry to academia. Its research opportunities have been expanded due to the implementation of real-time applications in computer vision, digital image acquisition, satellite image analysis, data transmission and medical image analysis [1]. Preprocessing a digital image for a real-time application is essential before deploying for any application. This article focuses on significant preprocessing work and recommends a statistics-based mathematical model to eliminate digital images' salt and pepper noise (SAPN). The previously proven filters have been discussed here: standard median filter (SMF)[2], an adaptive median filter (AMF) [3], decision-based filter (DBA) [4], and noise-adaptive fuzzy switching median filter (NAFSM)[5]. The most common disadvantage of denoising is that the image information is lost in higher noise density [6].

SMF has a fundamental mathematical structure and displays the best output on low-density noise removal [7]. A mask with a 3X3 size limits the processing ability to select noise-free pixels at high densities. Nevertheless, several recent additions to nonlinear filters have outgrown existing filter errors and reached a stage where they may be suitable for approved types of noise. Nonlinear methods such as trimmed media filters, noise adaptive fuzzy median filters [5], decision-based median filters [4] have become the norm for image analysis.

Related research indicates that the development of the filtering phase is gradual, and the opportunity motivates the researcher to strengthen the filters based on statistical reasoning. This article investigates the fusion of statistical logic to design an effective filter to remove noise from the highly damaged image.

Filters are loosely classified according to their ability to reduce the noise of any concentration: when retrieved, images begin to have jitter effects, noise type, recovery of image data, image type, and noise removal from pure black and white dot [8]. The adaptive median filter mask shows a different technique for selecting a noise-free pixel by extending the size

*Corresponding Author

of 2×2 but fails due to limited statistical analysis and displacement of the denoised pixels at high noise densities. The adaptive median filter fails to store local information properly, while the noise adaptive fuzzy switching median filter [9](NAFSM)[10] and the decision-based filter (DBA) [4] repeatedly replace the noisy pixels that are resulting blurred effect on the restored images. However, the output of any filter can be improved to a certain extent by giving more weight to a particular pixel based on the surrounding pixel intensity statistics [11][12][9] [13].

Statistical analysis for the improvements of AMF, DBA, NAFSM, and other filters was initiated. However, a proper mathematical model is not yet proposed. Instead, most researchers have suggested multiple-combined filters to eliminate high-density pulse noise from the grey and colour images, regardless of the image's black and white part, discussed in Section 2.3. And most of them are not defined as state-of-the-art filters, such as morphological-mean filter [14], decision-based asymmetric filter (DBUF) [15] and filter boundary discriminative noise detection [16]. Most of the improved filters have the following drawbacks:

1) Most advanced/hybrid filters replace noise pixels under certain predefined conditions. Eliminating high SAPN concentration confuses filters to replace noisy pixels.

2) Most recently improved decision-based filters use threshold concepts and retain high noise density (above 50% noise density) in the restoration process. A striking effect is included separately in the restored image. In comparison, such filters do not perform well in extracting features from the medical image.

3) Most sophisticated filters and derivatives of median-type filters repeatedly replace the noise pixels in recovered images, regardless of the pixels in the 3×3 Mask.

A fusion-based statistical model performs iterative analysis on the different processing mask sizes for statistical analysis for image retrieval with improved visual quality from high-density SAPN. The filter iteratively selects the different mask sizes and performs comparative analysis based on neighbours pixels information and the non-local pixels information at 11X11 Mask to understand the image texture. The proposed filter has the following distinctive features to overcome the limitations:

A. Optimization of Texture-based Accurate Pixel Classification

Misclassification of noise pixels at high densities is a significant limitation of most filters. Most recently enhanced hybrid filters such as decision-based filters, noise-adapted obscure medium filters, interpolation-based filters [17], morphological filters [18], inpainting-based filters [19] to classify noisy pixels at high noise density.

B. Optimization of Mask Sizes to Fuse Statistical Analysis into the Mathematical Model

A filtering mask plays a vital role in removing high-density noise from the image regardless of local information at high density. The image texture should be maintained from any shape of the processing mask when retrieving an image from corruption, and the mathematical model is inspired by proper image texture management. A recommended mask size is preferred from the recently proposed hybrid filters discussed in Section 2.3. Experiments already have identified the use of a 3×3 mask as the best filtering option for removing low noise levels (<30 % noise) [20]. However, there are no optimal rules for determining the size of a mask that can be applied to a noise concentration. The algorithm provides a statistical, mathematical model that can repeatedly use different shapes of the Mask according to the noise density.

C. Statistical Reasoning to Identify Image Texture

The pixel intensity of the edges is identical to the details of the SAPN, which assumes it as the noise feature instead of evaluating edge information. The concept of axis based statistical analysis [6] is used here for edge detection based on the horizontal and vertical resolution in the Mask. The filter utilizes a primary mask at low density to identify the edges and noise-free pixels. But at high-density, different mask shapes are used repeatedly to reconstruct the edges, noise-free pixels, and non-local image details from corruption to image details.

D. Need for the Fusion of Statistical Model

Statistical analysis provides high visual quality and high efficiency of quantitative measurement with other popular filters. The quantitative measurements are the structural similarity index (SSIM) [21], the image enhancement factor (IEF) and the peak signal-to-noise ratio (PSNR) [22].

It provides a systematic decision rule for selecting processing mask size based on noise density. The rule-driven model allows multiple filters to work in an iterative process to estimate pixel intensity.

The above features distinguish from other existing filers in terms of retrieving an image from low and high noise concentrations. The model can be applied at the image preprocessing stage for image preparation for any application.

The rest of the paper is fourfold: It gives an overview of related literature, provides complete implementation methods of the proposed algorithm, performance analysis, and conclusion.

II. Literature Review

A. Salt and Pepper Noise (SAPN)

Reconstruction of digital images from high-density salt and pepper sounds is induced by transmissions, faulty sensors, or analogue-to-digital converters. Noise is evenly distributed in the image where infected pixels are usually identical to their neighbours. A fixed number of pixels that change to 0 or 255 is found in the salt and pepper noise (SAPN).

B. SAPN Detection Criteria for Existing and Hybrid Filters

Algorithms such as Noise Adaptive Fuzzy Switching Media Filters, Decision-Based Algorithms, and Adaptive Media Filters are standard today, often showing noise on the pixel intensity rather than identifying the properties of the surrounding pixels. Thus, filters select 0 or 255 noise-free volumes as noise pixels, contributing to incorrect detection in the image. It affects the decision on the image recovered from the high noise in a severe case. A noisy image has the following features:

1) SAPN is evenly or randomly scattered and spread over distorted images with equal probability.

2) Noisy pixels are not distinguished from noise-free 0 or 255. Again, the pixel intensity in that vicinity can be 0 (or 255) in most cases.

3) The SAPN often distorts the black and white information in the grey or RGB images. The salt or pepper noise is assimilated in the image white or black detail and lost, respectively.

Problems with the above design are considered when improving the filter. But the use of weak statistical methods or multiple filters consistently provides a certain level of performance. Advanced logic-based mathematical models or machine learning methods can overcome current performance limitations. Fig. 1 provides a classification of filters based on noise concentration.

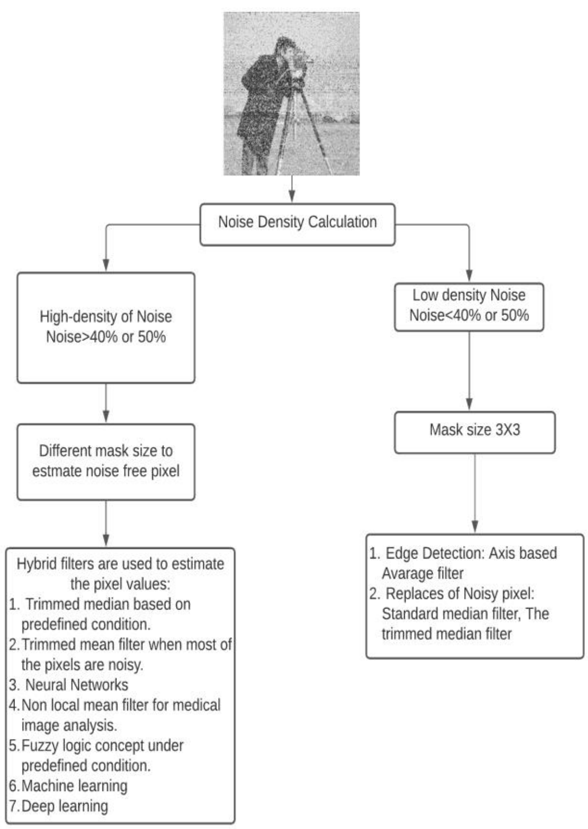


Fig. 1. Classification of Filters based on Noise Density.

C. Statistical or Mathematical Model used in Recently Published Journals

Recently published articles show that the most advanced filters are hybrid filter types, and a variety of filters act as a single algorithm applied to remove higher noise concentrations from images. The exact mathematical model for high-density pixel estimation has not yet been adequately defined. The statistical analysis of some recently published journals to identify trends in improving filters is summarized here in Table I. The table demonstrates the need to have an accurate mathematical model for eliminating noise at higher densities. Table I shows that filters for removing high-density noise are based on the systematic deployment of several filters in a single algorithm that improves the image quality at a certain level.

TABLE I. EXITING ALGORITHMS ACCURACIES

Filters	Statistical reasoning used for the filter design in recently published journals
The Axis-Based Filters[6]	The edges of the Mask are described with the aid of using a direct line passing via the primary pixel. Otherwise, its miles changed with the aid of using the expected value.
DBUTM Filter [23]	The mean of all four creaking neighbours replaces the noisy pels. If four neighbours aren't noisy, the corrupted pixel returns the unsymmetrical cut changed median value.
A CNN filter [24]	CNN model is used for denoising Salt and Pepper noise.
Modifed cascaded flter[8]	The trimmed median value replaces the noisy pixel while other pixel values, 0 and 255, are present at a high noise density in the selected window.
Improved Switching Median Filter [25]	The median value selection is based on the 2nd and 8th pixels of the 3 × 3, where all the pixels are arranged in a matrix in ascending order. Also, the window means is calculated from a 4X4 window.
Minimum-maximum median filter[26]	Selecting the mask size is entirely based on the noise density of the current processing window. Noisy pixel is replaced by the last processing pixel or median value.
Min-Max Filter[27]	Min-max average pooling is used here to remove SAPN noise.
Neural network[28]	The proposed algorithm utilizes a denoising convolutional neural network at a high noise density.
Tropical algebra-based adaptive filter[29]	The filter deployed tropical algebra to illustrate the adaptive principles to do away with salt and pepper noise.
A two-stage filter[30]	The filter behaves like a conventional filter; however, the high-density noisy pixel is changed using the median of the most repetitive pixels.
FFDNet: CNN based Image Denoising [34]	Proposed a CNN model.
Adaptive Algorithm and Wavelet Transform [33]	Utilizes Adaptive filering and Wavelet Transform

From Table I above, a summary of the most proposed hybrid filters between 2019 and 2021 can be summarized as follows:

1) The approximation of pixel intensity at high density is based on finding the noise-free pixel regardless of maintaining local features.

2) Install filters without comparing local information to different image properties at high density.

3) Sometimes pixel estimation relies on threshold value to estimate noisy pixels as noise-free.

4) Discontinuity of local pixels is not yet maintained at high density, which causes a blur effect in the image.

III. Methodology

In this section, a fusion of statistically based filters is presented, broadly divided into four subdivisions: the estimation of noise density, pre-edge filtering, and statistical reasoning based on ambient pixel intensity (the main algorithm of the filter), and post-smoothing filter (Robust Statistics). The block diagram of the proposed filter is demonstrated in Fig. 2.

A. *Estimation of the Noise Density*

The noise concentration of the processing mask is calculated to start the filtering process because it is an integral part of statistical analysis. Sound density (n_d)can be calculated using equations (1) and (2) [31].

$$Y_{(x,y)} = \begin{cases} 1, & q_{x,y} \in \{0,255\} \\ 0, & \text{Otherwise} \end{cases} \tag{1}$$

$$N_d = \frac{\sum_{y=1}^{x=R}\sum_{y=1}^{x=C} Y_{x,y}}{R \times C} \times 100\% \tag{2}$$

Where, $q_{x,y}$ is the pixels of the noisy image, and R and C are the image dimension.

B. *Filtration of the Edges of the Image Objects*

At first, a 3X3 mask was taken to classify the contaminated pixels based on the horizontal and diagonal directions in the initial stage. If the pair of pixels in the horizontal or diagonal direction is the same, it is declared as the underlying processing pixel, $q_{x,y}$ Noise-free. The matching figure is shown in equation 3 below:

$$q(x,y) = \begin{cases} q_{x+1,y+1}, & \text{if, } q_{x+1,y+1} = q_{x-1,y-1} \\ q_{x+1,y-1} & \text{if, } q_{x+1,y-1} = q_{x-1,y+1} \\ q_{x,y+1,} & \text{if, } q_{x,y+1} = q_{x,y-1} \\ q_{x,y} & \text{Trimmed Median} \end{cases} \tag{3}$$

It can be seen that the original values replace most of the pixels recovered in this process, and here the incorrect classification of pixels is reduced in a high density of noise.

This technique is considered when enough noise-free pixels are available to detect edges in the processing mask. Otherwise, the trimmed medium filter is placed in the processing mask. The block diagram of the vertical and horizontal edge detection scheme is given in Fig. 3.

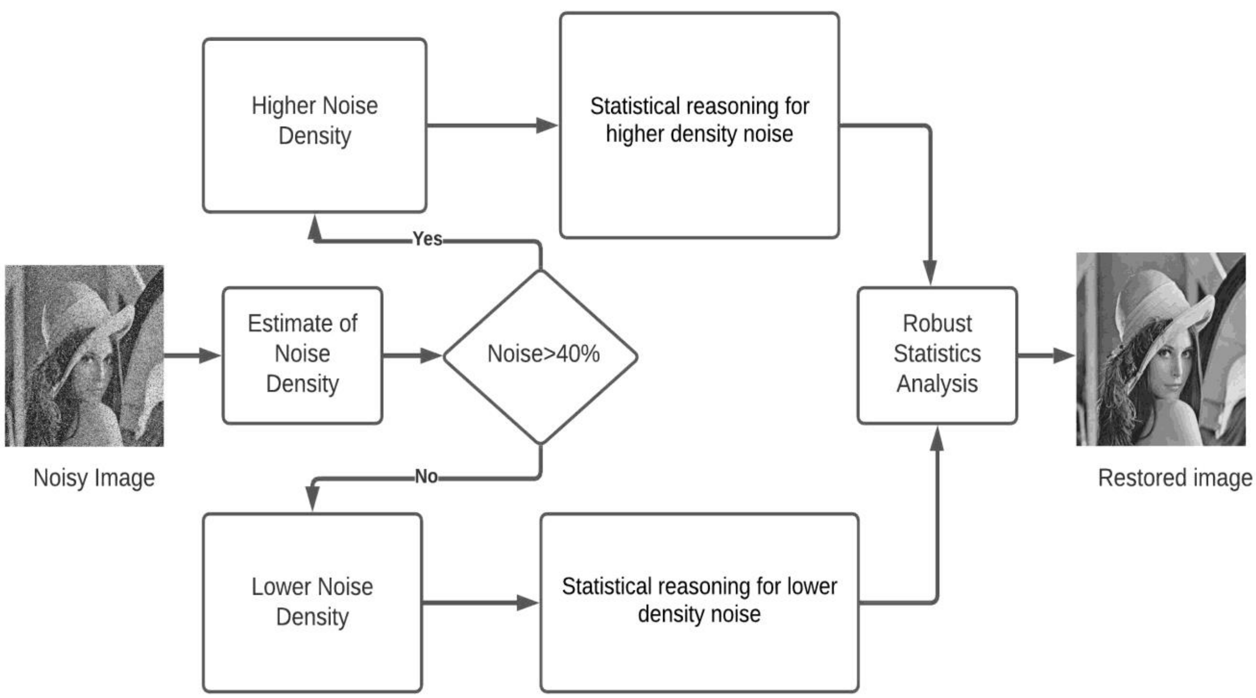


Fig. 2. Block Diagram of the Proposed Filter.

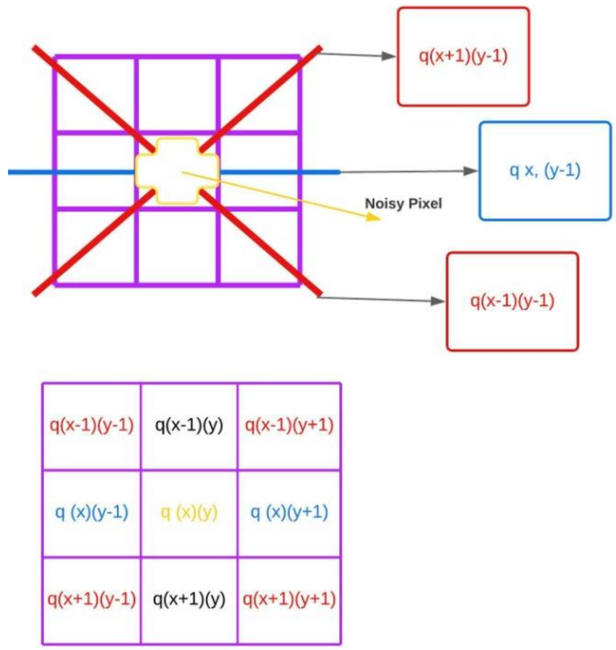


Fig. 3. Filtration of the Edges from the Image Objects.

C. Statistical Reasoning based on Ambient Pixel Intensity

At higher noise concentrations, the filter utilizes 3X3, 9X9, and 11X11 masks, and at the lower noise <40% density, only a 3X3 mask is applicable to perform edge detection and noise elimination. The trimmed median value shows the best result after detecting edges at lower noise density. Mood value is estimated from the 11X11 Mask for the pure image black and white data. Statistical analysis from 3X3 and 9X9 masks offer a function to calculate the pixel intensity of distorted images as given in an equation (4).

$$\mathrm{U_{x,y}} = \left(\propto\times\ \mathrm{q_{x-1,y-1}}\right) + (1.0-\propto)\mathrm{SMF}(11\times 11) \tag{4}$$

Here,

$\propto= (\frac{mood\,(3\times 3\ mask)}{9})$,

$1 - q_{x,y} = \ Last\ processed\ pixel$,

$SMF(11\times 11) = Median\ value\ of\ 9\times 9$ Mask.

If the last processed pixel is, $q_{x-1,y-1} =\ \ 0\ or\ 255$, and all the pixels at 11×11 Mask is $0\ or\ 255$, then the filter detects this image detail is pure black or white. Then the complete statistical equation is shown in (5).

$$\mathrm{q_{x,y}} = \begin{cases}\mathrm{U_{x,y}},\ \text{if } 1-\mathrm{q_{x,y}} \neq 0 \text{ or } 255,\ \ \text{Noise} > 40\% \\ \text{Mood}(9\times 9),\ \ \text{Otherwise}\end{cases} \tag{5}$$

Tests show that PSNR values drop sharply when $N_d >$ 40%. Statistical analysis of decision-making is based on the SAPN density of the Mask is considered here. Fig. 4 shows the different mask size sizes required to estimate intensity value. The proposed filter iteratively use the filter and perform median, mood and probability analysis on image data.

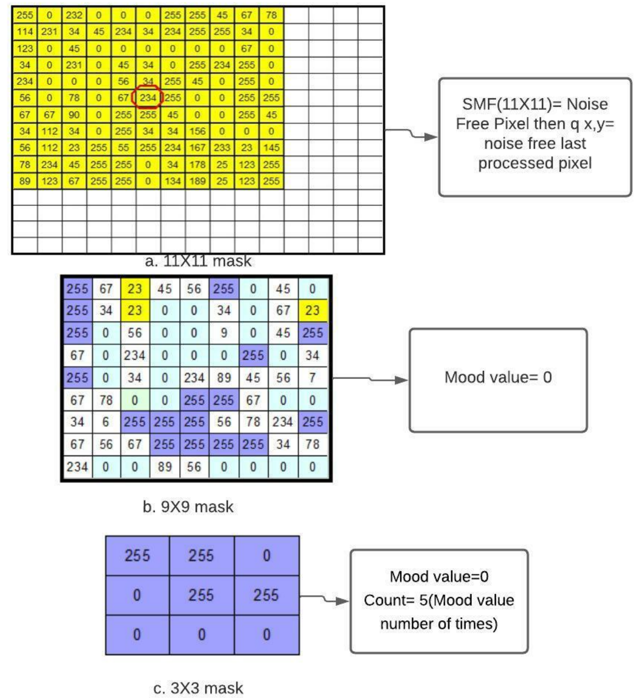


Fig. 4. Comparison Analysis on different Mask at Higher Density.

The need to select both masks of 9X9 and 11X11 is shown in a, b, and c of Fig. 5. The size of the Mask above 11X11 could not properly store local information. Thus, the statistical analysis is entirely dependent on 3×3, 9×9, and 11×11 masks. Experiment shows that, filter with size more than11×11 could not maintain the local image information at high density of noise. Fig. 5 visually demonstrates the trimmed median and the mood at e and f.

1) Post-smoothing filter: A smooth step has been added to the algorithm to improve the image quality further. The Robust Estimation is based on the principle that safety is more important than efficiency[32]. Here, the median is an estimator. Let q^1, q^2, q^3 ----- q^n indicate a random pixel from that processing image.

Algorithm 1(Removal of higher density SAPN):

$q_{x,y} \leftarrow Input\ \ Noisy\ image$
    $if\ \ \mathrm{N_d} < 40\ \&\&\ M_{3\times 3}$
  { $if\ \ Q_{xy} \leftarrow q(x,y)$ // Filtration of the edges
              else
        $Q_{xy} \leftarrow\ \mathrm{M_t}\,(\mathrm{q_{x,y}})$ // Trimmed Median
              }
elseif $\mathrm{N_d} > 40\ \ \&\&\ \ (1-\mathrm{q_{x,y}}) \neq 0\ ||\ 255\ \ \&\&\ M_{9\times 9}$
        $Q_{xy}\ \ \leftarrow \mathrm{U_{x,y}}$
  else $\mathrm{N_d} = 100\ \ \&\&\ \ M_{9\times 9}$
        $Q_{xy} \leftarrow\ \ Mood(\mathrm{q_{x,y}})$
Algorithm2 $\leftarrow Q_{xy}$

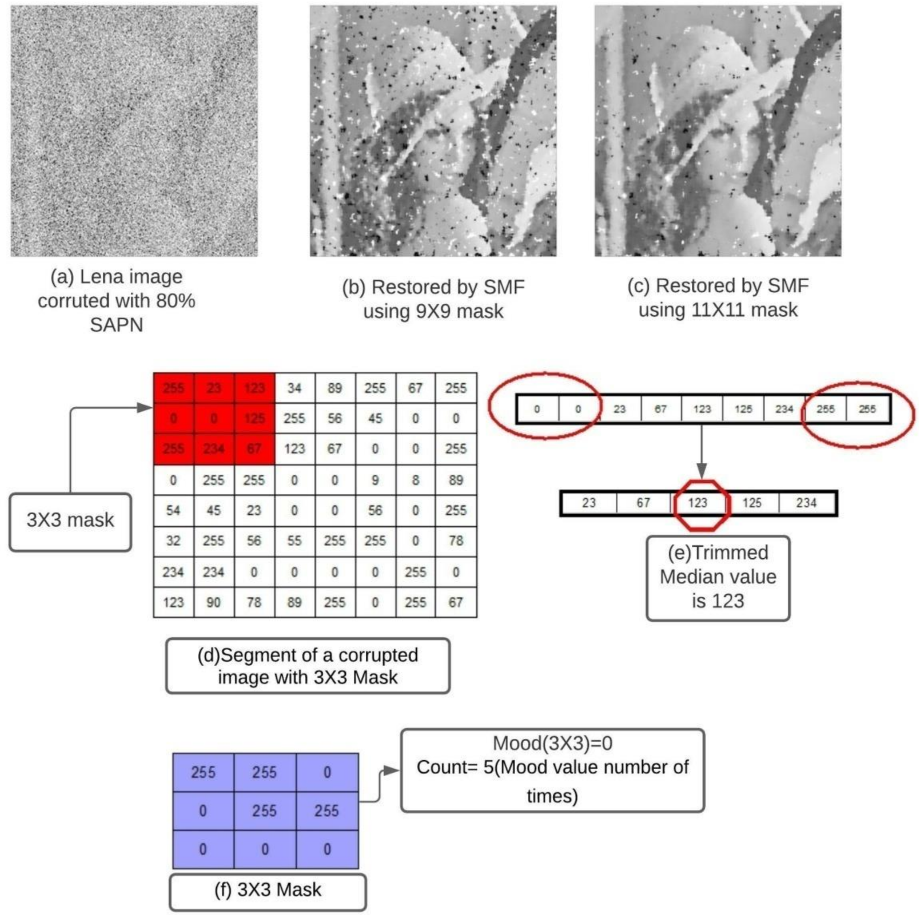


Fig. 5. Statistical Reasoning: (a) Lena Image Corrupted with 80% SAPN (b) Restored by SMF using 9 × 9 Mask (c) Restored by SMF using 11 × 11 Mask (d) the Segment of the Corrupted Image (e) the Trimmed Median (f) Mood Value.

U1 is the minor key of Qi, U2 is the following key in the sequence of Qi, and Un is the largest in the list (ascending order). The estimation of the new median, which removes pixel discontinuity in the image, is based on the L1 norm or Laplacian Distribution. The influence function is constant for the median, making the median an excellent and robust estimator. Robust Estimator is the M-estimators that are pretty good for distribution with outliers (scatter data value). The effects of robust estimators are compared in terms of influence function in the equation (6).

$$\varphi(q) = \begin{cases} 1 & q > 0 \\ -1 & q < 0 \end{cases} \tag{6}$$

Lorentz has an influence function of an estimator that tends to zero to increase the estimation distance and maximum value for the breakdown, which is used to the original pixel value from a corrupted image. The equations are given in (7) and (8).

$$P(q) = log\frac{(1+x^2)}{2\sigma^2} \tag{7}$$

$$\varphi = \frac{2q}{2\sigma^2+q} \tag{8}$$

Robust estimation is applied to minimize pixel discontinuity of the corrupted image, which eliminates blur effects from the image. The SAPN is first identified based on this method's minimum (0) and maximum (255) values. It is considered noise-free if the current pixel is within the dynamic range 0 and 255. Otherwise, it is seen as a noise pixel and is replaced by a value determined by the following algorithm using the Lorentzian Estimator.

Algorithm 2(Robust Statistics):

1: $Q_{x,y} \leftarrow Input\ preprocess\ image\ layer$
2: $if\ Q_{med} \leftarrow SAPN\ and\ Mask = 3 \times 3$
3: $P_{x,y} \leftarrow Q_{med} - Q_{xy}$
4: $if\ P_{x,y} \leftarrow SAPN$
//calculate the robust influence function
5: $\varphi(p) \leftarrow \frac{2p}{2\sigma^2+p^2}$
6: $s1 \leftarrow \sum_{l\in L} \frac{Pixel(l)\times\varphi(x)}{p}$
7: $s2 \leftarrow \sum_{l\in L} \frac{\varphi(x)}{p}$
//Estimated pixel value
8; $Q_{x,y} \leftarrow s1/s2;$
9: $R_{x,y} \leftarrow Restored\ image$

IV. Results

The statistics model and popular filters are programmed in MATLAB 2014 and then simulated with the 14 benchmark images used in related journals to study filter performance. The evaluation task is to calculate a denoised image with a noise density in the medium to high range (40% to 90%). The denoised images are judged in terms of visual quality and performance analysis matrices.

A. Evaluating by Quantitative Comparison

The visual reports are taken from the simulation to judge the significance of the proposed filter in terms of visual quality. The famous benchmark data for image denoising are Fishing Boat, Baboon, Pepper, Barbara, Bubble, and Camera Man. The visual quality shows that SAPN can be extracted from any digital image by the proposed filter.

Fig. 6 to 11 visually compare the proposed filter and other existing filters. Benchmark datasets are corrupted with SAPN at different levels of noise concentration, and the same datasets are restored from the stated filters. Recovered images are visually compared, and it is seen that the proposed filter has the best performance in creating high-quality images. As in the case of Fig. 6 and 7, the images are restored from 70% and 90% of the SAPN where the images are visually compared to the proposed filter outperforms the other filters. The optimal suppression power at all levels of noise concentration is shown in Fig. 8 and 9. Furthermore, Fig. 10 shows that 50% of bubbles have been recovered from noise.

The proposed filter cannot correctly restore the image because some parts have too rough transitions. It should be noted that most sophisticated filters, such as SMF, AMF, DBA, and MDBUTMF, are weak in effectively detecting noise pixels when the noise pollution in the image is above 40%. The visual comparison shows that they often fail to classify the pixel as noise-free at higher noise density, causing a blurring effect on images. In this case, the proposed filter applied post smoothing statistical analysis to remove image artefacts and blurring effects. The proposed filter did not perform well in the bubble image of Fig. 8 but shows excellent performance in removing noise from the cameraman's images in Fig. 11. Fig. 6 to 11 visually compare the proposed filter and other existing filters. Benchmark datasets are corrupted with SAPN at different levels of noise concentration, and the same datasets are restored from the stated filters. Recovered images are visually compared, and it is seen that the proposed filter has the best performance in creating high-quality images. As in the case of Fig. 6 and 7, the images are restored from 70% and 90% of the SAPN where the images are visually compared to the proposed filter outperforms the other filters. The optimal suppression power at all levels of noise concentration is shown in Fig. 8 and 9. Furthermore, Fig. 10 shows that 50% of bubbles have been recovered from noise. The proposed filter cannot correctly restore the image because some parts have too rough transitions.

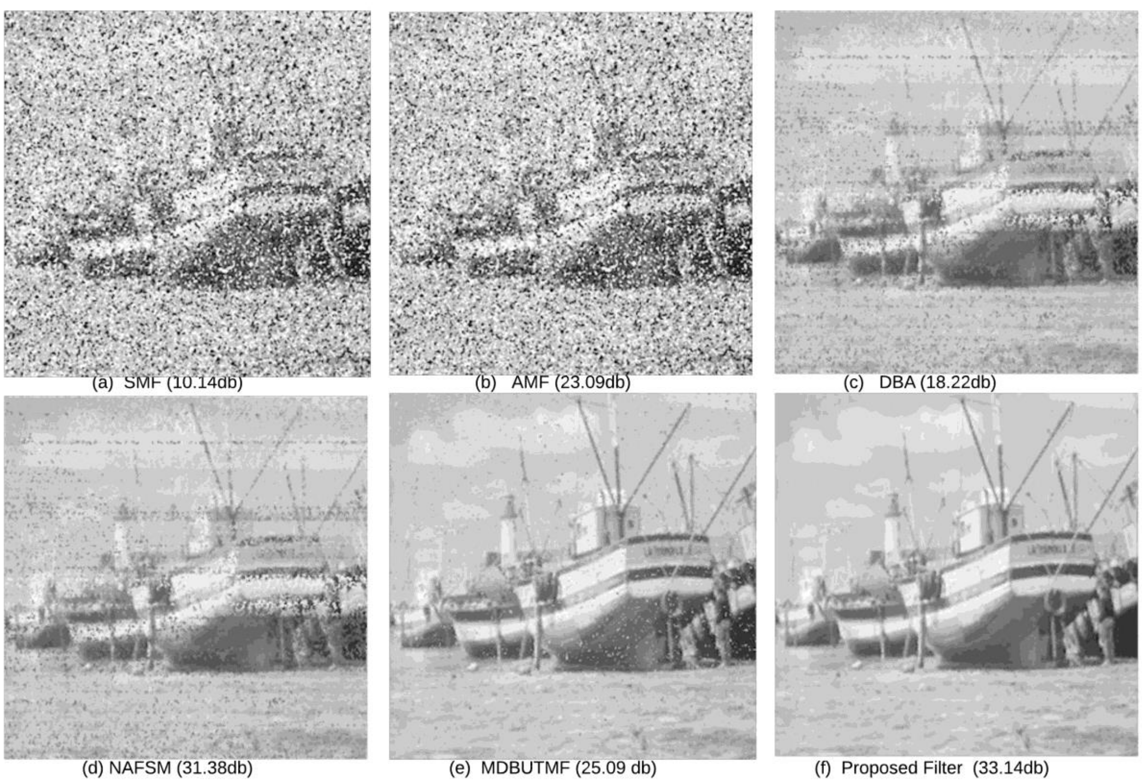


Fig. 6. (From Upper Left) a Fishing Boat Image Dnoised from 70% SAPN using ((a) SMF, (b) AMF, (c) DBA, (d) NAFSM Filter, (e) MDBUTMF and (f) PA.

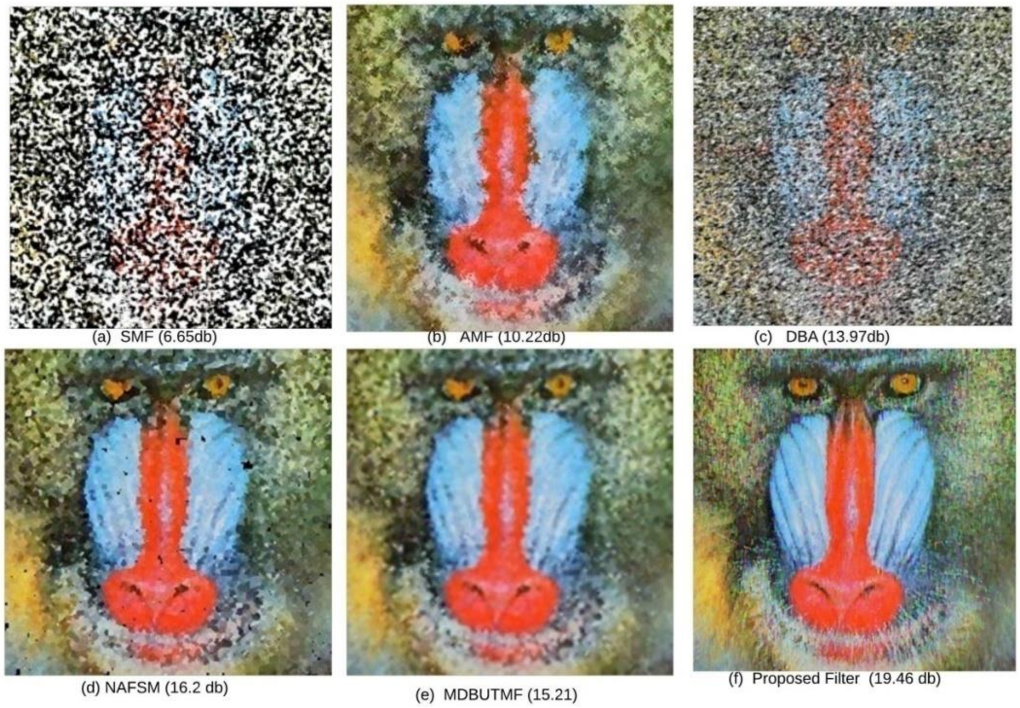


Fig. 7. (From Upper Left) Baboon Image Restored from 90% SAPN using ((a) SMF, (b) AMF, (c) DBA, (d) NAFSM Filter, (e) MDBUTMF and (f) PA.

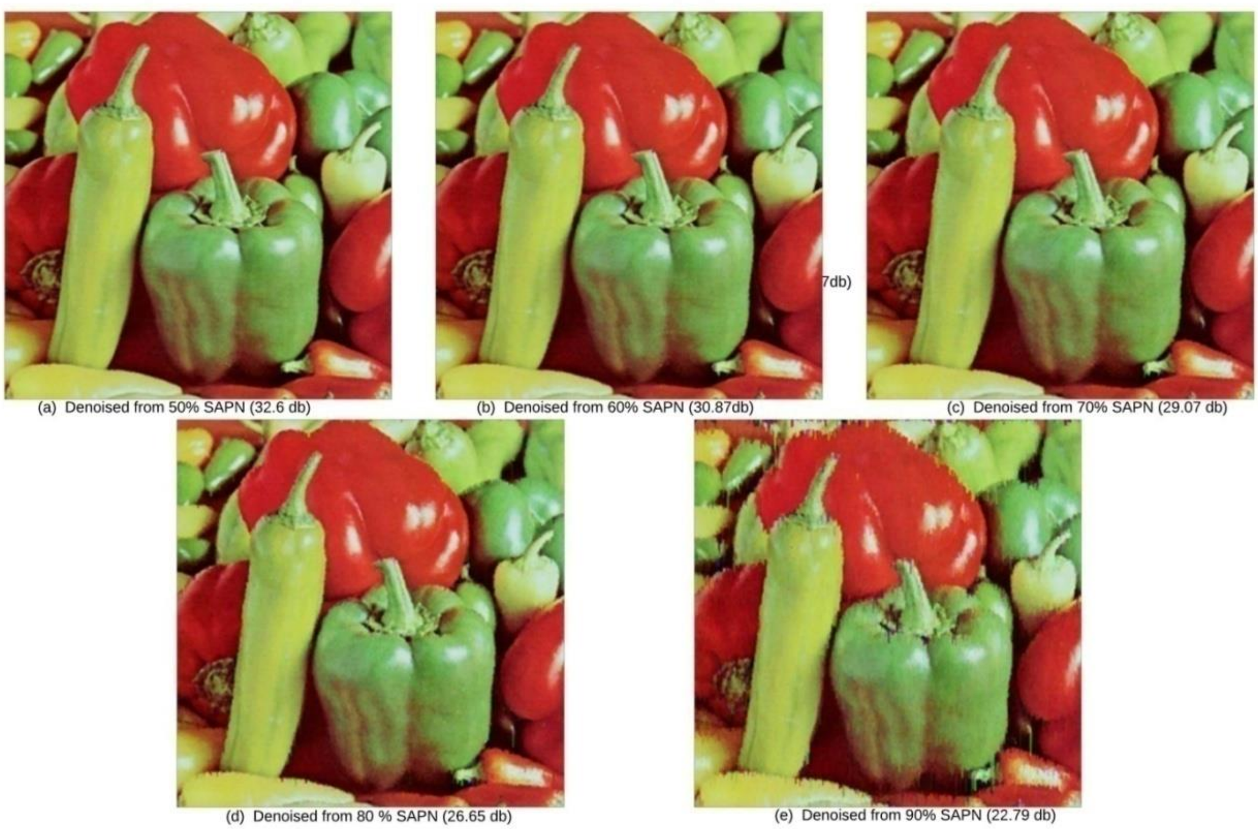


Fig. 8. (From Upper Left) Pepper Image Corrupted with 50% to 90% SAPN and Restored with PA.

Fig. 9. (From Upper Left) Barbara Image Corrupted with 50% to 90% SAPN and Restored with PA.

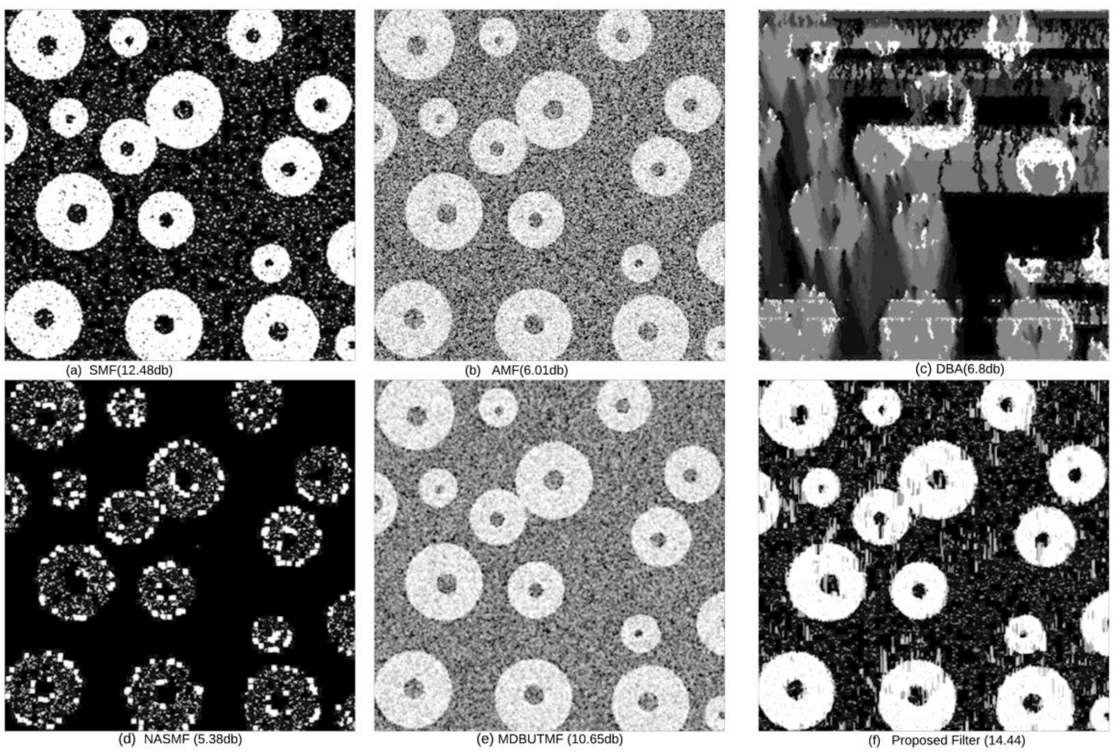


Fig. 10. (From Upper Left) Bubble Image Restored from 50% SAPN using (a) SMF, (b) AMF, (c) DBA, (d) NAFSM Filter, (e) MDBUTMF and (f) PA.

Fig. 11. (From Upper Left): Cameraman Image Denoised from 80% SAPN using (a) SMF, (b) AMF, (c) DBA (d) NAFSM Filter, (e) MDBUTMF, and (f) PA.

B. *Evaluating by Quantitative Comparison*

In this article, the Peak Signal-to-Noise Ratio (PSNR), Mean Square Error (MSE), Image Enhancement Factor (IEF), and Structural Similarity Index (SSIM) are considered for quantitative analysis. The PSNR is defined as the ratio between the original image's strength and the denoised image's strength, and the mathematical equation is given in (9).

$$\text{Peak Signal to Noise Ratio } = 10\log_{10}\frac{(MAX \times MAX)}{MSE} \tag{9}$$

Here , MAX is the maximum gray level of the image, and the Mean Square Error(MSE) is defined in the equation (10).

$$\text{Mean Square Error } = \frac{1}{RC}\sum(P_{x,y} - Q_{x,y})^2 \tag{10}$$

The R and C is the dimension of the image, $Q_{x,y}$ – original image and the $P_{x,y}$ - restored image. Another parameter is used, the Image Enhancement Factor (IEF). The mathematical expression of IEF is given in the equation (11).

$$\text{Image Enhancement Factor } = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(N_{ij} - R_{ij})^2}{\sum_{i=1}^{M}\sum_{j=1}^{N}(X_{ij} - R_{ij})^2} \tag{11}$$

Where R_{ij} , X and N_{ij} denote the original image, denoised image, and corrupted image, respectively. The final parameter used to evaluate the proposed filter's performance is Structural Similarity Index. The mathematical expression is in an equation (12).

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{x,y} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{12}$$

Here, μ_x and μ_y are the standard deviation used as an estimation of the signal contrast, σ_x and σ_y are the contrast comparison, and C1 and C2 are the constant value with a limit of 1.

C. *Experimental Results of Tabular Style for High Density of Noise for Baboon(Grey) and Lena(RGB) Images*

Tables II and III summarize the IEF, PSNR, and SSIM values for the six filters in the Baboon Grayscale and Lena RGB images. The experimental clearly show that the proposed algorithm has the edge over reported filters. Sometimes the Proposed filter cannot deliver optimal results in specific embodiments if certain image parts have very rough transitions or an utterly smooth area. Nevertheless, the average improvement of the proposed filter is the best from existing filters at SAPN density varies from 50% to 90%.

This, the performance of PSNR is exceptionally good for the noise concentration > 70%. Even at SAPN concentrations up to 95%, the proposed algorithm provides a fair average value compared with other existing filters. At medium noise concentrations, the effectiveness of NAFSM manages to surpass the Proposed algorithm but still exceeds existing strategies.

TABLE II. IEF, PSNR AND SSIM RESULTS ON THE BABOON.JPG FOR MEDIUM TO HIGH DENSITY

QA	Noise density	SMF	AMF	DBA	NAFSM	MDBUTMF	PA
IEF	50	3.96	23.96	12.49	30.94	24.19	39.33
	60	2.67	20.69	10.99	28.14	22.36	35.53
	70	1.9	15.32	9.34	25.16	18.57	30.55
	80	1.45	6.86	8.12	22.14	13.5	25.04
	90	1.18	2.68	6.36	16.81	8.44	19
PSNR	50	14.47	22.3	19.47	23.41	22.34	25.04
	60	11.96	20.85	18.1	22.19	21.19	23.89
	70	9.83	18.89	16.74	21.04	19.72	22.61
	80	8.05	14.81	15.54	19.9	17.75	21.19
	90	6.65	10.22	13.97	18.2	15.21	19.46
SSIM	50	0.21	0.69	0.49	0.74	0.69	0.86
	60	0.11	0.59	0.38	0.66	0.6	0.79
	70	0.06	0.46	0.28	0.57	0.48	0.69
	80	0.03	0.24	0.18	0.46	0.32	0.56
	90	0.01	0.06	0.09	0.29	0.16	0.39

TABLE III. IEF, PSNR AND SSIM RESULTS ON THE LENA.JPG FOR MEDIUM TO HIGH DENSITY

QA	Noise density	SMF	AMF	DBA	NAFSM	MDBUTMF	PA
IEF	50	4.9	145.35	18.63	261.33	138.33	387.83
	60	2.99	95.82	14.03	210.52	95.48	288.92
	70	2.02	34.39	10.98	176.14	50.04	224.04
	80	1.49	9.9	8.3	128.92	21.7	147.05
	90	1.19	2.94	6.18	62.23	9.14	79.96
PSNR	50	15.25	29.97	21.05	32.52	29.76	33.41
	60	12.31	27.37	19.02	30.79	27.35	31.78
	70	9.92	22.24	17.28	29.34	23.87	30.14
	80	8.02	16.25	15.49	27.4	19.66	27.79
	90	6.52	10.46	13.68	23.72	15.39	24.22
SSIM	50	0.16	0.75	0.4	0.8	0.72	0.92
	60	0.07	0.66	0.3	0.74	0.61	0.86
	70	0.03	0.49	0.22	0.66	0.43	0.78
	80	0.02	0.24	0.15	0.56	0.23	0.68
	90	0.01	0.04	0.08	0.39	0.09	0.51

D. Experimental Results of Graphical Style for Low Noise Concentration for Pepper Image

The performance of the proposed algorithm is tested with a pepper image, and the noise density varies from 10% to 40%. Restoration performance is quantitatively measured by metrics such as IEF, PSNR and SSIM, as shown in Fig. 12, 13 and 14. The SMF replaces the processing pixel by its median regardless of the type of pixel, which provides poor performance. However, the AMF [8] shows improved performance, but the loss of local image details is much more significant due to its adaptive nature. Predefined condition-based filters such as NAFSM and decision-based filters failed to retrieve images correctly.

The trimmed median filter always considers median values from the processing mask eliminating 0 and 255, which failed to store image edges and local information properly. Necessary features that may supressed noise for filters should also be taken care of for image processing tasks. The mathematical model also focused on this described in low density. Tests show that the exact pixel estimate of the proposed filters in low-density noise is the highest of the popular filters.

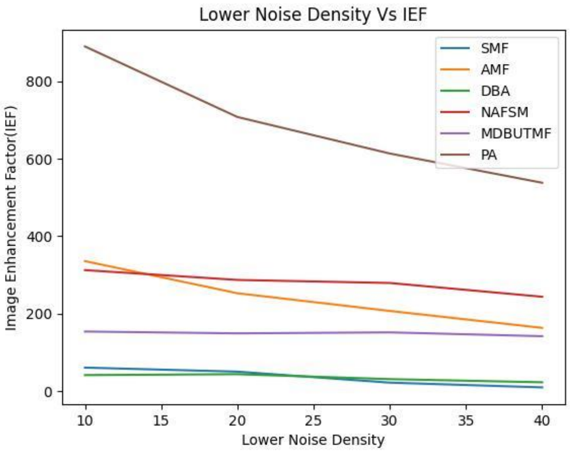


Fig. 12. Quantitative Analysis of Filters using IEF for Pepper Image.

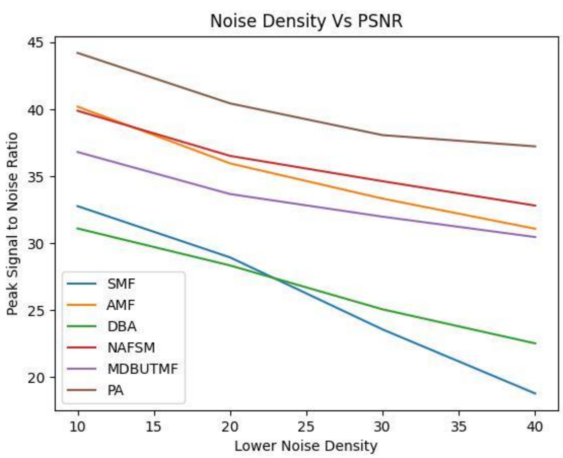


Fig. 13. Quantitative Analysis of Filters using PSNR for Pepper Image.

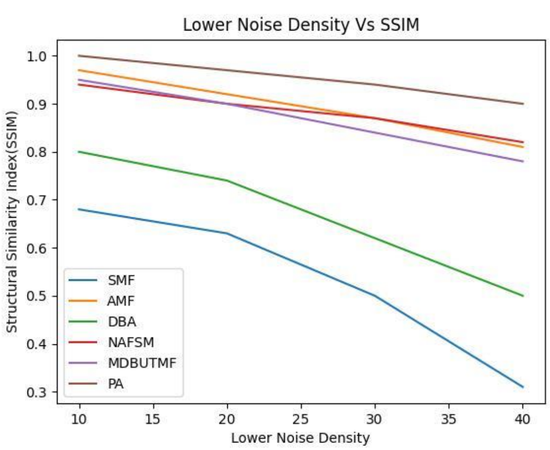


Fig. 14. Quantitative Analysis of Filters using SSIM for Pepper Image.

V. Conclusion

In this article, a complete and relative study of some sophisticated filters evaluates their performance in high-density noise removal. A fusion filter has been developed to deal with high-density noise. Statistical analysis was conducted on high-density noise, pixel estimation, data saving, white and black data identification of images, and removal of isolated pixels for data discontinuity without normalizing the image intensity. The proposed model provides the best results for noise removal at any concentration level. The main drawback of this filter is that the streaking effect occurs only when the high concentration of noise is removed from the pure black and white or medical images. Incorporating non-local statistical concepts into the proposed mathematical model could eliminate the streaking effect when removing high-density noise from medical images. The proposed algorithm can be placed on any digital devices as an image preprocessor.

References

[1] H. Liang, N. Li, and S. Zhao, "Salt and pepper noise removal method based on a detail-aware filter," Symmetry (Basel)., vol. 13, no. 3, 2021, doi: 10.3390/sym13030515.

[2] R. C. González, R. E. Woods, and S. L. Eddins, Digital Image Processing Using MARLAB. Pearson, 2004.

[3] H. Hwang and R. A. Haddad, "Adaptive median filters: new algorithms and results," IEEE Trans. image Process., vol. 4, no. 4, pp. 499–502, 1995.

[4] K. S. Srinivasan and D. Ebenezer, "A new fast and efficient decision-based algorithm for removal of high-density impulse noises," IEEE Signal Process. Lett., vol. 14, no. 3, pp. 189–192, 2007.

[5] K. K. V. Toh and N. A. M. Isa, "Noise adaptive fuzzy switching median filter for salt-and-pepper noise reduction," IEEE Signal Process. Lett., vol. 17, no. 3, pp. 281–284, 2009.

[6] A. Kundu, S. Banerje, C. Sarkar, and S. Barman, "An Axis Based Mean Filter for Removing High-Intensity Salt and Pepper Noise," 2020 IEEE Calcutta Conf. CALCON 2020 - Proc., vol. 1, pp. 363–367, 2020, doi: 10.1109/CALCON49167.2020.9106561.

[7] Z. Shen, J. Ni, and C. Chen, "Blind detection of median filtering using linear and nonlinear descriptors," Multimed. Tools Appl., vol. 75, no. 4, pp. 2327–2346, 2016.

[8] B. Karthik, T. Krishna Kumar, S. P. Vijayaragavan, and M. Sriram, "Removal of high density salt and pepper noise in color image through modified cascaded filter," J. Ambient Intell. Humaniz. Comput., vol. 12, no. 3, pp. 3901–3908, 2021, doi: 10.1007/s12652-020-01737-1.

[9] K. Ramamoorthy, T. Chelladurai, and P. N. Sundararajan, "Edge analysis for noise suppression in ultrasound kidney images using weighted median filter," Int. J. Syst. Signal Control Eng. Appl., vol. 7, no. 2, pp. 36–42, 2014.

[10] V. Jayaraj and D. Ebenezer, "A new switching-based median filtering scheme and algorithm for removal of high-density salt and pepper noise in images," EURASIP J. Adv. Signal Process., vol. 2010, no. 1, p. 690218, 2010.

[11] G. Balasubramanian, A. Chilambuchelvan, S. Vijayan, and G. Gowrison, "An extremely fast adaptive high-performance filter to remove salt and pepper noise using overlapping medians in images," Imaging Sci. J., vol. 64, no. 5, pp. 241–252, 2016.

[12] N. Singh, T. Thilagavathy, R. T. Lakshmipriya, and O. Umamaheswari, "Some studies on detection and filtering algorithms for the removal of random valued impulse noise," IET Image Process., vol. 11, no. 11, pp. 953–963, 2017.

[13] H. C. Bandala-Hernandez et al., "Weighted median filters: An analog implementation," Integration, vol. 55, pp. 227–231, 2016.

[14] R. Kunsoth and M. Biswas, "Modified decision based median filter for impulse noise removal," in 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 2016, pp. 1316–1319.

[15] D. Shekar and R. Srikanth, "Removal of High Density Salt \& Pepper Noise in Noisy Images Using Decision Based UnSymmetric Trimmed Median Filter (DBUTM)," Int. J. Comput. Trends Technol., vol. 2, no. 1, pp. 109–114, 2011.

[16] P.-E. Ng and K.-K. Ma, "A switching median filter with boundary discriminative noise detection for extremely corrupted images," IEEE Trans. image Process., vol. 15, no. 6, pp. 1506–1516, 2006.

[17] C. Kalyoncu, Ö. Toygar, and H. Demirel, "Interpolation-based impulse noise removal," IET Image Process., vol. 7, no. 8, pp. 777–785, 2013.

[18] M. Gonzalez-Hidalgo, S. Massanet, A. Mir, and D. Ruiz-Aguilera, "Improving salt and pepper noise removal using a fuzzy mathematical morphology-based filter," Appl. Soft Comput., vol. 63, pp. 167–180, 2018.

[19] K. Gupta, N. Goyal, and H. Khatter, "Optimal reduction of noise in image processing using collaborative inpainting filtering with Pillar K-Mean clustering," Imaging Sci. J., vol. 67, no. 2, pp. 100–114, 2019, doi: 10.1080/13682199.2018.1560958.

[20] D. Guo, Z. Tu, J. Wang, M. Xiao, X. Du, and X. Qu, "Salt and Pepper Noise Removal with Multi-Class Dictionary Learning and L0 Norm Regularizations," Algorithms, vol. 12, no. 1, p. 7, 2019.

[21] U. Sara, M. Akter, and M. S. Uddin, "Image quality assessment through FSIM, SSIM, MSE and PSNR—A comparative study," J. Comput. Commun., vol. 7, no. 3, pp. 8–18, 2019.

[22] A. Tanchenko, "Visual-PSNR measure of image quality," J. Vis. Commun. Image Represent., vol. 25, no. 5, pp. 874–878, 2014.

[23] V. kishorebabu and R. Varatharajan, "A decision based unsymmetrical trimmed modified winsorized variants for the removal of high density salt and pepper noise in images and videos," Comput. Commun., vol. 154, pp. 433–441, 2020, doi: 10.1016/j.comcom.2020.02.048.

[24] B. Fu, X. Zhao, Y. Li, X. Wang, and Y. Ren, "A convolutional neural networks denoising approach for salt and pepper noise," Multimed. Tools Appl., vol. 78, no. 21, pp. 30707–30721, 2019, doi: 10.1007/s11042-018-6521-4.

[25] S. Anwar and G. Rajamohan, "Improved Image Enhancement Algorithms based on the Switching Median Filtering Technique," Arab. J. Sci. Eng., vol. 45, no. 12, pp. 11103–11114, 2020, doi: 10.1007/s13369-020-04983-9.

[26] B. Garg, "An adaptive minimum-maximum value-based weighted median filter for removing high density salt and pepper noise in medical images," Int. J. Ad Hoc Ubiquitous Comput., vol. 35, no. 2, pp. 96–116, 2020, doi: 10.1504/IJAHUC.2020.109795.

[27] P. Satti, N. Sharma, and B. Garg, "Min-Max Average Pooling Based Filter for Impulse Noise Removal," IEEE Signal Process. Lett., vol. 27, pp. 1475–1479, 2020, doi: 10.1109/LSP.2020.3016868.

[28] A. Noor, Y. Zhao, R. Khan, L. Wu, and F. Y. O. Abdalla, "Median filters combined with denoising convolutional neural network for Gaussian and impulse noises," Multimed. Tools Appl., vol. 79, no. 25–26, pp. 18553–18568, 2020, doi: 10.1007/s11042-020-08657-4.

[29] A. Abdurrazzaq, I. Mohd, A. K. Junoh, and Z. Yahya, "Tropical algebra based adaptive filter for noise removal in digital image," Multimed. Tools Appl., vol. 79, no. 27–28, pp. 19659–19668, 2020, doi: 10.1007/s11042-020-08847-0.

[30] D. N. H. Thanh, N. H. Hai, V. B. S. Prasath, L. M. Hieu, and J. M. R. S. Tavares, "A two-stage filter for high density salt and pepper denoising," Multimed. Tools Appl., vol. 79, no. 29–30, pp. 21013–21035, 2020, doi: 10.1007/s11042-020-08887-6.

[31] N. Sharma, P. J. S. Sohi, B. Garg, and K. V. Arya, "A novel multilayer decision based iterative filter for removal of salt and pepper noise," Multimed. Tools Appl., vol. 80, no. 17, pp. 26531–26545, 2021, doi: 10.1007/s11042-021-10958-1.

[32] M. B. Ebrahimi, Saviz and Setoudeh, Farbod and Tavakoli, "A New hybrid Method for Noise Robust Estimation of Image Fractal Dimension," Majlesi J. Electr. Eng., vol. 14, no. 2, pp. 25--34, 2020.

[33] Jani, Md & Islam, Md. (2018). De-noising and Feature Extraction of ECG and EEG Signal Using Adaptive Algorithm and Wavelet Transform. 41. 43-56.

[34] K. Zhang, W. Zuo and L. Zhang, "FFDNet: Toward a Fast and Flexible Solution for CNN-Based Image Denoising," in IEEE Transactions on Image Processing, vol. 27, no. 9, pp. 4608-4622, Sept. 2018, doi: 10.1109/TIP.2018.2839891.

Framework of Infotainment using Predictive Scheme for Traffic Management in Internet-of-Vehicle

Reshma S[1]
Department of Computer Science & Engineering
Global Academy of Technology
Bangalore, India

Chetanaprakash[2]
Department of Computer Science & Engineering
Bapuji Institute of Engineering & Technology
Davangere, India

***Abstract*—Infotainment system potentially contributes towards controlling accident fatalities in the era of Internet-of-Vehicles (IoV). Review of existing system is carried out to find that irrespective of various methods towards infotainment system, the quality of data being retrieved as well as issues associated with power and traffic congestion in vehicular communication is still an impending challenge. Therefore, this manuscript introduces a novel predictive scheme that offers enriched set of information from the environment to assists in decision making. Reinforcement learning is adopted for controlling traffic signal and power while the proposed system introduce augmented Long Short Term Memory scheme in order to predict the best possible traffic scenario for assisting the infotainment system to make a precise decision. The simulation is carried out for proposed system with existing learning schemes to find out proposed scheme offers better performance in every respect over challenging scene of an IoV.**

Keywords*—Infotainment system; internet-of-vehicle; reinforcement learning; decision making; power; long short term memory

I. Introduction

The concept of vehicular communication system arrives from vehicular adhoc network two decades back in order to facilitate comfortable and safer driving experience [1]. However, owing to the complex structural implementation and various problems associated with such forms of adhoc network, a reliable and safer communication cannot be guaranteed. So, the most recent innovations of Internet-of-Things (IoT) have introduced an Internet-of-Vehicle (IoV) system that is mainly formed to minimize the event of fatal accidents on road [2]. This is carried out by installing IoT objects within the vehicle which is known to facilitate various functionalities. One such form of system which creates a bridge of communication between the vehicle and external entities is infotainment system [3][4]. The contribution of infotainment system is quite significant especially when deployed over an IoV with respect to essential data transmission [5]. It doesn't only pertain to data transmission based on real-time data, but it also carry out various analytical operation to judge the traffic system. This analyzed outcome is disseminated to drivers via infotainment system in order to ensure safer driving over road [6]. A study shows that out of all deaths happened in country of Sri Lanka, maximum of deaths were due to road accidents [7]. Among the road accidents, maximum of them occur during morning hours of 9:00 AM to 10:00 AM and also during evening hours of 6:00 PM to 7:00 PM [8]. This clearly indicates that during rush hours, there will be more accidents and fatalities. Hence an efficient traffic management system is required to manage the traffic and avoid congestion and ultimately accidents. The travel time of the emergency vehicles is also an important factor. It is shown that risk of death due to cardiac arrest will increase by 95% during initial 3 hours of time. Hence, the travel time of emergency vehicles should be much lesser compared to travel time of ordinary vehicles. The infotainment system can be used to transfer vital information and entertainment information together. Since the Infotainment system is aware of GPS co-ordinates and health information of the vehicle, the same information can be used to perform several important tasks e.g. redirecting the driver to roads which have less traffic congestion with connected Infotainment system this can be used to manage entire city's traffic. CityFlow provides an excellent platform for simulating city's traffic and urban mobility in general [9]. The CityFlow platform is built in python and is 20 times faster than the other popular alternatives [10]. It is also found to be compatible with Reinforcement Learning (RL) techniques and hence it can be used along with RL agents. Reviews show that studies towards infotainment system and IoV still demands lot of improvisation that motivates to carry out proposed study.

The proposed study presents a unique computational framework of an infotainment system that is meant for data dissemination over congestion-free traffic using machine learning. The contributions of this study are: i) a unique traffic model is implemented for an IoV considering power consumption, ii) a better traffic management is presented to control power and traffic signaling operation in distributed manner applicable for an IoV operation, and, iii) an analytical model is built which is responsible for carrying predictive analysis of data dissemination for infotainment system with an effective decision making system considering the dynamicity of practical IoV environment. The organization of the manuscript is as follows: Section II discusses about existing investigation towards infotainment system followed by research problem highlights in Section III. Briefing of adopted research methodology is carried out in Section IV while an elaborated discussion about the system implementation is carried out in Section V. Discussion of Result analysis is carried out in Section VI while conclusive remarks of proposed contribution is carried out in Section VII.

II. Related Work

This section discusses about the existing studies being carried out towards IoV with a special emphasis towards infotainment system supportability. Recent studies towards IoV have been reviewed with respect to various methodologies and its effectiveness is studied.

The recent work carried out by Wu et al. [11] has constructed a hybrid communication system which mainly targets towards energy-efficient data transmission system using infotainment. The study also introduces a selection of cache nodes for all the intelligently connected vehicles. The limitation of the study is associated with non-inclusion of spatial complexity associated with streaming over such caching system. Adoption of machine learning towards communication system via infotainment system is carried out by Xu et al. [12] where a reinforcement learning algorithm has been used. The purpose is to encapsulate fluctuating patterns of channel condition in order to select a specific frame. The limitation of the study is that it emphasizes mainly on achieving throughput without consideration of vehicle density or emergency condition. Din et al. [13] have developed a caching system which assists in placing an appropriate content over the target vehicle. The limitation of this model is its non-consideration of uncertain traffic situation which could adversely affect the caching process. Vasudev et al. [14] have developed a unique communication system that emphasize over the mutual authentication scheme in vehicle-to-vehicle communication system. The limitation of this approach is that it uses cryptographic operation over a constraint device, which cannot be considered over a long run without performing any form of optimization of key management. Benarous et al. [15] have implemented a secure communication scheme in IoV where maintains privacy of location-based services utilized by the vehicles. The limitation of this scheme is that it doesn't present any identification system towards intruders and implementation is carried out considering known adversarial scenario.

A robust infotainment system over an IoV also demands an efficient resource management scheme as seen in work of Ni et al. [16]. The study uses allocation of resources as well as broadcasting of beacons over arbitrary access points for performing congestion control. However, the limitation of the study is that the model carry out the resource allocation without considering dynamic traffic scenario as well as it doesn't cater up any emergency services too during communication. Adoption of deep learning is witnessed in work of Chang et al. [17] where a model for accident detection system is developed. Upon detecting the collision, the information is transmitted to cloud-services which release notification. The limitation of the study is that it response time of notification completely depends upon the traffic and priority system, which may fail to cater up emergency transmission of accident notification. Silva et al. [18] have carried out a study towards social IoV system which uses conventional communication system in order to perform exchange of data among the vehicles. The paper concludes that there is still an unsolved problem associated with ethical guidelines about such communication in IoV. The work carried out by Sharma and Liu [19] have addressed the problem of misbehavior detection using machine learning in IoV. The study has used supervised learning model for this purpose. The limitation of this work is it is applicable only for specific attack. The work carried out by Wang et al. [20] has developed a behavioral modelling that predicts the driving strategy for safer driving. However, this model completely lacks associating with traffic system in order to exchange such information using infotainment system.

The work presented by Qureshi et al. [21] has presented a mechanism of data propagation using clustering approach in IoV. The method calls for a selection of a cluster head using self-assessment approach as well as routing attributes for data exchange within one-hop nodes. The limitation of this paper is that it consumes too much time in clustering process and does leads to delay in case of heavy traffic in IoV. The work carried out by the Fu et al. [22] have presented a transcoding operation for multimedia streaming in IoV over fog computing. The study uses a reinforcement learning scheme which assists in optimizing the allocation of an appropriate resource for facilitating streaming in IoV. The limitation of this work is its it cannot be used for large stream of data in dense traffic. Mechanism of content caching is implemented in work of Xue et al. [23] where a dynamic programming has been used for minimizing the problem of content caching in data transmission of vehicular network. Irrespective of reduced delay, the study model suffers from poor scalability issues in presence of massive number of vehicle density as well as there is no scheme for prioritization of certain vehicle that seems quite impractical. Existing system has also witnessed modelling of task orchestration in vehicular network as reported in work of Sonmez et al. [24]. The study has used machine learning approach considering the success score of task completion. However, the limitation of the study is its non-inclusion of traffic-lights or centralized controlling system, without which the model is not practical to implement. Hong et al. [25] have presented a cost optimization based scheme using analytical framework in order to enhance the transmission time in IoV network. The model suffers from pitfall of using static threshold for cost, which is impractical in real-world traffic. The work carried out by Hou et al. [26] has used Q-learning-based strategy for content management in IoV. Although, the model is capable of making prediction for movement of vehicle, it doesn't have any inclusion of multiple path decision over urban traffic. Apart from these, there are various work carried out by Xia et al. [27], Su et al. [28], Ni et al. [29], and Heo et al [30] towards improving communication system with respect to infotainment system in an IoV.

Existing approaches discussed about offers a claim to better outcome; however, they are also associated with some significant issues. The next section outlines the research problems explored from this review.

III. Motivation for the Research

- The traffic congestion is a daily day problem especially in a country like India. The traffic congestion can be easily mitigated with the existing infrastructures and roads. The issue is not of the infrastructure but of the poor management of the infrastructure. Hence an efficient system needs to be designed to manage all the infrastructure and get a better results for the same.

- The connected vehicles are no longer a dream of the future. With several car companies like Tesla and Morris Garage supporting the connected cars by default, the system does not require a hardware upgrade anymore. The computing power of the cars is more than the horsepower of themselves as of now. Since the computing power already exists in the car they only need a software update to support the smart Internet of Vehicle infrastructure now a days.
- The Road accidents due to congestion are a serious cause of the concern now a days. With the systems like autopilot from tesla and several other cutting edge technologies, it is possible to automatically redirect the city's traffic easily with the modern technological systems. And hence there is a scope for a system that can plan the city's traffic and redirect it to a suitable destination on the go.
- The heart attack and other serious emergencies must be addressed immediately. There is a limit to which people cooperate with the emergency vehicles. Green corridor is a very much common phenomenon in cities where an emergency vehicle is given zero traffic and fully allowed to pass through. This is traditionally done for VIP vehicles. However for the genuine emergency vehicles, people give way with understanding. This is possible to be modified based on traffic signal where the emergency vehicles can be made to reach destinations much faster that they are now.
- The Infotainment systems make the driving experience less of a hassle but more of an enjoyable experience. This can be achieved with the proposed study.

IV. Research Problem

After reviewing the existing system in infotainment system in IoV, following are the open end problems identified in proposed study:

- Restricted Coverage Issues in IoV: The conceptual definition of an IoV calls for an interconnected vehicles; however, they still have a dependencies towards a fixed infrastructure at one point. It could be in the form of a hotspots mounted on road or embedded within traffic signal in order to guide the vehicles for congestion free direction. However, the existing studies don't report to consider this coverage issue from Road Side Units (RSU) and mainly focus on vehicle to vehicle communication. This is incomplete implementation for any IoV system to assists the infotainment system within the vehicle.
- Non-inclusion for Density Monitoring: Majority of existing studies on IoV and infotainment system fixes the number of vehicles on specific route. However, in real-time, there are fair possibilities of either increase or decrease of such density over an uncertain instance of time. Without this consideration, the infotainment system will either faces congestion issue or face scarcity of information to undertake decision of data transmission.
- More Focus on Navigation: Majority of existing studies towards infotainment system only focuses on route navigation, whereas infotainment system can also be used for various other forms of data transmission at same time. This requires a dedicated and congestion-free communication channel to be explored by the infotainment system in vehicles. Even if this concept is implemented within present state of implementation in infotainment system, it will significantly cause a serious bottleneck condition for the traffic among the vehicles.
- Less Emphasis towards Data Quality: In IoV system, there are numerous numbers and types of data being required to fulfill the process of data dissemination within an infotainment system. Although usage of mobile edge computing and cloud services makes the operation easier, but still there is a serious pitfalls of almost all the existing architecture of data transmission in IoV. This generates a massive set of traffic data which pertains to road attributes as well as vehicle attributes. Apart from this, there is also a need of multi-objective function to develop a model, which can extract only the productive traffic-related information within IoV system. Hence, ensuring data quality is quite a challenging scenario within current state of infotainment system.
- Uses of Sophisticated Technique: Existing system adopts sophisticated technique targeting for data transmission within vehicles ignoring the resource efficiency of the infotainment system. Adoption of machine learning demands higher training, which is again not much reported to be resource friendly for all implementation carried out in IoV till date.

Therefore, it can be seen that above mentioned issues do exist in present time of IoV deployment scenario. From practical viewpoint, this problem is much dominantly seen in a road network R_n with multiple junction point. It is because of the decision to find the optimal path owing to the problems identified in this section. Therefore, the prime problem formulation of the proposed system can be stated as follow:

$$R_n(S) \rightarrow [opt(r_i)] \asymp A_i \quad (1)$$

In the above problem formulation, the core idea is to obtain a better form of road network R_n for all state attributes associated with intersection points. The idea is to optimize the set of reward r_i parameters for all the set of actions A_i considered in environment of IoV.

This problem is tackled by developing a computational framework that implements a conditional logic for vehicles considering its properties. Further reinforcement learning approach is used to redefine various state attributes that resolves the decision making problem further using LSTM attention network.

V. Research Methodology

The core aim of the proposed system is to design and develop a smart traffic system which is capable of facilitating enriched information to the infotainment system embedded within the vehicle in IoV. Adopting an analytical research methodology, the proposed system make use of machine learning approach in a unique manner which assists in better decision making in the form of direction as well as seamless data transmission in IoV. The proposed system emphasizes more on data quality, where data is associated with both traffic and vehicles in order to assist the infotainment system to undertake correct decision of route formation and resource-efficient seamless data dissemination in IoV. The architecture developed for this notion is highlighted in Fig. 1 as follows:

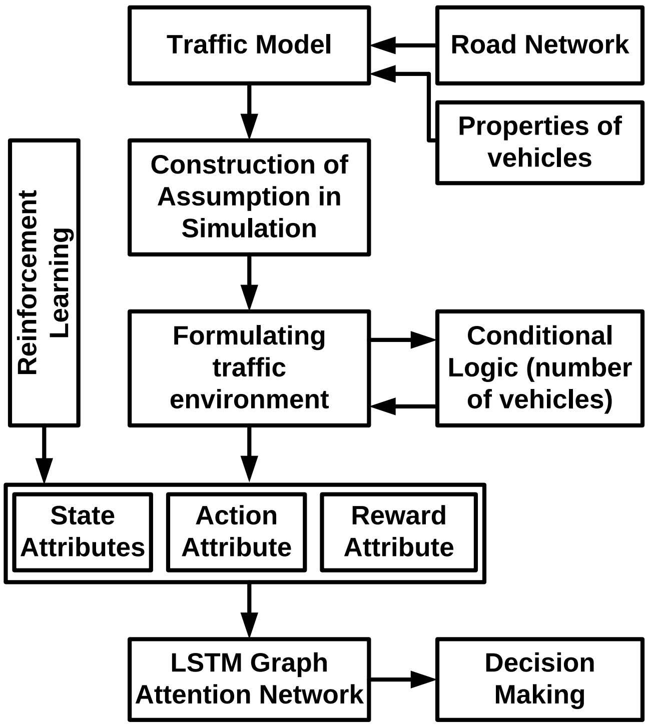


Fig. 1. Architecture of Proposed System.

According to Fig. 1, the proposed system develops a traffic model that mainly consists of formation of road network and properties of vehicle. The proposed modelling considers various attributes in order to develop the topology of traffic model. Further the proposed scheme also constructs assumption which is used for simulation study followed by considering all the essential challenges involved in developing this model. A traffic environment is formed where specific conditional logic is constructed. The proposed system makes use of reinforcement learning scheme which is used over framing up state attributes, action attributes, and reward attributes. Further, Long Short Term Memory (LSTM) graph attention network is utilized which is basically used for decision making towards opting for congestion free and reserving resources while performing vehicular communication in IoV. The next section elaborates further about the operation being carried out by each block towards infotainment system.

VI. System Implementation

In order to design an infotainment platform/scenario, it is required to realize that all vehicles in IoV are required to be strongly interconnected with each other in order to make a seamless transmission. There might be some vehicles connected directly to internet via 4G/5G; however it is required to ensure that the connectivity is given to all vehicles in order to ensure transmission of vital data like traffic and emergency data. Before simulation of the infotainment system itself, the traffic and the congestion is needed to be simulated first and therefore the proposed system is simulated using standard CityFlow simulator [9]. This section discusses about the various aspects of the system implementation.

A. *Traffic Model*

In order to simulate the traffic scenario, A road network R_n with 4 junctions as {J_1, J_2, J_3, J_4} is considered with three level of congestion as i) highly congested, ii) moderately congested, and iii) less congested. The model defines a vehicle *V* with characteristic elements from the set of properties viz. length, width, maximum positive acceleration, maximum negative acceleration, usual positive acceleration, usual negative acceleration, minimum gap, maximum speed, headway time. The brief highlights of these properties are as follows:

- Length refers to the length of vehicle including the luggage space and bumpers.
- Width of vehicle refers to physical width of the vehicle including mirrors.
- Maximum positive acceleration is the change in speed of the vehicle when accelerator is applied in full throttle.
- Maximum negative acceleration is change in speed of the vehicle when sudden break is applied.
- Typical positive acceleration is the usual acceleration of the vehicle.
- Typical negative acceleration is usual change in speed occurred when breaks are applied.
- Minimum gap is recommended gap that should be maintained between the vehicles.
- Maximum speed is top speed of the vehicle.
- Headway time is the time taken by the following vehicle to reach the position of leading vehicle.

B. *Assumptions on Traffic Simulation*

The assumptions being considered while developing the proposed schema of infotainment are as follows:

- Everyone respects traffic rules and lane discipline.
- It is assumed that no mishaps happen like accidents.
- All roads are in good condition.
- There are no two wheelers and three wheelers. All the vehicles are assumed to be cars or emergency vehicles.
- Everyone tend to move at similar speeds.

Another, important properties defined for a vehicle are: {Interval, Start time, End time} with default values of {5.0, 0, -1} respectively. These values are considered using 5.0 Likert Scale which signifies 5 as highest and -1 as lowest score. The design process of the model defines a definite simulation time (T_s). If the start time is equal to zero, it means that at the beginning of simulation, the vehicles will appear at their respective junction, however if the end time is equal to -1, it means that it is uncertain to say that when again a particular vehicle will re-appear on the same junction. Moreover, if the interval is defined say interval=5, it means that at every 5 units of time, that vehicle will re-appear on respective junction.

C. Challenges of Modelling

In the present study, the problem is being formulated as a Markov chain model. Each intersection in the system is controlled by an agent. The infotainment system which is present within the vehicle is an embedded system hence it only has a routing table to forward the information. Since all the information is encrypted only the end node can see the required information. SNR of the multimedia signals are noted at the cars and average SNR is calculated. SNR is calculated for 4 different types of data viz. i) text data, ii) video data, iii) audio data, and iv) security data.

D. Traffic Environment from CityFlow

The CityFlow simulator is used to generate the traffic data for three years' time period. Apart from the traffic scenarios of junctions and vehicle characteristics, the program (simulator) also keep adding vehicles with random start time and end time parameters over the span of simulation. Hence, the number of vehicles on the roads keep increasing and creates a dynamic and uncertain stage of congestion. In addition the simulator is internally programmed to model seasonal traffic in such a way that the number of vehicles on the road will be comparatively lower in the month of July and august due to rainy season. The problem of mitigating the congestion, require information as in the Table I.

In the Table I, the parameter of MED is computed as following expression (1),

MED (S)=

$$\begin{cases} S\left[\frac{1}{2}\times\left(\sum_{J=1}^{n} V_J\right)\right] & if \text{ number of vehicles are even} \\ \frac{S\left[\frac{1}{2}\times\left(\sum_{J=1}^{n} V_J-1\right)\right]+S\left[\frac{1}{2}\times\left(\sum_{J=1}^{n} V_J+1\right)\right]}{2} & if\ number\ of\ vehicles\ are\ odd \end{cases} \quad (1)$$

TABLE I. FORMULATION USED IN PROPOSED STUDY

Junction ID	Total number of vehicles	Speed	Congestion
{J1,J2,J3,J4, .. Jn}	$\sum_{J=1}^{n} V_J$	MED (S)	$\left(\sum_{J=1}^{n} V_J, speed = 0 \Big/ \sum_{J=1}^{n} V_J\right) \times 100$

E. Formulating State Attribute

State is definitive term that represents the state of the particular intersection. Since it has multiple values, it is represented in form of a vector *S* as follows,

$$S = \{\vec{S_1}, \vec{S_2}, \vec{S_3} \dots \vec{S_n}\}$$

$$\vec{S_i} = [L_q, P] \quad (2)$$

In the above expression (2), the variable L_q represents the average queue length of the intersection that is mathematically represented as follows,

$$L_q = \frac{1}{4}\sum_{i=1}^{4} u_i \quad (3)$$

u_i is the queue length of the individual road in the intersection.

F. Formulating Actions Attribute

Actions are execution attribute that Reinforcement Learning RL agent can perform on the environment. Since a single RL agent is assigned to an intersection, there are possibilities of *n* number of actions *A* as follows,

$$A = \{A_1, A_2, A_3 \dots A_n\}$$

$$A_i = [\vec{T_i}, P_i] \quad (4)$$

In the above expression (4), the variable P_i represents the power input of the base station. If the RL agent sets a higher power then the signal can be transmitted further and results in a higher useful information ratio. At the same time, it also results in higher overall power consumption. The first variable in expression (4) is represented as follows,

$$\vec{T_i} = [X_1, X_2, X_3, X_4] \quad (5)$$

In the above expression (5), the variable X_i represents the traffic signal. Since there are 4 signals in each intersection it is represented by X_1 to X_4 and its generalized form is as follows,

$$X_i \in \{R, Y, G\} \quad (6)$$

In the above expression (6), the variable R, Y, and G represents three different lights in the traffic signal. Red, Yellow and Green.

G. Formulating Reward Attribute

Reward r_i is a real number representing the overall performance of the system.

$$r_i = -\frac{\sum_{i=1}^{4} u_i}{4} - P_i + \frac{\sum_{i=1}^{N} s_i}{N} \quad (7)$$

The RL system proposed in this study is programmed in such a way that both traffic congestion as well as information SNR are optimized. The information is passed through a software Defined Network (SDN) created by Mobile Adhoc Networks (MANET) by the vehicles. The parameters which are being optimized here are,

- Useful information ratio of 4 different types of data (varying preference) (MAX).
- Traffic congestion (Average queue length) (MIN).

- Average Travel time of the regular vehicles (MIN).
- Average Travel time of the emergency vehicles (Only ambulances are considered) (MIN).
- Overall Power consumption by the base stations (MIN).

H. Methodology for Implementing LGAT Neural Network

The proposed system implements a neural network in the form of regular Long Short-Term Memory LSTM itself; however, one of the hidden layers in this network is common for all the networks over the grid. This essentially makes each neural network to be aware of its surroundings. Hence this is named as LSTM Graph Attention Network (LGAT). LGAT has two parts involved in its module i.e. i) First part which is before the GAT layer and ii) second part is after the GAT layer. Before GAT layer rectified linear unit ReLU Activation function is used whereas after GAT layer, Sigmoid function is used. This is done since the output is always expected to be residing between 0 and 1. The power input of the base station is controlled by considering the input of the percentage of maximum power consumption. The traffic signals are always controlled by considering the input of 1 or 0 to each signal lamp with the one hot encoding strategy. The Adam optimizer is used to train the neural network and the loss function used here is MSE. It should be taken into special attention that MSE is used here instead of commonly used binary cross entropy. This is due to the fact that the network should output an analog value for power consumption in terms of percentage.

Fig. 2 shows the structure of the proposed LGAT neural network where the second hidden layer is the shared layer whose weights and biases are shared with all the other networks. The weight sharing mechanism here is very similar to that of the Siamese neural networks. The output layer contains 13 outputs 12 of which corresponds to the traffic signals and one corresponds to the base station input power percentage.

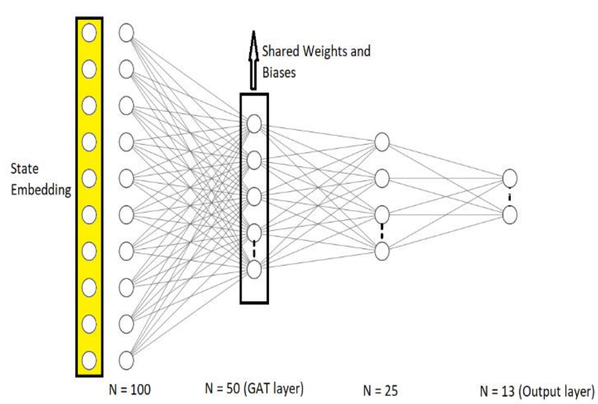


Fig. 2. Structure of Proposed Neural Network.

I. Training the Agent

The agent is trained for 4000 episodes in the study and the ANN shown here is trained for one epoch in every episode. Effectively the ANN is trained for 4000 epochs with dynamically changing data. However, since this is an LSTM network, the changing nature of the data is also learnt by the network. The central shared hidden layer enables multiple agents to cooperate with each other. Due to this cooperation the agents are able to produce a good output. Here, the competition behind agents must be avoided at all costs as that will results in selfish agents which may just shut down their base stations in order to save power.

VII. Result Analysis

This section discusses about results being obtained from the simulation study by implementing the proposed scheme discussed in prior section. The recommended hardware and software stack for training the agent are as follows.

- CPU: Intel Core I7 10th Generation.
- GPU: Nvidia GeForce RTX 2060.
- OS: Kali Linux 2021.
- C compiler: GCC 10.2.1.
- GPU C library: Nvidia CUDA 10.1.
- GPU python bridge: Nvidia CuDNN.
- Python: 3.8.2.
- TensorFlow: 2.5.0.

The above-mentioned stack is used in order to get the best results since the TensorFlow works better when it is executed on GPU. Above stack must be used in order to run TensorFlow over GPU. Table II highlights about the properties of vehicles considered for proposed scheme.

TABLE II. Properties of a Vehicle (V)

#	Property	Value	Units
1	Length	5	Feet
2	Width	2	Feet
3	Maximum Positive Acceleration	2	m/sec^2
4	Maximum Negative Acceleration	4.5	- m/sec^2
5	Typical Positive Acceleration	2	m/sec^2
6	Typical Negative Acceleration	4.5	- m/sec^2
7	Minimum Gap	2.5	Feet
8	Maximum Speed	16.67	m/sec
9	Headway Time	1.5	second

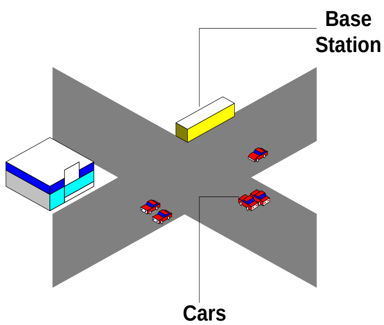


Fig. 3. Simulation Environment.

As it can be observed from Fig. 3 that the simulation is a set up in such a way that there is always a base station in every intersection. Fig. 4 exhibits the higher level overview of system implementation. It is powerful enough to transmit till next base station. However, since the signal strength of the base station can be controlled by the RL agent, the base station's power consumption will vary and the range also varies. If there are vehicles closer to each other, then it is enough if the base station transmits the signal to nearest car. That car can act as a repeater and transmit the message further to other cars. Hence if a particular junction is congested, then the base station may spend less amount of power to transmit the signal further.

The reward depends on both travel time and power consumption. Hence the agent is expected to optimize both of these parameters. The environment is built in such a way that it can support one agent per every intersection. Hence this is a multi-agent environment. Every agent can perform the optimization of power and congestion in their own intersections however, they are expected to co-operate with each other and optimize the entire city's power consumption and travel time of entire city in average.

There are several base stations are present in the city as well as vehicles act as relay to the signal. If the vehicles are far apart, then the base stations have to send signal far hence there will be more power consumption by the base stations. The system must optimize the over power consumption as well.

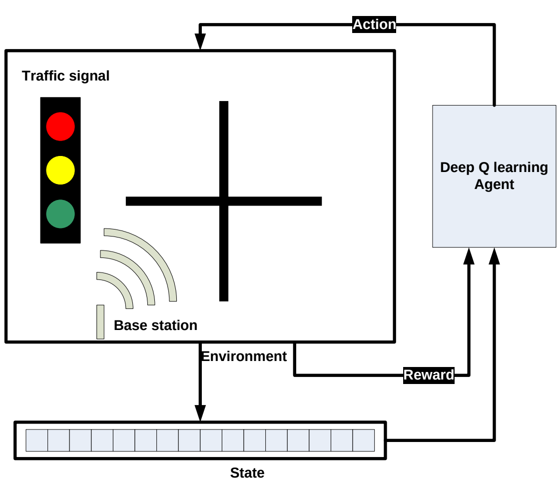


Fig. 4. Higher Level Overview of the System.

Following are the parameters which are being optimized

- Useful Information Ratio: Every vehicle needs to receive the information required for itself. More it acts as a relay, more the battery consumption and lower the bandwidth utilization for itself. Hence base station must provide higher power for better data transmission.
- Queue Length: Queue length is defined as the distance between front of the first stopped car in the intersection to the back of the last stopped car in the intersection (Feet). Traffic congestion is the average of all 4 incoming queue lengths in each intersection.
- Average Travel Time: Average travel time is nothing but average time taken by all cars to travel from source to destination (Entry intersection, Exit intersection). It is considered for Regular vehicles and Emergency vehicles.
- Overall Power Consumption: This is the sum of power consumed by all base stations in the city. The order in which the priority is given to the parameters viz. Travel time of Emergency vehicles, Congestion, Travel time of Regular vehicles, Useful info ratio for emergency data, Useful info ratio for text data, Useful info ratio for audio/video data, Power consumption Simulation parameters are set as following.

Fig. 5 highlights the consideration of 6X6 grid for proposed simulation with 36 intersections in total while there are two simulations done using this layout. Uniflow assumes that the traffic moves in a single direction during morning and opposite direction in the evening. Biflow assumes that the traffic moves in both directions during all times of the day.

Fig. 6 highlights the map to shows the area considered in Hangzhou junction of China. It contains total of 16 junctions and traffic is the real recorded traffic.

Fig. 7 highlights the map to shows the area considered in New York city that contains 196 junctions in total. The proposed system is assessed with existing system of learning-based model of vehicular network in IoV.

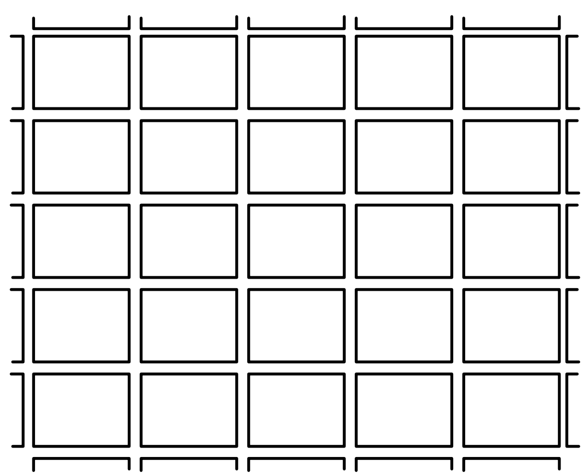


Fig. 5. 6X6 Grid for Simulation.

Fig. 6. Hangzou Simulation Setup.

Fig. 7. New York Simulation Setup.

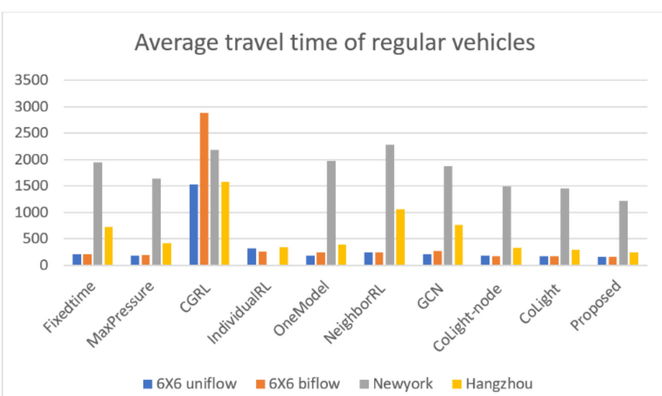


Fig. 8. Average Travel Time for Various Methods.

The above graph in Fig. 8 clearly indicates that the proposed method performs better for every scenario. As it can be observed, New York City is the most difficult scenario. For individual RL, data isn't available. The overall travel time is reduced because of the LGAT architecture. The vehicular traffic follows a particular pattern during the day. LGAT can learn the temporal patterns as well and be able to predict the future vehicular traffic.

An extra parameter which is considered in this study is that average travel time of emergency vehicles (Fig. 9) in which the proposed system is performing better compared to CoLight model. The performance is better in the proposed system since the LSTM layer is used From Fig. 10, can be observed that the travel time of emergency vehicle is half of regular vehicles.

The graph in Fig. 11 shows that the overall power consumption is less for proposed method. This evidently shows that proposed system has better performance score when evaluated with existing CoLight model in perspective of different available dataset of data dissemination in vehicles of urban scenario.

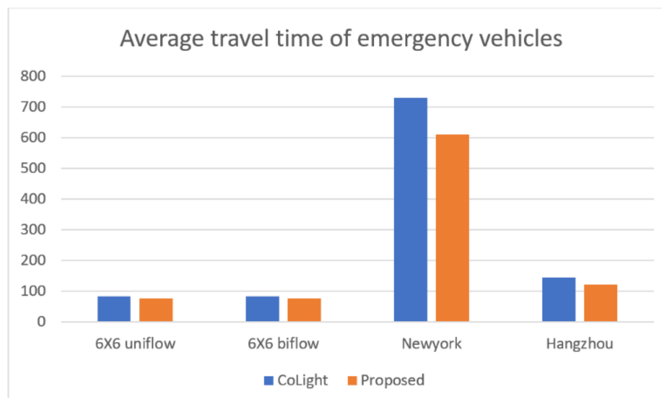


Fig. 9. Comparison of Travel Time for Emergency Vehicles.

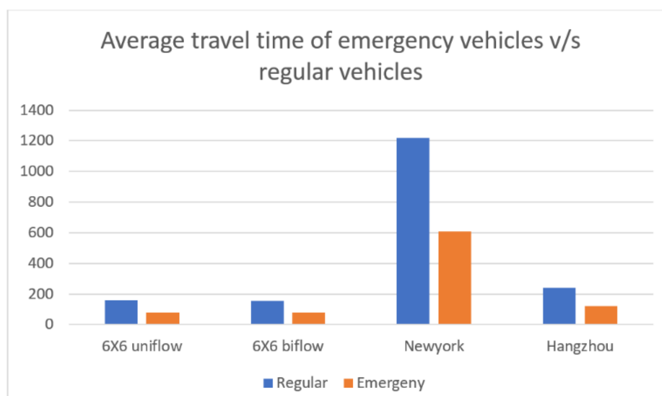


Fig. 10. Average Travel Time for Emergency Vehicle vs Regular Vehicle.

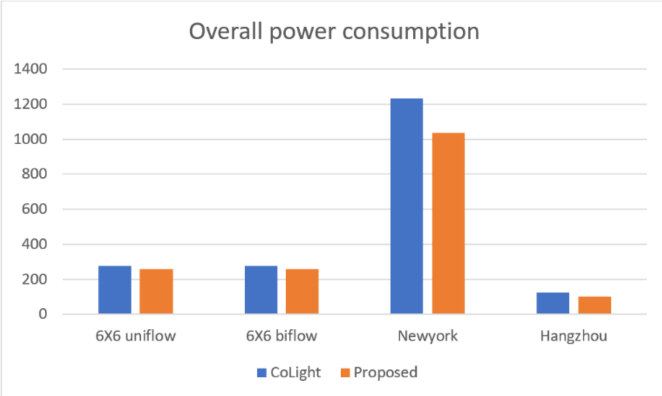


Fig. 11. Overall Power Consumption in MWH.

VIII. Conclusion

This study presents a novel mechanism to manage the urban traffic system and prevents the congestion at the same time. The system can be used to redirect emergency vehicles to shorter paths where there is less congestion and reduce their travel time. The presented model considers the aspect of power consumption by the system. This is carried out in order to address the problem about infotainment system that not only consists of the in-vehicle system but also the sensors, gateways and signal repeaters. Such forms of devices consume a lot of power in order to make sure the quality information is transferred. The implementation of proposed study also optimizes the power consumption by the infotainment system and ensures overall transmission efficiency. The proposed system constructs an optimal environment for the city traffic management and its reward system so that the system rewards the agent based on both power consumption and the travel time. While the environment is a single environment, this is a multi-agent system. An RL agent is assigned at every intersection and they control the traffic signal and power input of the transmitter at the intersection. An agent will also be aware of actions and states of other agents through a novel neural network architecture proposed in this study, LGAT architecture. The proposed system implements an LGAT that is a special type of LSTM in which one of the layer's weights and biases are shared with all other agent's weights and biases. The neural network here uses the DQN architecture for RL. The DQN architecture means the NN takes the action as input and outputs Q values for all possible actions. Q values are nothing but the future rewards for the system. The system is trained over multiple episodes with a single epoch per episode. The proposed study considers several existing methods and considers various parameters to study the traffic. There are two synthetic environments and two realistic environments in this study. The two realistic environments are the traffic data from Network city and Hongzow junction from Hong Kong. The synthetic environment contains two different environments which are 6x6 uniflow and 6x6 Biflow.

References

[1] N. Chowdhury, L. M. Mackenzie, Vehicular Communications for Smart Cars Protocols, Applications and Security Concerns, Taylor & Francis Limited, ISBN: 9780367457440, 036745744X, 2021.

[2] G. Zhao, G. Zhu, Big Data Transportation Systems, World Scientific Publishing Company, ISBN: 9789811236013, 9811236011, 2021.

[3] J. Aulinas, H. Sjafrie, AI for Cars,CRC Press, ISBN: 9781000417166, 1000417166, 2021.

[4] W. Xu, H. Zhou, X. Shen, Internet Access in Vehicular Networks, Springer International Publishing, ISBN: 9783030889913, 3030889912, 2021.

[5] W. Shi, L. Liu, Computing Systems for Autonomous Driving, Springer International Publishing, ISBN: 9783030815646, 3030815641, 2021.

[6] A.Khanna, D.Gupta, P.L. Mehta, V.H.C. de Albuquerque, Smart Sensors for Industrial Internet of Things Challenges, Solutions and Applications, Springer International Publishing, ISBN: 9783030526245, 3030526240, 2021.

[7] https://www.worldbank.org/en/news/feature/2021/11/04/sri-lanka-s-journey-to-road-safety.

[8] S. K. Singh, "Road Traffic Accidents in India: Issues and Challenges", Elsevier-ScienceDirect, Transportation Research Procedia, vol.25, pp.4708-4719, 2017.

[9] https://cityflow-project.github.io/.

[10] L. Raes, P. Michiels, T. Adolphi, C. Tampere, "DUET: A Framework for Building Secure and Trusted Digital Twins of Smart Cities", IEEE Internet Computing, 2021, DOI:10.1109/MIC.2021.3060962.

[11] H. Wu, J. Zhang, Z. Cai, F. Liu, Y. Li and A. Liu, "Toward Energy-Aware Caching for Intelligent Connected Vehicles," in IEEE Internet of Things Journal, vol. 7, no. 9, pp. 8157-8166, Sept. 2020, doi: 10.1109/JIOT.2020.2980954.

[12] W. Xu, S. Guo, S. Ma, H. Zhou, M. Wu and W. Zhuang, "Augmenting Drive-Thru Internet via Reinforcement Learning-Based Rate Adaptation," in IEEE Internet of Things Journal, vol. 7, no. 4, pp. 3114-3123, April 2020, doi: 10.1109/JIOT.2020.2965148.

[13] I. Ud Din, B. Ahmad, A. Almogren, H. Almajed, I. Mohiuddin and J. J. P. C. Rodrigues, "Left-Right-Front Caching Strategy for Vehicular Networks in ICN-Based Internet of Things," in IEEE Access, vol. 9, pp. 595-605, 2021, doi: 10.1109/ACCESS.2020.3046887.

[14] H. Vasudev, V. Deshpande, D. Das and S. K. Das, "A Lightweight Mutual Authentication Protocol for V2V Communication in Internet of Vehicles," in IEEE Transactions on Vehicular Technology, vol. 69, no. 6, pp. 6709-6717, June 2020, doi: 10.1109/TVT.2020.2986585.

[15] L. Benarous, S. Bitam and A. Mellouk, "CSLPPS: Concerted Silence-Based Location Privacy Preserving Scheme for Internet of Vehicles," in IEEE Transactions on Vehicular Technology, vol. 70, no. 7, pp. 7153-7160, July 2021, doi: 10.1109/TVT.2021.3088762.

[16] Y. Ni et al., "Toward Reliable and Scalable Internet of Vehicles: Performance Analysis and Resource Management," in Proceedings of the IEEE, vol. 108, no. 2, pp. 324-340, Feb. 2020, doi: 10.1109/JPROC.2019.2950349.

[17] W. Chang, L. Chen and K. Su, "DeepCrash: A Deep Learning-Based Internet of Vehicles System for Head-On and Single-Vehicle Accident Detection With Emergency Notification," in IEEE Access, vol. 7, pp. 148163-148175, 2019, doi: 10.1109/ACCESS.2019.2946468.

[18] R. Silva and R. Iqbal, "Ethical Implications of Social Internet of Vehicles Systems," in IEEE Internet of Things Journal, vol. 6, no. 1, pp. 517-531, Feb. 2019, doi: 10.1109/JIOT.2018.2841969.

[19] P. Sharma and H. Liu, "A Machine-Learning-Based Data-Centric Misbehavior Detection Model for Internet of Vehicles," in IEEE Internet of Things Journal, vol. 8, no. 6, pp. 4991-4999, 15 March15, 2021, doi: 10.1109/JIOT.2020.3035035.

[20] X. Wang, S. Han, L. Yang, T. Yao and L. Li, "Parallel Internet of Vehicles: ACP-Based System Architecture and Behavioral Modeling," in IEEE Internet of Things Journal, vol. 7, no. 5, pp. 3735-3746, May 2020, doi: 10.1109/JIOT.2020.2969693.

[21] K. N. Qureshi, M. M. Idrees, J. Lloret and I. Bosch, "Self-Assessment Based Clustering Data Dissemination for Sparse and Dense Traffic Conditions for Internet of Vehicles," in IEEE Access, vol. 8, pp. 10363-10372, 2020, doi: 10.1109/ACCESS.2020.2964530.

[22] F. Fu, Y. Kang, Z. Zhang, F. R. Yu and T. Wu, "Soft Actor–Critic DRL for Live Transcoding and Streaming in Vehicular Fog-Computing-

Enabled IoV," in IEEE Internet of Things Journal, vol. 8, no. 3, pp. 1308-1321, 1 Feb.1, 2021, doi: 10.1109/JIOT.2020.3003398.

[23] Z. Xue, Y. Liu, G. Han, F. Ayaz, Z. Sheng and Y. Wang, "Two-Layer Distributed Content Caching for Infotainment Applications in VANETs," in IEEE Internet of Things Journal, vol. 9, no. 3, pp. 1696-1711, 1 Feb.1, 2022, doi: 10.1109/JIOT.2021.3089280.

[24] C. Sonmez, C. Tunca, A. Ozgovde and C. Ersoy, "Machine Learning-Based Workload Orchestrator for Vehicular Edge Computing," in IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 4, pp. 2239-2251, April 2021, doi: 10.1109/TITS.2020.3024233.

[25] X. Hong, J. Jiao, A. Peng, J. Shi and C. -X. Wang, "Cost Optimization for On-Demand Content Streaming in IoV Networks With Two Service Tiers," in IEEE Internet of Things Journal, vol. 6, no. 1, pp. 38-49, Feb. 2019, doi: 10.1109/JIOT.2018.2873085.

[26] L. Hou, L. Lei, K. Zheng and X. Wang, "A Q -Learning-Based Proactive Caching Strategy for Non-Safety Related Services in Vehicular Networks," in IEEE Internet of Things Journal, vol. 6, no. 3, pp. 4512-4520, June 2019, doi: 10.1109/JIOT.2018.2883762.

[27] S. Xia, F. Lin, Z. Chen, C. Tang, Y. Ma and X. Yu, "A Bayesian Game Based Vehicle-to-Vehicle Electricity Trading Scheme for Blockchain-Enabled Internet of Vehicles," in IEEE Transactions on Vehicular Technology, vol. 69, no. 7, pp. 6856-6868, July 2020, doi: 10.1109/TVT.2020.2990443.

[28] Z. Su, M. Dai, Q. Xu, R. Li and H. Zhang, "UAV Enabled Content Distribution for Internet of Connected Vehicles in 5G Heterogeneous Networks," in IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 8, pp. 5091-5102, Aug. 2021, doi: 10.1109/TITS.2020.3043351.

[29] Y. Ni et al., "Toward Reliable and Scalable Internet of Vehicles: Performance Analysis and Resource Management," in Proceedings of the IEEE, vol. 108, no. 2, pp. 324-340, Feb. 2020, doi: 10.1109/JPROC.2019.2950349.

[30] S. Heo, W. Yoo, H. Jang and J. -M. Chung, "H-V2X Mode 4 Adaptive Semipersistent Scheduling Control for Cooperative Internet of Vehicles," in IEEE Internet of Things Journal, vol. 8, no. 13, pp. 10678-10692, 1 July1, 2021, doi: 10.1109/JIOT.2020.3048993.

Internet of Things (IoT) Application for Management in Automotive Parts Manufacturing

Apiwat Krommuang, Opal Suwunnamek
King Mongkut's Institute of Technology Ladkrabang Business School
King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand

***Abstract*—Automotive parts manufacturing focuses on sustainable development manufacturing capabilities with future technology changes. Internet of Things (IoT) plays an important role in applying internet technology to machines and equipment in manufacturing processes for transformation towards Industry 4.0 as well as creating values added and higher competitive advantage for the sustainability of the industries. This research aims to study factors that influence decision-makers in selecting IoT applications for managing auto parts production and they consist of connectivity, telepresence, intelligence, security, and value including their fifteen sub-factors. In this research, The Fuzzy Analytic Network Process (FANP) is a Multiple Criteria Decision Making (MCDM) technique used to analyze, identify, and prioritize factors in selecting IoT applications for managing production processes. The questionnaire is designed based on the FANP technique to survey the importance of weight for each factor from executives of 88 auto parts manufacturers who authorize as the decision-makers for selecting IoT applications. The results have indicated that telepresence is the most important factor that will assist them in controlling production to guarantee that production capabilities meet the objective and connectivity is the second important factor that must ensure that IoT applications are compatible with their machinery and equipment can be controlled smoothly and precisely. Meanwhile, performance is the most important sub-factor and other sub-factors are ranked as functional orientation, data management, control, and compatibility respectively. Therefore, manufacturers can use this research as a criterion for selecting appropriate IoT applications for controlling their manufacturing for sustainable effectiveness.**

***Keywords*—Internet of things; auto parts industry; FANP; MCDM; sustainable manufacturing**

I. Introduction

The Internet of Things (IoT) is admitted as one of the most essential areas of future technology and is gaining various attention from several industries. IoT is considered a key driver for industries to transform industry 3.0 to industry 4.0 [1]. It is also crucial for integrating manufacturing processes for the fourth industrial revolution [2]. The gap between the physical and digital worlds can be bridged by synchronizing the flow of information with the physical flow for greater supply chain integration [3]. This is because IoTs devices play a crucial role in driving businesses to create superior connectivity and progressive operations to increase their agility [4]. It is an integration of interconnected sensing and communication technology that enables equipment, machines, sensors, robots, and automatic actuators [5]. IoT is determined into three groups: People to people, people to machine/things, Things/machine to things/machine, and Interacting through the internet [6]. The industrial internet of things is a concept based on the same principles as the IoT, but for connection machines in factories. The primary communication between machine-to-machine and automation based on the information exchanged between them [7]. Smart factory solutions are designed the enable sensors and connected edge devices to improve product quality and real-time factory operational efficiency. The potential to address the data captured and exchanged in real-time has multiplied [4]. Some devices are operating their functionality for progressive production, performance, and maintenance in the production [5] that can generate greater customer value through offered services. The efficacy can provide identifying, sensing, networking, and processing abilities for communicating with other devices and services through the internet.

Several organizations are searching for new technologies which can support their sustainable businesses [8] such as the Internet of Things (IoT), Big Data Analytics (BDA), Cloud Computing [5], and Blockchain (BC). Meanwhile, these new technologies can be used to support production and services in other sectors [9]. Besides, the mentioned technologies are currently applied in several industries for transforming towards Industry 4.0 which is a crucial industrial revolution in this century [10]. To improve supply chain structures and achieve sustainability, industry 4.0 becomes the most prominent solution in current manufacturing processes [11]. Therefore, industry 4.0 plays a critical role in bringing manufacturers to sustainable production [12]. Sustainable production covers important components such as production, process, and system which create value-added and sustainable growth for the business [13]. Therefore, this integration is a challenge for improving sustainable production in the future [14]. Moreover, Industry 4.0 also entails data exchange and autonomous systems, as well as manufacturing activities centered on the development of smart factories to respond to dynamic production [15]. Because the investment in IoT applications is expensive, decision-makers in selecting IoT applications in the auto parts production process confront a challenge in determining which relevant IoT apps should be chosen. Therefore, this study aims to answer this research question with the following objectives. 1) This research aims to study the influences of factors in selecting IoT applications in the auto parts industry, 2) the importance of each factor will be identified and prioritized, and 3) trends of decision-making for selecting IoT applications are studied for improving the management of the auto parts production. Thus, the first section of this article identifies factors that influence decision-making in selecting IoT applications by reviewing relevant literature. Subsequently, quantitative research is conducted by surveying decision-makers in the auto parts industry with a questionnaire developed based on the Fuzzy Analytical Network approach (FANP). The next section defines and formulas of FANP are introduced, and it can effectively identify and prioritize factors in selecting IoT applications for the auto parts production. Finally, the results of data analysis by FANP express and indicate important factors that influence decision-makers in selecting IoT applications for auto parts production.

II. Literature Review

A. Connectivity

Currently, the Internet is the interconnection hub of smart devices and can create an intelligent network throughout the value chain to machines, products, and systems [16]. This consequence is the connection and interaction between people and machines which leads to more interconnected production systems [17], the connection of physical devices and machines, and the integration of applications to the software or database [18]. The information was gathered through the IoTs to enhance the productivity of business processes [19]. Therefore, it is the industrial processes to connectivity technologies for monitoring the production activities, inventory management, sales management, and after-sales services. Compatibility is the connectivity and automation of collaboration activities [20]. and accessibility in entering the network while providing the general ability to generate data [6]. Data exchange is supported by different networks and data is gathered from networks connecting equipment, production line, and sensor network. Standardization is an important prerequisite in digitization. [21]. The connection of any object to the Internet could be one of the standardization challenges and the success of the development of interoperable. Standardization also enables the optimization between functionality, costs, and quality of applications and solutions. The key IoT infrastructures can help customers in industries to transform businesses and services successfully [22]. The combination of numerous sensors in wireless technologies can communicate with the Internet infrastructure and can collect information for different industrial applications. It is converting compatible data to interoperability between internet devices, appliances, and objects.

B. Telepresence

Telepresence can presentation where all employees can access their current operational status simultaneously, for convenient control and decision making by IoT applications. [21]. Thus, the services with the telepresence function would positively influence workers [23], and the performance of IoT devices in real-time from different locations at any time. Therefore, reliable telepresence helps consumers feel positive about their products and positive Internet experiences. The IoT System can improve management efficiency with more controllable and auditable management processes [24]. Synchronization of production systems between virtual and real systems depends on data from sensors and connected smart devices, and real-time data updates [25]. Connectivity automation of collaboration activities enables synchronization through IoT-enabled productivity [20]. It focuses on the synchronization of the production process in optimizing the production requirements [26]. Time and energy can be saved by using an internet device to synchronize things or products during the work process. IoT monitoring with dynamic system analysis and event processing integrates between device and business [20]. Communication between machines and connected machines provides information about the status of controlling or monitoring systems [27]. For example, the area of condition monitoring, predictive maintenance, estimating the remaining lifetime of the components, making decisions about urgent, short-term, and real-time actions [28]. Smart process monitoring is realized by connecting cloud computing and services the machine tools [29]. Internet of Things (IoT) dynamically controls and makes decisions, as well as the real-time monitoring and controlling of the manufacturing process [30]. The smart factory uses self-controlled machines and automated robots that have the module of adaptive intelligence to take necessary actions depending on pre-defined instructions [31]. It has an automatic device control system to continuously monitor performance in real-time anytime and anywhere. This data of smart factories will be available to humans or systems for the monitoring, prediction, and control of the production systems [32].

C. Intelligence

The intelligence of IoT is the capability to derive real-time data. The purpose of smart manufacturing into intelligence systems is to improve the positive production processes in all aspects [33]. Smart manufacturing information systems changed from digitization and networking to intelligence [34]. IoT applications must be embedded with intelligence to enhance devices that can monitor the environment, identify problems, communicate with each other, and potentially address problems without human assistance [35]. It more profoundly integrates manufacturing operations systems with communication, information, and intelligence technologies [36]. IoT devices are functional or personal assistants for organizations and are functions that control devices in real-time from different locations by connecting the internet devices of things to the employees [37]. Smart devices are focused on functional orientation, application, and analysis to be able to solve problems. The internet devices coordinate things, and function as personal assistants for the organization. Therefore, the service of this functional orientation will benefit employees. Challenges and applications of artificial intelligence are components of the concept of the Internet of Things. Relevant to Industry 4.0, it involves automation including artificial intelligence. The information system of smart manufacturing is based on the new technology of industrial artificial intelligence. Cloud Computing provides infrastructure, software, and platform as services [38]. Cloud computing and IoT are the main support technologies in Cloud manufacturing and can interact with each other in a cloud-based intelligent environment [39]. Therefore, cloud computing can continually access available data that requires manufacturing management applications and enables the effectiveness of processes.

D. Security

Security and privacy of sensitive data in Industrial IoT systems are necessary nowadays, and management trust is crucial for data security in IoT. [40]. Transparency, conflict of interests, data confidentiality, network security, and the safety of IoT devices are among the most important security and privacy problems. [41]. The security for industrial process measurement and control network and system security [42]. High-value industrial data can be harmed by a single click, potentially riskiness the entire process of industrial plants. Traceability and monitoring of systems are the results of real-time data management to prevent certain events or potential system failures [43]. As a result, higher-layer applications can avoid having to handle unneeded data and lower the danger of information privacy being compromised. Employees and businesses that share information between departments are protected by security, which encompasses information privacy, data management, and permission. [44]. It is the permission of employees and businesses when they communicate information, as well as a feature that allows an authorized user to search the database for specific information when needed. The ability to manage the information flow is known as data management. Information can be retrieved, integrated, and controlled using data management under the management service layer [43]. It is capable of handling large amounts of data and synchronizing operations in real-time.

E. Value

When connected devices can communicate with one another, the actual potential of the IoT for businesses may be completely realized. The connection of humans, parts, and systems creates self-guided, dynamic, real-time value-added interconnections across the value chain [45]. Several processes happen, and value is added to every process [46]. Improved product or service delivery, increased productivity, lower labor costs, lower energy usage, and a shorter build-to-order cycle are all examples of this benefit. There is also a need for Internet of Things integration in the operation of marketing activities, seasonal variations or market trends, and fluctuations in market demand or custom orders [28]. The benefits of the Internet of Things can be used to improve manufacturing processes and create new products. After receiving data from IoT devices, the organization can learn about some business concerns such as changes in client preferences or worker performance [34]. Similarly, The data, the convenience, and the ease with which it is collected and evaluated are the key drivers of this industrial revolution [47]. Using IoT devices or goods during the work process can help employees save time and energy [21]. [13]. It can also refer to services that help people manage their jobs more effectively and make quick judgments. The effectiveness of IoT enables stakeholders to make dynamic decisions based on real-time data [48]. It can manage the use of technology to fulfill company objectives [49]. The information could be utilized to assess different performance levels and provide statistical analysis reports. Data on machine tools will improve as a result of increased machine use, which can be used to improve overall performance [50].

TABLE I. INTERNET OF THINGS (IOT) APPLICATION WITH REFERENCES

Factors	Reference
Connectivity	[7], [18], [19], [20], [21], [22], [65], [66]
Telepresence	[21], [23], [25], [27], [28], [29], [30], [32], [43]
Intelligence	[24], [33], [34], [35], [36], [39], [53], [71]
Security	[40], [41], [42], [43], [44], [52], [69]
Value	[13], [21], [28], [34], [45], [46], [48], [50]

F. Factors Influence IoT Applications

IoT devices have been recognized as useful in a variety of situations and environments [51]. Furthermore, IoT could connect disparate networks while also addressing security concerns [52]. It is a technology and innovation that focuses on motion and environmental detection [53]. Therefore, the main purpose of IoT was to provide a that could accept, store, and transmit data over the Internet within computer devices including sensors, desktops, laptops, smartphones, and other electronic devices, some applications related to the IoT had provided value for consumers [54]. Moreover, these devices could potentially link to the Internet and keep track of personal and user data. The devices would then evaluate the data and make sensible recommendations to the users. From the literature review, IoT applications with references as the Table I and factors have been derived and reviewed to determine the factors that influence IoT applications for auto parts production as connectivity, telepresence, intelligence, security, and value which are shown in Fig. 1.

III. MATERIALS AND METHODS

A. Data Collection

This study was based on the quantitative research principle by using the survey method for data collection. The population of auto parts manufacturers came from the automotive institute and the institute identified that there were 720 auto parts firms. Subsequently, the sample size for this research was 88 samples as Yamane's formula [55]. The research tool in this study was a questionnaire that was developed based on factors of connectivity, telepresence, intelligence, security, and value as the framework shown in Fig. 1.

After the questionnaire was developed completely, the research distributed the questionnaire to 88 respondents sampled by snowball method who were the decision-makers in implementing IoT in automotive part manufacturers. At the beginning of this survey, five experts the auto parts manufacturers, which had deployed IoT applications in the auto parts production, were identified and called in for asking their permission to answer the questionnaire.

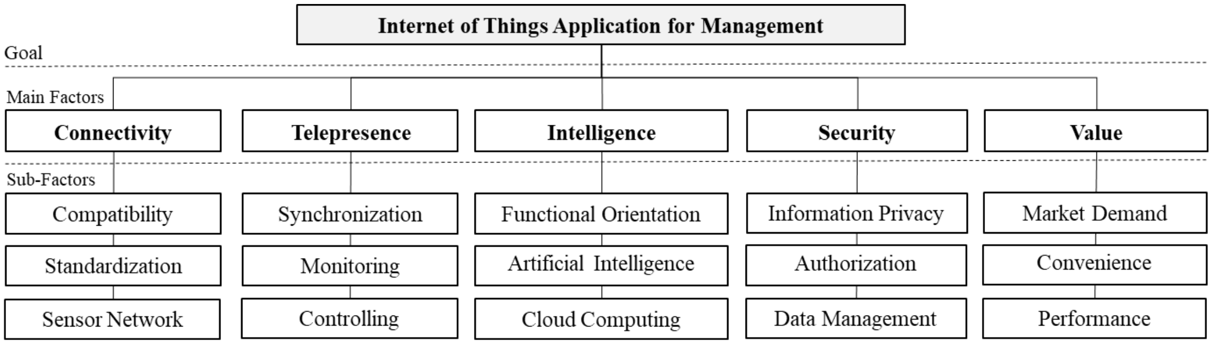


Fig. 1. Conceptual Framework of Internet of Things Application for Management.

B. Analytic Network Process

The analytic hierarchy process (AHP) is a well-known multiple criteria decision-making (MCDM) method that has been widely used in a variety of decision-making situations. The principle of AHP is to distinguish problems in a linear top-bottom format as a hierarchy. While the upper levels of the constructed hierarchy are independent of all lower levels, and the components in each level are likewise independent. However, some decision-making problems cannot be structured as a hierarchy, or some components have relations to other components in the structure. Thus, the analytic network process had been developed to address the mentioned problems of AHP and deal with dependencies between criteria and alternatives [57]. The components in the same level can be functionally dependent and the lower levels can reversely affect the upper level in each component. [58]. The advantages of ANP for considering dependencies between criteria and alternatives are listed as follows [59].

- The approach can deal with various, and complicated criteria.
- It supports decision-makers to build a decision-making model and to study relationships in the matrix form.
- It can evaluate both quantitative and qualitative criteria with the pairwise comparison technique to identify the importance of criteria.

Therefore, ANP is the appropriate MCDM approach than AHP for identifying and prioritizing factors that influence decision-makers in selecting IoT applications because the dependencies between each main factor and its sub-factors will be analyzed.

C. Fuzzy Analytic Network Process

1) Fuzzy set theory: To deal with situations involving uncertainty, imprecision, and vagueness, Zadeh (1965) created a fuzzy set theory. The capacity to represent ambiguous data is a major contribution of fuzzy set theory. [60]. The theory also permits mathematical operations and programming to be applied to the fuzzy domain. Generally, the fuzzy set is defined by a membership function, which represents the grade of any element x of X that has the partial membership to M. The value between zero and one determines the degree to which an element belongs to a set. So an element x belongs to $M, \mu_M(x) = 1$ and $\mu_M(x) = 0$ [56, 60, 61].

2) Triangular fuzzy number: A triangular fuzzy number (TFN) is presented as (l, m, u), where $l \leq m \leq u$. The membership function $\mu_M(x)$ of TFN is obtained by Equation (1).

$$\mu_M(x) = \begin{cases} (x-l)/(m-l) & l \leq x \leq m \\ (u-x)/(u-m) & m \leq x \leq u \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

The membership functions of triangular fuzzy numbers are used for the pairwise comparison of decision variables from “Very bad” to “Excellent”, and the middle preference values between them. The membership functions of TFNs, $M_i = (m_{i1}, m_{i2}, m_{i3})$, where $i = 1,2, \cdots, n$ and m_{i1}, m_{i2}, m_{i3} are the lower, middle, and upper values of the fuzzy number M_i, respectively [56, 61]. According to the concept of extent analysis, each object is taken and extent analysis for each goal g_i is performed, respectively. Therefore, the m extent analysis values for each object are obtained as the following signs [61]:

$$M_{g_i}^1, M_{g_i}^2, \cdots, M_{g_i}^m, \text{ i=1,2,}\cdots\text{,n} \tag{2}$$

where $M_{g1}^j (j = 1, 2, \ldots, m)$ are presented in triangular fuzzy numbers.

3) Fuzzy analytic network process: Fuzzy analytic network process (Fuzzy ANP) is a multiple criteria decision-making (MCDM) approach that uses both interdependences of criteria and inner dependences of criteria with the pairwise comparison matrix. The steps of Chang’s (1996) extent analysis can be given as follows [61, 62]. The value of fuzzy synthetic extent for the i^{th} object is defined as:

$$S_i = \sum_{j=1}^{m} M_{gi}^j \otimes \left[\sum_{i=1}^{n} \sum_{j=1}^{m} M_{g_i}^j\right]^{-1} \tag{3}$$

To obtain $\sum_{j=1}^{m} M_{gi}^j$, perform the fuzzy addition operation of m extent analysis values for a particular matrix such that:

$$\sum_{j=1}^{m} M_{gi}^j = \left(\sum_{j=1}^{m} l_j, \sum_{j=1}^{m} m_j, \sum_{j=1}^{m} u_j\right) \tag{4}$$

And to obtain $\left[\sum_{i=1}^{n} \sum_{j=1}^{m} M_{g_i}^j\right]^{-1}$, perform the fuzzy addition operation of $M_{g_i}^j (j = 1,2, \ldots, m)$ values such that:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} M_{g_i}^{j} = \left(\sum_{j=1}^{m} l_j, \sum_{j=1}^{m} m_j, \sum_{j=1}^{m} u_j\right) \tag{5}$$

Then the inverse of the vector in eq. (5) is calculated as follows:

$$\left[\sum_{i=1}^{n}\sum_{j=1}^{m} M_{gi}^{j}\right]^{-1} = \left(\frac{1}{\sum_{i=1}^{n} u_i}, \frac{1}{\sum_{i=1}^{n} m_i}, \frac{1}{\sum_{i=1}^{n} l_i}\right) \tag{6}$$

The degree of possibility of $M_2 = (l_2, m_2, u_2) \geq M_1 = (l_1, m_1, u_1)$ is defined as:

$$V(M_2 \geq M_1) = \sup[\min(\mu_{M_1}(x), \mu_{M_2}(y)] \tag{7}$$

and can be equivalently expressed as follows:

$$V(M_2 \geq M_1) = \begin{cases} 1 & \text{if } m_2 \geq m_1, \\ 0 & \text{if } l_1 \geq u_2, \\ \frac{l_1 - u_2}{(m_2 - u_2) - (m_1 - l_1)} & \text{otherwise} \end{cases} \tag{8}$$

The degree of possibility for a convex fuzzy number is greater than k convex fuzzy numbers $M_i (i = 1,2, \ldots, k)$ that can be defined by:

$$V(M \geq M_1, M_2, \ldots, M_k) = \min V(M \geq M_i), \quad i = 1, 2, \ldots, k \tag{9}$$

Assume that:

$$d(S_i) = \min V(S_i \geq S_k), \quad k = 1, 2, \ldots, n; k \neq i \tag{10}$$

Then the weights vector is calculated by:

$$W' = (d(S_1), d(S_2), \ldots, d(S_n))^T \tag{11}$$

And weight vectors are normalized as follows.

$$W = (d(A_1), d(A_2), \ldots, d(A_n))^T \tag{12}$$

Where W is a real number, and $A_i = (i = 1, 2, \ldots n)$ are n elements.

IV. Results

This paper, Internet of Things application is divided into five categories consisting of Connectivity, Telepresence, Intelligence, Security and Value. Each category contains several IoT applications for management which are shown in Fig. 1. This study proposes the Internet of Things application for management related to the auto parts manufacturers in Thailand. This section can be divided into six steps to evaluate the industry. The first step is to determine the Internet of Things framework of main factors and sub-factors, which is performed using a questionnaire survey that uses the fuzzy ANP approach to calculate the weights of the factors and sub-factors. The application is based on the methods outlined in the preceding part and is explained step by step, along with the outcomes.

Step 1. Determine the fuzzy scale for the important weight of factors and sub-factors from a decision-maker with the responsibility for the Internet of Things application management in a company. Choose appropriate linguistic variables for the relative weight of the factors as defined by Table II [61,72] prepare a pair-wise comparison to each factor from expert opinion by using the linguistic scale.

Step 2. Conduct the questionnaire survey for exploring internet of things application for management in practices by collecting data from 88 respondents in auto parts companies.

Step 3. Calculate the consistency ratio (CR) by Equation (13).

$$CR = \frac{CI}{RI} \tag{13}$$

When $CI = \frac{(\lambda_{max} - n)}{(n-1)}$ and $\lambda_{max} = \sum_{i=1}^{n}\left[\sum_{j=1}^{n} a_{ij} W_j\right]$. The CR value should be lower than 0.1 which means the weights are determined by decision-makers have consistency. If the CR value is greater than 0.1, it expresses those weights determined to have no consistency. Therefore, the research should review the evaluation of main factors and sub-factors in the first step.

Step 4. Calculate the average fuzzy evaluation matrix from the questionnaire survey and the local weights of the main factors and sub-factors as shown in Tables III to VIII.

Step 5. Calculate the inner dependence weights and the dependencies among the factors that are considered the main factors as shown in Table IX.

Step 6. Calculate the global weights of sub-factors by multiplying the local weights of the sub-factors with the interdependent weights of main factors as shown in Table X.

From the results, the CR ratios of five main factors and fifteen sub-factors were lower than 0.1 which expressed weights determined by decision-makers of auto parts companies had consistency. In this study, the decision-makers made the pairwise comparison for main factors and sub-factors and they were analyzed by fuzzy ANP. The results indicated that telepresence was the most important factor with the interdependent weight at 22.9 percent. Next, connectivity was the second rank of the important factor with a weight of 20.5 percent. While value, security, and intelligence had interdependent weights at 19.7, 18.7, and 18.2 percent respectively.

For the importance of sub-factors, the results indicated that performance was the most important sub-factor that influenced the decision-makers with the global weight at 9.20 percent. The second rank was the functional orientation with a global weight of 8.80 percent, while data management, controlling, and compatibility had the global weights of 8.60, 8.30, and 8.28 percent respectively. For the least importance, the bottom five sub-factors had consisted of information privacy, cloud computing, authorization, artificial intelligence, and market demand with their global weights at 5.88, 5.41, 4.30, 3.91, and 2.84 percent respectively. Therefore, these sub-factors had less impact on the decision-maker during selection to apply IoT in their production management.

TABLE II. LINGUISTIC SCALE FOR RELATIVE WEIGHTS OF FACTORS

Risk Level	Triangular Fuzzy Scale	Triangular Fuzzy Reciprocal Scale
Just equal	(1, 1, 1)	(1, 1, 1)
Equally important	(1/2, 1, 3/2)	(2/3, 1, 2)
Weakly more important	(1, 3/2, 2)	(1/2, 2/3, 1)
Strongly more important	(3/2, 2, 5/2)	(2/5, 1/2, 2/3)
Very strongly more important	(2, 5/2, 3)	(1/3, 2/5, 1/2)
More important	(5/2, 3, 7/2)	(2/7, 1/3, 2/5)

TABLE III. LOCAL WEIGHT AND PAIR-WISE COMPARISON MATRIX OF "MAIN FACTORS"

Main Factors	Connectivity	Telepresence	Intelligence	Security	Value	Weights
Connectivity	1.000, 1.000, 1.000	0.876, 1.209, 1.563	0.996, 1.344, 1.731	1.086, 1.398, 1.742	0.665, 1.029, 1.537	0.206
Telepresence	0.789, 1.010, 1.389	1.000, 1.000, 1.000	1.054, 1.385, 1.811	1.069, 1.414, 1.786	1.044, 1.329, 1.744	0.234
Intelligence	0.770, 1.039, 1.408	0.667, 0.880, 1.185	1.000, 1.000, 1.000	0.688, 0.974, 1.311	0.869, 1.216, 1.593	0.179
Security	0.948, 1.224, 1.547	0.710, 0.894, 1.163	1.076, 1.415, 1.835	1.000, 1.000, 1.000	0.754, 1.098, 1.473	0.184
Value	0.759, 1.128, 1.773	0.722, 0.897, 1.138	0.759, 0.983, 1.324	0.923, 1.255, 1.864	1.000, 1.000, 1.000	0.198

TABLE IV. LOCAL WEIGHT AND PAIR-WISE COMPARISON MATRIX OF "CONNECTIVITY"

Sub-factors	Compatibility	Standardization	Sensor network	Weights
Compatibility	1.000, 1.000, 1.000	1.221, 1.624, 2.059	0.982, 1.269, 1.572	0.404
Standardization	0.633, 0.839, 1.112	1.000, 1.000, 1.000	0.760, 1.121, 1.623	0.287
Sensor network	0.781, 0.947, 1.195	0.723, 1.063, 1.574	1.000, 1.000, 1.000	0.309

TABLE V. LOCAL WEIGHT AND PAIR-WISE COMPARISON MATRIX OF "TELEPRESENCE"

Sub-factors	Synchronization	Monitoring	Controlling	Weights
Synchronization	1.000, 1.000, 1.000	0.828, 1.202, 1.638	0.828, 1.167, 1627	0.345
Monitoring	0.706, 1.004, 1.458	1.000, 1.000, 1.000	0.619, 0.835, 1.223	0.293
Controlling	0.782, 1.108, 1.515	0.985, 1.424, 1.879	1.000, 1.000, 1.000	0.363

TABLE VI. LOCAL WEIGHT AND PAIR-WISE COMPARISON MATRIX OF "INTELLIGENCE"

Sub-factors	Functional Orientation	Artificial Intelligence	Cloud Computing	Weights
Functional Orientation	1.000, 1.000, 1.000	1.367, 1.814, 2.364	1.133, 1.551, 2.011	0.486
Artificial Intelligence	0.507, 0.693, 0.938	1.000, 1.000, 1.000	0.679, 0.992, 1.441	0.215
Cloud Computing	0.567, 0.772, 1.061	0.887, 1.256, 1.739	1.000, 1.000, 1.000	0.298

TABLE VII. LOCAL WEIGHT AND PAIR-WISE COMPARISON MATRIX OF "SECURITY"

Sub-factors	Information Privacy	Authorization	Data Management	Weights
Information Privacy	1.000, 1.000, 1.000	0.952, 0.303, 1.761	0.652, 0.900, 1.239	0.314
Authorization	0.693, 0.954, 1.294	1.000, 1.000, 1.000	0.539, 0.706, 0.979	0.230
Data Management	1.131, 1.554, 2.059	1.287, 1.744, 2.212	1.000, 1.000, 1.000	0.456

TABLE VIII. LOCAL WEIGHT AND PAIR-WISE COMPARISON MATRIX OF "VALUE"

Sub-factors	Marketing demand	Convenience	Performance	Weights
Marketing Demand	1.000, 1.000, 1.000	0.578, 0.752, 1.023	0.557, 0.726, 0.970	0.144
Convenience	1.397, 1.839, 2.305	1.000, 1.000, 1.000	0.663, 0.853, 1.222	0.390
Performance	1.330, 1.773, 2.231	0.917, 1.304, 1.705	1.000, 1.000, 1.000	0.466

TABLE IX. INNER DEPENDENCE WEIGHT OF THE FACTORS

Factors	Connectivity	Telepresence	Intelligence	Security	Value
Connectivity	1.000	0.255	0.254	0.258	0.264
Telepresence	0.310	1.000	0.286	0.278	0.294
Intelligence	0.223	0.240	1.000	0.228	0.205
Security	0.237	0.247	0.208	1.000	0.237
Value	0.231	0.259	0.252	0.236	1.000

TABLE X. TOTAL WEIGHT OF MAIN FACTORS AND SUB-FACTORS

Factors	Interdependent weights	Ranking	Sub-factors	Local weights	Ranking	Global weights	Ranking
Connectivity	0.205	2	Compatibility Standardization Sensor network	0.404 0.287 0.309	1 3 2	0.083 0.059 0.063	5 10 9
Telepresence	0.229	1	Synchronization Monitoring Controlling	0.345 0.293 0.363	2 3 1	0.079 0.067 0.083	6 8 4
Intelligence	0.182	5	Functional Orientation Artificial Intelligence Cloud Computing	0.486 0.215 0.298	1 3 2	0.088 0.039 0.054	2 14 12
Security	0.187	4	Information Privacy Authorization Data Management	0.314 0.230 0.456	2 3 1	0.059 0.043 0.086	11 13 3
Value	0.197	3	Market Demand Convenience Performance	0.144 0.390 0.466	3 2 1	0.028 0.077 0.092	15 7 1

V. DISCUSSION

Results from this study have indicated that the importance level of each variable is different by the factor of telepresence is the most important because there is telepresence for considering proposed data in working units that are connected to IoT for tracking operations in the production process. The data should be presented in real-time to executives and employees to track circumstances in production more conveniently. Moreover, it has impacted the controlling, decision-making, and planning that is done more efficiently. The second factor is connectivity which is important for integrating the production process of auto parts to control processes connected. Furthermore, the connectivity is for processing and exchanging data with suppliers, production process, and warehouse management following Just-In-Time production for more flexibility and accuracy. For the performance factor, it can be connected to IoT which is important in controlling and tracking operations more quickly. The accuracy of data enhances forecasting and decision-making, as well as the ability to improve working units throughout the supply chain, following specified plans. The function orientation is a crucial part that must control flows of produced auto parts continuously throughout the process and lay down guidelines with relevant departments both inside and outside the organization. Besides, data that is connected to IoT can improve function orientation more efficiently. Finally, data management is for managing big data from IoT through a data center that relates to internal and external networks that enhance efficient operations in the production process of auto parts. However, the results can be discussed in detail for each factor as follows.

A. Telepresence

The telepresence can consolidate and analyze data for presenting reports from several perspectives as requirements and scopes of each department such as the productivity report of each production line versus production plan or failure reports. The reports are for relevant departments can receive information and prepare for any case from the defined plan delayed. It can be summarized and proposed to executives as management dashboards on important issues for industrial management based on changing circumstances or needs. It can also be used as information for each department to have a picture of their performance or service levels of others to improve their processes more efficiently [63], such as presenting information in the production process, auto parts approval data for delivery, notifying customers, or forecasting stock more precisely. Therefore, the presentation of data enables all relevant segments to evaluate and track actual production reports against targets in real-time.

1) Controlling: Operating processes following targets or defined plans is necessary to ensure continuity and accuracy in the production of auto parts for stable working throughout production operations, enabling dynamic control, and situational assessment for data analysis in decision-making more quickly and anywhere [30, 64]. Besides, controlling through several Programmable Logic Control (PLC) systems can deal with machines that have complex functions in the production line of various auto parts to ensure that operations are in line with the production plan. Therefore, it expresses that the automated control of the production system supports operations through the working steps, controlling robots to comply with requirements is for operating continuous production, such as plating and painting robots that are controlled to work relative to the movement of parts on production belts, or robots at the packing belts which must be controlled to operate accurately and precisely as the determined functions to meet the customer orders. Additionally, robots are controlled to welt joints of auto parts as specified standards as well as control interaction or transmission environments between devices.

2) Synchronization: The auto parts manufacturing process is for connecting data to work consistently throughout the supply chain, such as ordering, controlling, or transmitting data without any interruption in the working process. Therefore, all units must precisely synchronize data into the central operating system for a holistic view of all productions. As a result, management is more convenient and faster [25, 26]. Process synchronization can create balance in production lines and units because the production of auto parts includes the main production line for assembly parts and the second production line for assembly sub-parts as well as parts from suppliers that are a part of the production.

3) Monitoring: The manufacturing process of auto parts is crucial due to the large number of components and suppliers in the supply chain. It can dynamically monitor and perform real-time analysis efficiently [20]. Therefore, applying the internet of things is to continuously monitor operations of each unit on the production line, such as monitoring the assembly process to track operations as planned, monitoring the operating systems for avoiding potential risks and system failures, or doing preventative maintenance for machines, etc.

B. Connectivity

Connectivity is a crucial integration in the production of auto parts for control of various stages to detect machines and equipment for the exchange of data in analytical processing [65]. Particularly, the auto parts industry produces parts with a Just-In-Time production system to meet the needs of the production line in terms of quantity, time, and quality. Therefore, data connections with suppliers have real-time access to various information, and coordination with suppliers is streamlined and accurate [66]. Enable efficient operation, such as planning production requirements and delivering raw material parts in due course, managing raw materials in warehouses to be alerted reducing raw material shortages, and reliability of the delivery of raw material parts, etc.

1) Compatibility: The Auto parts manufacturing process is for centralizing data of operational statuses quickly and up to date. Particularly, several parts produced from different suppliers need to be linked with the same platform [6], for example, linking customer production data to assemble a vehicle that lets suppliers know what materials to produce or order during what period. In a part of compatibility for quality inspection data of vehicle chassis parts, when it is compared to the customer-specified standards, assembling auto parts that must match the production plan according to the specified model. In case of the product does not meet standards, it will inform details of errors for revisions, etc.

2) Sensor network: Linking sensors to computer systems in the same network is to communicate with each other inside and outside the organization [22,67]. This method enables data transfer, controlling and tracking working statuses, or detecting errors that occur in the production process, such as errors from machines in the manufacture of auto parts or detecting parts that do not meet the standard or track statuses in the logistic process effectively. Besides, the sensor network can build trust by connecting external sensor networks to auto parts suppliers for coordination through the enterprise resource planning system (ERP) that allows data from suppliers to be centralized into the data center such as production planning, quality management of parts, purchasing and delivery statuses, disbursement of raw materials, etc. The logistic management in the transportation of auto parts that are linked to the sensor network can control and trace the services by connecting to the Global Positioning System (GPS). Furthermore, it can check the driver's status and physical condition for safety and efficiency purposes.

3) Standardization: Standards for auto parts manufacturers must build confidence throughout the supply chain process [21]. Effective control under the same standards, such as standardized performance of parts, is created and delivered to both inside and outside involved units through the correct digital system as designed. It can be connected to various measurements to detect non-compliant auto parts, such as checking the coating standard of all standard vehicle parts in terms of thickness and shades. It must be precise checking because the color is very close, or the traceability process of auto parts (vehicles have a wide range of different models and parts, there must be standards for tracking the production process and sources), etc.

C. Value

Creating value in the auto parts manufacturing industry can create value-added through tracking, recording, and data processing. Exchange of information and display insights that enable more accurate and faster decision making by enabling organizations to dynamically recognize the device connection that can communicate [45] such as linking data from suppliers, production, logistic system, and services. It supports the ability of data analysis throughout the production process or can operate automated systems that are linked to production plans in the production process of auto parts. Therefore, costs have been reduced from removing wastes in the production and increasing value-added in operations. Moreover, decision-making and strategies can be appropriately operated from the quality data.

1) Performance: The basic need of the auto parts industry to operate. Using an appropriate production factor and reducing waste in the working process can be done. IoT is a tool that can support operations work more accurately and quickly as well as reduce waste of time and errors [50, 68]. Moreover, tracking devices and connecting data can help to produce quality auto parts that meet the customer's needs, data from IoT can evaluate performances in departments for improving their performances consistently.

2) Convenience: The access insights that are analyzed and processed, can be used in operating the quality production because IoT can create convenience in terms of data collection from resources [47, 65]. It can facilitate operators or executives in working such as convenience in tracking real-time performances of different product lines at the same time, convenience in tracking the shipment of auto parts from suppliers, and convenience in trust-building to the customers by enabling tracking of production information and services.

3) Market demand: The application in marketing helps auto parts manufacturers to get the correct data on customer needs [28, 34]. The auto parts can be adjusted to meet the customer's needs as well as quality standards or other requirements that the customers accept. Change circumstances and market volatility through devices and data networks can exchange the data with customers in perspectives such as quality issues, customer complaints about quality issues, and production plans changing. It occurs from the data exchange in the same operating system including determining the delivery cycle to meet the just-in-time system or anticipating new products developed.

D. Security

Information security plays a crucial role which restricting access to only relevant data to prevent espionage from unauthorized persons [40, 69]. The auto parts industry because data in business operations must be kept confidentially for the highest benefit of the organization, particularly the auto parts industry has a large network of people involved. Therefore, permissions of data access and data management require a security system to prevent data leakage from employee or system errors such as production data, logistics data, customer service data, etc.

1) Data management: The auto parts industry where data collection is dynamically changing. Center data management must have the capability to deal with big data, filter standard data, and manage data more efficiently [70]. Therefore, it is necessary to have an efficient storage system for managing data formats, data display, and managing data as the operation characteristic to prevent failures of the system that may occur from various specifications or productions from different suppliers. Besides, failures can be occurred by capability data of different production lines, data of the automated system, or robots that need data characteristics in terms of the readable language by robots.

2) Information privacy: Information privacy is a challenge in the development and services of IoT solutions that require strict compliance with standards of data privacy, access statuses of using devices in production, encrypting, and controlling data under the security policy [43]. including reducing the risk of disclosure because the industry has more trade secrets such as production technology, product details, and features as well as developing new products in the future. The considered information privacy management is the controlling of connecting between devices in both internal and external organizations. Preventive management that maintains the confidentiality of process data in the supply chain access the secured cloud system in connecting production data.

3) Authorization: Data access in levels of departments is an important part that must be operated consistently with information privacy to avoid obstacles to access to information excluded the permission for practitioners to access self-information. In the industry, the cross-function team can work with other employees from other departments in terms of developing new products that should have the team from R&D, marketing, production, or suppliers. Therefore, people in the team should be able to access information about each other. Therefore, assigning permissions to temporarily grant access to the necessary information of other entities but there must be a controller of authority to access only specified data for preventing data security that may impact the total production of auto parts such as standard parameters in the quality inspection or accessibility of trade secrets. The importance of accessing sensitive information can prevent problems or adverse events in the organization.

E. Intelligence

Intelligence is directly involved in the production of auto parts that uses IoT to interact between machine and machine, or machine and human to understand patterns or behaviors of wise operations. Furthermore, It can adjust production systems of auto parts to digital systems to increase their intelligence [24]. Ssuch as intelligence in evaluation and analysis for the volume of produced auto parts in each production line and using data to measure performances as planned against the past data as well as forecasting its trend. Intelligence in simultaneous controlling complex machines and equipment in the main and second production lines is for the operations can be traced and controlled more efficiently.

1) Functional orientation: The auto parts production that can be applied to use IoT for providing information that is beneficial to operators can work accurately and make decisions efficiently. Furthermore, functional orientation for tracking auto parts from suppliers or customers can be connected to logistics processes for continuous monitoring [65]. This emphasizes the exact time of delivery according to the specified cycle [24]. The functional orientation has been used to align works and be an assistant for the organization in multi-tasks such as connecting ERPs in the supply chain, determining collaborative networks clearly, and set-up clear process flows. Besides, automated data recording with RFID can check the status of auto parts from suppliers and the production more effectively that expresses production data and quality of raw material for planning.

2) Cloud computing: The auto parts industry is a service that provides an external IT infrastructure for connecting software and platforms that can process big data from IoT to respond to intelligent processing that can access data anytime and anywhere. This can respond to the needs of customers and related parties more comprehensively [34,71]. Internal and external organizations under collaboration have a data center to access data conveniently such as the production report of auto parts through the organization’s web browser and software and all data will be simultaneously kept on the service provider’s server. Therefore, the user can access the system all time the most services are the type of web applications to support data-driven. When there is sufficient data for creating innovation in the manufacturing process of auto parts.

3) Artificial intelligence: Artificial Intelligence (AI) is a cutting-edge technology that has changed the way of working for the auto parts industry and has been involved with IoT directly. AI also helps to increase customer satisfaction by analyzing big data from the customer relationship management (CRM) system that supports continuous production and creates value-added for the organization [34, 71]. It supports the industry to operate with complex matters more quickly such as quality inspection of auto parts by running a simulation of used auto parts in a real situation instead of real inspections that spends more time and resources than the simulation. Besides, it can propose the best alternative to the operators to consider and make decisions more precisely such as selecting the best route for transportation purposes or predicting possible incidents that may occur in the production line in advance. Moreover, AI can be developed into an intelligent operating system that can interact with humans, intelligent production that can predict or notify any incident in the production, and planning and inspection for auto parts to be stocked and utilize spaces in the warehouse more effectively. With another application, raw materials in the process can be scanned to trigger the automated reimbursement system to feed raw materials into the production line continuously and support the non-stop production. Therefore, the production data can be transferred to other departments such as planning, purchasing, production, and marketing departments to know the statuses in production at the same time and use the data to plan future operations more precisely.

VI. Conclusion

In this research, five factors have been identified that influence the auto parts industry for applying Internet of Things (IoT) technology to integrate production management with connectivity, telepresence, intelligence, security, and value respectively. The study has revealed that telepresence is the most important factor for decision-makers of auto parts manufacturers in selecting IoT applications for managing production. While the importance of connectivity, value, security, and intelligence are prioritized in descending order respectively. For the importance of sub-factors, the top five ranks have consisted of performance, functional orientation, data management, controlling, and compatibility respectively which influence the auto parts manufacturers to decide for applying IoT. Meanwhile, information privacy, cloud computing, authorization, artificial intelligence, and market demand have less influence on decision-makers for selecting IoT to apply to their auto parts production. To the results of this study, auto parts manufacturers can make use of factors and sub-factors as selection criteria for procuring internet devices or relevant equipment to IoT to improve their production management as well as the firm's performances in terms of financial and production benefits.

In implications for management, this research benefits executive (Decision-makers) in auto parts manufacturers who are seeking IoT applications to improve production capabilities to increase value-added and competitive advantages. The first criteria that they need to consider when selecting IoT applications are telepresence. Telepresence will assist them in controlling production lines to guarantee that production capabilities meet the objective, and all data from each unit in the production lines should be integrated for daily performance monitoring and early detection of any events that could stop production. Connectivity is the second criterion that needs to consider for selecting IoT applications. Decision-makers must ensure that IoT applications are compatible with their machinery and equipment so that they can be controlled smoothly and precisely, and they are also connected to the existing computer network. Moreover, considered IoT applications should have a standard that is both trustworthy and market acceptance. Next, the values that decision-makers should consider, are to improve performance in production, be convenient to use, and be desired by end-users. Therefore, if consumers can perceive the value and benefits of IoT applications, they will be more likely to accept them, and their obstacles to using them will be greatly lowered. Security is the next criterion to consider in the selection of IoT applications. To avoid the risk of data disclosure or leaking, IoT applications should have an efficient data storage system to manage sent data from devices, as well as a data privacy standard. Besides, data access should be set with authorization to avoid unauthorized access to personal information or commercial secrets. Finally, intelligence is considered the least important criteria for selecting IoT applications because most auto parts manufacturers in use manpower in production. Moreover, the industry still needs skilled manpower for producing high-quality auto parts. Intelligence may be necessary for the firms that are looking for cost reduction and automated processes. These firms want to expand their capacity and productivity to supply auto parts for large automotive firms with increased demand.

The auto parts manufacturers who have fully implemented and used IoT applications in their manufacturing lines are limited to very large enterprises, which is a limitation of this study. Therefore, small and medium-sized auto parts manufacturers may choose to integrate some IoT applications in their production as a way to save money. For future research, these factors in this study will be used as the selection criteria to evaluate and select appropriate IoT applications for managing auto parts production.

References

[1] A. J. C. Trappey, C. V. Trappey, U. H. Govindarajan, A. C. Chuang., J. J. Sun, "A review of essential standards and patent landscapes for the Internet of Things," Advanced Engineering Informatics. Vol. 33, pp. 208-229, August 2017.

[2] N. G. P. Carvalho, O. Chaim, E. Cazarini, M. C. Gerolamo, "Manufacturing in the fourth industrial revolution: A positive prospect in sustainable manufacturing," Procedia Manufacturing. Vol. 21, pp. 671-678, January 2018.

[3] P. Lou, Q. Liu, Z. Zhou, H. Wang, "Agile supply chain management over the internet of things," IEEE international conference on management and service science, August 2011.

[4] L. Atzori, A. Iera, G. Morabito, "The internet of things: A survey," Computer networks. Vol. 54, pp. 2787-2805, May 2010.

[5] K. N. Qureshi, S. Din, G. Jeon, F. Piccialli, "Link quality and energy utilization based preferable next hop selection routing for wireless body area networks. Computer Communications, Vol. 155, pp. 382-392, January 2020.

[6] K. K. Patel, S. M. Patel, "Internet of Things-IOT: Definition, Characteristics, Architecture, Enabling technologies, Application & Future Challenges," International journal of engineering science and computing, Vol. 6, pp. 6122-6131, April 2016.

[7] S. Gierej, The Framework of Business Model in the Context of Industrial Internet of Things. Procedia Engineering, Vol. 182, pp. 206-212, April 2017.

[8] J. C. Bendul, H. Blunck, "The design space of production planning and control for industry 4.0," Computers in Industry. Vol. 105, pp. 260-272, February 2019.

[9] S. Seuring, C. Brix-Asala, R. U. Khalid, "Analyzing base-of-the-pyramid projects through sustainable supply chain management," Journal of Cleaner Production. Vol. 212, pp. 1086-1097, March 2019.

[10] M. Ghobakhloo, "Industry 4.0, digitization, and opportunities for sustainability," Journal of cleaner production. Vol. 252, pp. 119-869 April 2020.

[11] K. P. Lin, M. L. Tseng, P. F. Pai, "Sustainable supply chain management using approximate fuzzy DEMATEL method," Resources, Conservation and Recycling. Vol. 128, pp. 134-142, January 2018.

[12] C. Liu, W. Cai, S. Jia, M. Zhang, H Guo, L. Hu, Z. Jiang, "Emergy-based evaluation and improvement for sustainable manufacturing systems considering resource efficiency and environment performance," Energy conversion and management. Vol. 177, pp. 176-189, December 2018.

[13] I. S. Jawahir, R. Bradley, "Technological elements of circular economy and the principles of 6R-Based Closed-loop Material Flow in Sustainable Manufacturing," Procedia Cirp, Vol. 40, pp. 103-108, February 2016.

[14] A. D. Jayal, F. Badurdeen, O. W. Dillon Jr, I. S. Jawahir, "Sustainable manufacturing: Modeling and optimization challenges at the product, process and system levels," CIRP Journal of Manufacturing Science and Technology, Vol. 2, No. 3, pp. 144-152, April 2010.

[15] J. C. Trappey, C. V. Trappey, U. H. Govindarajan, A. C. Chuang, J. J. Sun, "A review of essential standards and patent landscapes for the Internet of Things," Advanced Engineering Informatics, Vol. 33, pp. 208-229, August 2017.

[16] C. Katsma, H. Moonen, J. Hillegersberg, "Supply Chain Systems Maturing Towards the Internet-of-Things: A Framework," 24th Bled eConference, June 2011.

[17] D. Mourtzis, S. Fotia, N. Boli, E. Viachou, "Modelling and quantification of industry 4.0 manufacturing complexity based on information theory: a robotics case study," International Journal of Production Research, Vol. 57, pp. 6908-6921, Jan 2019.

[18] D. Chen, G. Doumeingts, F. Vernadat, "Architectures for enterprise integration and interoperability: Past, present and future," Computers in Industry, Vol. 59, pp. 647-659, September 2008.

[19] L. Uden, W. He, "How the Internet of Things can help knowledge management: a case study from the automotive domain," Journal of Knowledge Management, Vol. 21, pp. 57-70, February 2017.

[20] M. Al-Talib, W. Y. Melhem, A. I. Anosika, J. A. G. Reyes, S. P. Nadeen, A. Kumar, "Achieving resilience in the supply chain by applying IoT technology," Procedia CIRP, Vol. 91, pp. 752-757, August 2020.

[21] P. T. M. Ly, W. H. Lai, C. W. Hsu, F. Y. Shih, "Fuzzy AHP analysis of Internet of Things (IoT) in enterprises," Technological Forecasting and Social Change, Vol. 136, pp. 1-13, Novenber 2018.

[22] J. A. Luis, J. A. G. Galan, F. G. Bravo, M. S. Raya, J. A. Espigado, P. M. T. Rovira, "An Efficient Wireless Sensor Network for Industrial Monitoring and Control," Sensors, Vol. 18, No, 1, pp. 182, January 2018.

[23] P. P. Ray, "A survey on Internet of Things architectures" Journal of King Saud University-Computer and Information Sciences, Vol. 30, No.3, pp. 291-319, July 2018.

[24] W. Wu, C. Cheung, S. Y. Lo, R. Y. Zhong, G. Q. Huang, "An IoT-enabled Real-time Logistics System for A Third Party Company: A Case Study. Procedia Manufacturing, Vol. 49, pp. 16-23, July 2020.

[25] M. Garetti, P. Rosa, S. Terzi, "Life Cycle Simulation for the Design of Product-Service Systems," Computers in Industry, Vol. 63, No. 4, pp. 361-369, May 2012.

[26] M. Tavana, V. Hajipour, S. Oveisi, "IoT-based enterprise resource planning: Challenges, open issues, applications, architecture, and future research directions," Internet of Things, Vol. 11, pp. 100-262, September 2020.

[27] D. Kiel, C. Arnold, K. I. Voigt, The influence of the Industrial Internet of Things on business models of established manufacturing companies-A business level perspective. Technovation, Vol. 68, pp. 4-19, December 2017.

[28] M. Ebrahimi, A. Baboli, E. Rother, "The evolution of world class manufacturing toward Industry 4.0: A case study in the automotive industry," IFAC-PapersOnline, Vol. 52, No. 10, pp. 188-194, November 2019.

[29] A. Caggiano, T. Segreto, R. Teti, "Cloud Manufacturing Framework for Smart Monitoring of Machining," Procedia CIRP, Vol. 55, pp. 248-253. November 2016.

[30] S. Keivanpour, D. A. Kadi, "The Effect of "Internet of Things" on Aircraft Spare Parts Inventory Management," IFAC-PapersOnLine, Vol. 52, No. 13, pp. 2343-2347, December 2019.

[31] S. Wang, J. Wan, D. Zhang, D. Li, C. Zhang, "Towards smart factory for industry 4.0: a self-organized multi-agent system with big data based feedback and coordination," Computer networks. Vol. 101, pp. 158-168, May 2016.

[32] Y. Chen, "Integrated and Intelligent Manufacturing: Perspectives and Enablers," Engineering, Vol. 3, No. 5, October 2017.

[33] P. O. Donovan, K. Leahy, K. Bruton, D.T.J. O'Sullivan, "An industrial big data pipeline for data-driven analytics maintenance applications in large-scale smart manufacturing facilities," Journal of Big Data, Vol. 2, No.1, pp. 1-26, November 2015.

[34] X. Zhang, M. Xinguo, "An implementation for Smart Manufacturing Information System (SMIS) from an industrial practice survey," Computers & Industrial Engineering. Vol. 151, pp. 106938, February 2021.

[35] B. S. Sahay, J. Ranjan, "Real time business intelligence in supply chain analytics," Information Management & Computer Security. Vol. 16, No. 1, March 2008.

[36] Y. Wang, M. S. Ma, J. H. Yang, K. S. Wang, "Industry 4.0: a way from mass customization to mass personalization production," Advances in Manufacturing," Vol. 5, pp. 311-320, November 2017.

[37] J. Wu, J. Chen, W. Dou, "The Internet of Things and interaction style: the effect of smart interaction on brand attachment," Journal of Marketing Management, Vol. 33, pp. 61-75, September 2017.

[38] A. A. F. Saldivar, Y. Li, W. Chaen, Z. Zhan, J. Zhang, L. Y. Chen, " Industry 4.0 with cyber-physical integration: A design and manufacture perspective," IEEE, 21st international conference on automation and computing (ICAC). November 2015.

[39] R. Y. Zhong, S. Lan, C. Xu, Q. Dai, "Visualization of RFID-Enabled Shopfloor Logistics Big Data in Cloud Manufacturing," The International Journal of Advanced Manufacturing Technology, Vol. 8, pp. 5-16, August 2016.

[40] Y. Zhang, P. Zhang, A. V. Vasilakos, "A survey on trust management for Internet of Things," Journal of network and computer applications, Vol. 42, pp. 120-134. June 2014.

[41] A. K. Mohammadzadeh, S. Ghafoori, A. Mohammadian, R. Mohammadkazemi, B. Mahbanooei, R. Ghaemi, "A Fuzzy Analytic Network Process (FANP) approach for prioritizing internet of things challenges in Iran," Technology in Society, Vol. 53, pp. 124-134. May 2018.

[42] S. Li, L. D. Xu, S. Zhao, "5G Internet of Things: A survey," Journal of Industrial Information Integration. Vol. 10, pp. 1-9. June 2018.

[43] A. Gilchrist, Industry 4.0: The Industrial Internet of Things. Springer, 2016.

[44] Y. Chang, X. Dong, W. Sun, "Influence of Characteristics of the Internet of Things on Consumer Purchase Intention," Social Behavior and Personality An International Journal, Vol. 42, pp. 321-330, March 2014.

[45] D. Gorecky, M. Schmitt, M. Loskyll, D. Zuhlke, Human-machine-interaction in the industry 4.0 era. 12th IEEE international conference on industrial informatics (INDIN). July 2014.

[46] Z. Pang, Q. Chen, W. Han, L. Zheng, "Value-centric design of the internet-of-things solution for food supply chain: Value creation, sensor portfolio and information fusion," Information Systems Frontiers, Vol.17, pp. 289-319, August 2015.

[47] S. L. Molnar, N. Mohamed, "Reliability Assessment in the Context of Industry 4.0: Data as a Game Changer," Procedia Computer Science, Vol. 151, pp. 691-698, May 2019.

[48] M. Rezaei, M. A. Shirazi, B. Karimi, "IoT-based framework for performance measurement: A real-time supply chain decision alignment," Industrial Management & Data Systems, Vol. 117, pp. 688-712, May 2017.

[49] P. Y. K. Chau, K.Y. Tam, "Factors Affecting the Adoption of Open Systems: An Exploratory Study," MIS Quarterly, Vol. 21, pp. 1-24, Mar 1997.

[50] A. Pater, A. V. Gils, "Stimulating Ethical Decision-making in a Business Context: Effects of Ethical and Professional Codes," European Management Journal, Vol. 21, pp. 762-772, December 2013.

[51] S. K. Choi, C. H. Yang, J. Kwak, "System Hardening and Security Monitoring for IoT Devices to Mitigate IoT Security Vulnerabilities and Threats," KSII Transactions on Internet and Information Systems, Vol. 12, pp. 906-918, February 2018.

[52] J. Jin, J. Gubbi, S. Marusic, M. Palaniswami, "An information framework for creating a smart city through internet of things," IEEE Internet of Things Journal, Vol. 1, pp. 112-121, April 2014.

[53] Y. H. Wang, C. C. Hsieh, "Explore technology innovation and intelligence for IoT (Internet of Things) based eyewear technology," Technological Forecasting and Social Change, Vol. 127, pp. 281-290, February 2018.

[54] K. Haseeb, K. A. Bakar, A. H. Abdullah, T. Darwish, "Adaptive energy aware cluster-based routing protocol for wireless sensor networks," Wireless Networks, Vol. 23, pp. 1953-1966, April 2016.

[55] T. Yamane, "Statistics: An introductory analysis," New York, 1967.

[56] A. Taweesangrungroj, R. Rattanabanchuen, S. Sinthupinyo, "A Decision-Making Model of Technological-Focused Government Agency Selection of Technological Start-Up Businesses," Trends in Sciences, Vol. 19, pp. 1749-1749, January 2022.

[57] L. Mikhailov, M.G. Singh, "Fuzzy analytic network process and its application to the development of decision support systems," IEEE Transactions on Systems, Man, and Cybernetics, Vol. 33, pp. 33-41, April 2003.

[58] T. Partani, S. V. Marashi, M. H. Alishahi, "Using Fuzzy Analytic Network Process (FANP) in a SWOT Analysis," Global Journal of Computer Science and Technology, Vol. 13, pp. 16-32, June 2013.

[59] N. Nezamoddini, F. Kianfar, F. H. Tash, "Integrated strategic decision making using ANP," IEEE Chinese Control and Decision Conference. May 2011.

[60] L. A. Zadeh, "Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems," Advances in Fuzzy Systems – Applications and Theory, Vol. 6, pp. 394-432, May 1996.

[61] S. Sinrat, W. Atthirawon, "Integrated factor analysis and fuzzy analytic network process (FANP) model for supplier selection based on supply chain risk factors," Research Journal of Business Management, Vol. 9, pp. 106-123, May 2015.

[62] D. Y. Chang, "Applications of the extent analysis method on fuzzy AHP," European Journal of Operational Research, Vol. 95, pp. 649-655, December 1996.

[63] R. Y. Zhong, S. Lan, C. Xu, Q. Dai, "Visualization of RFID-Enabled Shopfloor Logistics Big Data in Cloud Manufacturing," The International Journal of Advanced Manufacturing Technology, Vol. 84, pp. 5-16, August 2016.

[64] T. Qu, Y. Pan, X. Lin, K. Kang, C. Li, M. Thurer, G. Q. Huang, "Internet of Things-based real-time production logistics synchronization mechanism and method toward customer order dynamics," Transactions of the Institute of Measurement and Control, Vol. 39, pp. 429-445, April 2017.

[65] S. Bag, J. H. C. Pretorius, "Relationships between industry 4.0, sustainable manufacturing and circular economy: proposal of a research framework," International Journal of Organizational Analysis, July 2020.

[66] C. L. Hsu, J. C. C. Lin, "An empirical examination of consumer adoption of Internet of Things services: Network externalities and concern for information privacy perspectives," Computers in Human Behavior, Vol 62, pp. 516-527, September 2016.

[67] S. E. Yoo, T. Kim, "Industrial Wireless Sensor Networks: Protocols and Applications,"Sensors, Vol. 20, pp. 1-3, October 2020.

[68] Z. Bi, L. D. Xu, C. Wang, "Internet of things for enterprise systems of modern manufacturing," IEEE Transactions on industrial informatics, Vol. 10, pp. 1537-1546, January 2014.

[69] F. Tao, J. Cheng, Q. Qi, "IIHub: An industrial Internet-of-Things hub toward smart manufacturing based on cyber-physical system," IEEE Transactions on Industrial Informatics, Vol. 14, pp. 2271-2280, May 2017.

[70] M. A. Elkheir, M. Hayajneh, N. A. Ali, "Data management for the internet of things: Design primitives and solution," Sensors, Vol. 13, pp. 15582-15612, July 2013.

[71] W. M. P Kumari, "Artificial intelligence meet Internet of Things," Sensors, Vol. 4, pp. 51-55, 2017.

[72] C. Kahraman, T. Ertay, G. Buyukozkan, "A fuzzy optimization model for QFD planning process using analytic network approach," European Journal of Operational Research, Vol. 171, pp. 390–411, June 2006.

Investigation of Hybrid Feature Selection Techniques for Autism Classification using EEG Signals

S.Thirumal[1]
Research Scholor, Department of CSE
Hindustan Institute of Technology and Science
Chennai, India

J.Thangakumar[2]
Associate Professor, Department of CSE
Hindustan Institute of Technology and Science
Chennai, India

***Abstract*—Autism Spectrum Disorder (ASD), the non-uniform neurodevelopment condition that is characterized by the impairment of behaviour in communication and social interaction with some restricted their repetitive behaviour. Today, to measure the voltage created during brain activity is measured using electroencephalography (EEG). The wavelet transform is used for decomposing the time-frequency of the EEG signal. Feature Selection is the process that significantly reduces feature space dimensionality, while maintaining the right representation of their original data. In this work, metaheuristic algorithm is utilized for feature selection. The proposed feature selection is based on River Formation Dynamics (RFD) and a hybrid Greedy RFD is presented. Support Vector Machine (SVM) can be a concept consisting of a set of methods of supervised learning to analyze pattern recognition that is a successful tool in the analysis of regression and classification. Experimental results show the proposed Greedy RFD feature selection improves the performance of the classifiers and enhance the accuracy of classifying ASD.**

Keywords*—Autism spectrum disorder (ASD); electroencephalo graphy (EEG); feature selection; River Formation Dynamics (RFD); Support Vector Machine (SVM); hybrid greedy RFD

I. Introduction

All the existing medical conditions in human are recognised with the support of a medical professional [1]. But, owing to any variation in their physiological signals, the assessments can result in human error and are not similar while being met by a medical professional. Autism has been identified as a new spectrum disorder that can affect every individual in a different manner in different degrees. This is characterized by means of social deficits, poor communication, and repetitive behavioural patterns. Nevertheless, all problems with children that have a condition of Autism Spectrum Disorder (ASD) can contribute to this. Even though the ministry has been focusing on its early identification resulting in screening and intervention, there are some challenges faced. Different tools are available to diagnose the condition of ASD using this. Human brain have specific paths to send, receive and to interpret all information to bring about a response through its synaptic activation.. These pathways are known as sensory systems [2, 3].

The electroencephalogram (EEG) is a signal acquired form individuals that are scientifically and clinically sound. As a consequence, quantification has a vital role in the study of the human brain [4]. The frequency content and examination of EEG signals, in particular, were identified as a preponderant approach to the knowledge extraction problem. These were used in the research of brain processes related to Motor Imagery (MI). It shows there were tremendous development in the Brain –computer Interface (BCI) technology, which aims to replace compromised human neuromuscular system capability. The majority of BCIs depend on physiologically well-defined EEG properties, such as oscillations in neural networks or their potentials generated by specific stimuli. Various methods that assess the energy of the signal that is dispersed in this frequency, which is t-f or t-s domains, dominated the extraction of EEG features.The Discrete Wavelet Transform (DWT) further decomposes this EEG segment into various sub-bands [5]. Different statistical features and entropy functions were used for the extraction of features from sub-bands.

Multi-variate relations in data can be captured using Machine learning techniques [6] and are very suitable for the detection distribution and subtle differences of certain found in data. Therefore, when compared to the other univariate approaches, the machine learning approach will be able to perform better in terms of classifying the EEG, especially in terms of conditions such as ASD [7, 8]. Therefore, they hold plenty of promise in bringing about knowledge improvement in diagnosing ASD. The Support Vector Machines (SVM) tends to have some merits like higher accuracy and not needing large numbers of training samples that avoid overfitting. Therefore, the SVM has now aroused plenty of concern among the neuroimaging research community [9]. GBM iteratively builds by combining prediction from several weak learners and achieves good performance [10]. In this work, greedy RFD has been investigated to classify the EEG and ASD.

The main contribution of this work is:

The features of EEG and the behavioural data is fused to achieve better accuracy in classifying ASD.

The River Formation Dynamics (RFD) metaherustic is used for selecting features from the wavelet transforms.

The proposed Greedy RFD improves the performance of RFD.

II. Related Work

Sudirman et al [11] aimed to build a new sensory profile using the EEG bio-signal and its potential for distinguishing between various sensory responses. All EEG signals needed here were useful in identifying various emotional states like super learning, light relaxation, and positive thinking. These were inside a frequency range falling between 8 and 12 Hertz. A total of 64 children were part of this research, and from among them, about 34 children were given vestibular sensory, taste, sound, and visual simulations. All raw EEG data had been filtered using an Independent Component Analysis (ICA), and wavelet transforms with the EEGLAB software. To build its sensory profile, standard deviations means and entropy approximation had been extracted from filtered EEG signals.

Herman et al [12] had performed a study which compares spectral signal representation like Power Spectral Density (PSD) techniques, continuous and discrete wavelets, atomic decompositions, and Time-Frequency (t-f) energy distributions. The main emphasis was on the identification of certain differentiated properties in the feature sets that represent EEG trials that are recorded at the time of imagination, that is, for the left-hand or the right-hand movement. Separating features can be quantified in an offline study by making use of the accuracy of classification based on a rate that is obtained using linear as well as non-linear classifiers. There are PSD approaches that demonstrate a consistent level of robustness along with effectiveness that is useful in certain distinctive spectral patterns. They are used to differentiate whether MI induced EEGs are belongs to left and right. The observation has been based on data analysis from a total of eleven subjects in two different sessions. Additionally, the capabilities of generalization of these classifiers found in their intersession performance have been discussed.

Cheong et al [13] studied the Discrete Wavelet Transform (DWT) for extraction of features from their EEG signals that were obtained based on a sensory response from children suffering from autism. For the purpose of this study, the DWT was used in order to decompose the filtered EEG signal into their components with a statistical DWT coefficient feature computed within the time domain. Such features were employed for training Multilayer Perceptron (MLP) based neural network for the classification of signals into three different classes based on the severity of autism (whether mild, moderate, or severe). The results of training in terms of accuracy had achieved about 92.3% with an MSE of 0.0362.

Fan et al [14] had proposed a data-driven method to design the RFD. Speaking technically, the RFD was constructed using threshold responses using receptive fields from many candidates based on their distinctiveness and their correlations. By means of employing two types of such receptive fields (rectangular pooling and Gaussian pooling), two different binary descriptors, the RFDR and RFDG, were chosen. The experiments of image matching experiments on the Patch Dataset and the Oxford Dataset proved that the RFD was able to outperform them, and its work was comparable to the float-valued descriptors at a short time. Experiments on object recognition proved that the RFDR and RFDG were able to show better performance than their competitors.

The human genome can be used to extract vast amounts of data. Autism is a type of neurobehavioral disorder in which a person's capacity to interact and communicate is impaired. It has a solid genetic foundation. There are various gene variants linked to autism, and these changes can disrupt the functioning of the brain that begins before birth. Mutated genes are passed down from their immediate ancestors to their offspring, and this is a risk factor for autism. Reeta et al [15] had proposed another novel approach for ranking such diseased genes that are found in an autistic individual. This system will predict the autistic behaviour of individuals by means of comparing certain similarities among the genes of individuals and the diseased for training that is implemented by making use of the method known as the Naive Bayesian classification. For example, in case the DNA of an individual will be tested, and in case it is found that the DNA consists of certain diseased genes in the training set, autism may be predicted. The approach also makes the process of diagnosis of the condition simpler and earlier as well.

In Vaishali and Sasikala [16], the diagnosis dataset of the ASD will have 21 features that are obtained from the repository of the UCI machine learning, and this has been experimented with by using a swarm intelligence-based binary firefly feature selection wrapper by using the same, there was an identification made that about 10 features among a total of 21 ASD datasets were enough to distinguish between the patients that had ASD or did not. The obtained results, along with the proposed approach, were able to justify the new hypothesis by means of producing a certain amount of average accuracy within the range of about 92.12%-97.95% along with optimum feature subsets that were equal to the accuracy of the ASD.

In Alzubi et al [17], a hybrid mechanism to select feature which is more accurate was proposed to detect informative SNPs to be chosen for an optimal SNP subset. This method was based on the fusion of that of a filter and also a wrapper method along with the Conditional Mutual Information Maximization (CMIM) method and Support Vector Machine Recursive Feature Elimination (SVM-RFE). The proposed method’s performance had been evaluated based on three different state-of-the-art methods, which are the Minimum Redundancy Maximum Relevance (mRMR), ReliefF, and the CMIM. It also used four classifiers which were the Naive Bayes (NB), SVM, Linear Discriminant Analysis (LDA), and the k Nearest Neighbors (kNN) on an ASD-SNP dataset that was obtained from the Gene Expression Omnibus and National Center for Biotechnology Information genomics data repository. The results of the experiment had demonstrated that the efficiency of this approach of feature selection outperformed other methods and achieved a classification accuracy of 89%.

III. Methodology

Feature selection refers to a technique that was employed for pre-processing data, and this is preferable at the time of performing machine learning. Selection by mean is to choose attributes and variables within a dataset that is fit into a particular model and is tested for performance. The section further details the extraction of features by making use of the Wavelet Transform, Support Vector Machine, RFD, Greedy RFD, Naïve Bayesian Classification, and K- Nearest Neighbours.

A. Datasets

The techniques are evaluated using autism dataset obtained from King Abdulaziz University (KAU) Brain Computer Interface (BCI) Group. The EEG data is recorded using all the electrodes with 16 channels. This dataset had been filtered using a band-pass filter along with pass band frequency (0.1–60Hz) and a notch filter that had stop band frequency (60Hz) and at 256Hz frequency sampling is digitized. The recording time of EEG is varied from 12 to 40 minutes among autistic subjects up to 173 minutes, and from 5 to 27 minutes up to 148 minutes for normal subjects.

The behavioural dataset related to autism screening of toddlers containing features that is used for further analysis to determine autistic traits. Ten behavioural features is recorded with other individuals characteristics.

B. Feature Extraction using Wavelet Transform (WT)

The These wavelets were in use in the recent decade for different tasks of image processing. For the purpose of image compression, fractals compression, fractals, resolution enhancement, denoising, and image enhancement frequency domain and time analysis is used. The basic idea behind all of this is the analysis of the signal in accordance with the scale. The main advantages of such Wavelet transforms are that compared to the Fourier transforms that represent functions; there are some discontinuities with sharp peaks that help in the accurate deconstruction or reconstruction of non-periodic, non-stationary finite signals. The images based on wavelet, enhancement, de-noising, and so on had better performance owing to the properties of sparsity or multi-resolution structure. Wavelet transform also has the trait of multi-resolution analysis aside from the ability to express a local feature of this signal in the domain of time and frequency. Therefore, it is found to be fit to detect the flash state or the irregularity of the signal and in setting out the composition [18]. The wavelet ψ having a compact support and vanishing moment n is as in (1):

$$\int_{-\infty}^{+\infty} t^k \psi(t)dt = 0, \text{ for } 0 \leq k \leq n \quad (1)$$

This has another function θ that has fast decay and is shown as (2).

$$\psi(t) = (-1)^n \frac{d^n\theta(t)}{dt^n} \quad (2)$$

Once this is done, the wavelet transform for signal f is given as in (3):

$$Wf(u,s) = s^n \frac{d^n}{du^n}(f * \overline{\theta}_s)(u) \quad (3)$$

Wherein, $\overline{\theta}_s(t) = s^{-1/2}\theta(-t/s)$, are the time and space coordinate and s the scale. The wavelet transforms Wf (u, s) is the nth order derivative of f that has $\overline{\theta}_s$ on a domain proportional to s.

The most important frequencies of EEG exist between 0.1 to 30 Hz. The standard EEG clinical bands are the delta (0.1 to 3.5 Hz), theta (4 to 7.5 Hz), alpha (8 to 13 Hz), and beta (14 to 30 Hz) bands. A sample EEG image and corresponding alpha, beta, delta and theta waveforms are shown in Fig. 1. 30Hz waves will termed as gamma waves.

In this work, the number of decomposition levels is taken as 5. Thus, the EEG signal is decomposed into D1-D5 details. As a result output coefficients of mean, standard deviation, variance, skewness and kurtosis feature vectors are used for classifying the signal.

C. River Formation Dynamics (RFD) Feature Selection

The algorithm of River Formation Dynamics was earlier used in solving problems that are NP-hard to find paths within a graph [19]. Furthermore, the algorithm is also capable of optimizing the distance of the path and also considers certain other dependencies like restrictions to acceleration, velocity, pathfinding tasks, and so on, which are considered to be NP-hard problems [20]. The RFD algorithm can be depicted as given below. The actual amount of soil that is assigned to each of the nodes will drop as they keep moving and erode the paths or depositing the carried sediment (thereby increasing node altitude). The descending slope makes more dependence on the probability of deciding the next node, and which can be relative to the actual the node's altitude dissimilarity, the altitude of its adjacent node, and the drop position. The environment created in the early stages will be flat. This means that throughout the process, the nodes will have the same height, except in the case of a target node with a height of zero. There is a blotch placed at the starting node to navigate the entire site, which determines the optimal path. At every step there is another set of droplets are traversed in space sequentially and erosion is marked at the nodes traversed.

The nodes represent features, and the routes between them reflect the decision of the next feature when utilising RFD to optimise feature selection. The drop travelling through with the least number of nodes visited that meets the stopping requirement is used to find the best feature subset.

The first step will be the initialization of the nodes of the algorithm and define the set that is formed using the cell decomposition of the site. Every node here will possess all information, and in case it consists of an obstacle aside from additional data, the determination of the time that is required for traveling across the entire distance to its goal is considered. When there is a drop initialization there may be a suitable number of drops that are kept to the first node. After this, the algorithm will be executed until such time the last condition is met.

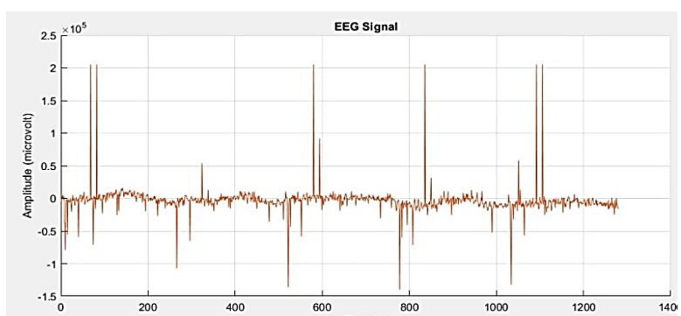


Fig. 1. Sample EEG Images.

The statements says that the drops were in the same path, and in addition to that, for reducing the time taken for computation, a upper limit on the actual number of iterations was introduced along with the condition that verifies whether the last n loops have improved the solution or not. Path analysis will involve identifying the right solution. For the purpose of discovering the best drop that can conduct any additional erosion, it has to be performed on travelled paths by means of reducing the altitude of the nodes. The final step will be to add some sediment to the nodes in order to overcome any circumstances in which all the altitudes are near 0, thus making the gradients can be ignored, which may destroy the paths formed. This will diminish slowly with each loop in the algorithm.

D. Proposed Greedy River Formation Dynamics Feature Selection

The Greedy algorithm is the one that can determine this problem by means of making a choice that appears to be the best at that moment. There are several problems of optimization that may be solved by means of using this algorithm, and some may not have an efficient solution. However, the Greedy algorithm can implement such a solution which is nearly optimal. The Greedy algorithm also reflects a problem-solving heuristic that makes a locally optimal choice in every stage, hoping to be able to find a global optimum. For most of these issues, the greedy strategy may not be able to produce an optimal solution, but it can bring about a locally optimal one that can approximate another globally optimal solution within a reasonable time frame.

```
    Initialize Nodes(), Drops(), Greedy (D, n)
    Solution <- 0
    while (not endConditionMet())
    For i<-1 to n do {
        S<-Select (D)
        If (Feasible (solution, s)) then
                Solution <-Union (Solution, s)}
        Return solution
    moveDrops()
    analyzePaths()
    erodePaths()
    depositSediments()
end while
```

E. K- Nearest Neighbors (KNN)

Computing the closest distance between the neighbours is depicted as the K value, and for making use of this algorithm, a few other elements in the initial subject set, which is the K number (of nearest neighbours), is observed, and the K parameter and distance were considered.

These training tuples have been described in the form of n attributes. In this, every tuple will indicate the point within the n-dimensional space. So, all training tuples will be stored within an n-dimensional pattern space. For a certain unknown tuple, the k-NN classifier will lookout for a pattern space in the k training tuples that are the closest to the unknown one. The k training tuples represent the k-nearest neighbours of this unknown tuple. "Closeness" has been defined as the distance metric like the Euclidean distance. Euclidean distance between that of two points or tuples will be$X = (x_1, x_2, \ldots., x_n)$ and $Y = (y_1, y_2, \ldots., y_n)$ is (4):

$$d(X,Y) = \sqrt{\sum_{i=1}^{n}(X_i - Y_i)^2} \tag{4}$$

The step by step process of K-NN algorithm is given below:

- Computing the distance between that of the new sample and all earlier samples that have been grouped in the form of clusters;
- Sorting the distance in increasing order to choose k samples having smaller distance values;
- Applying a voting principle and adding a new sample to the largest cluster of the k samples.

F. Naive Bayesian Classification

Naive Bayes' is very popular method in the categorization of the texts and in identifying documents of a certain type as to whether they are legitimate or spam. The Naive Bayes' Classifier can be very scalable and needs linear parameters in the variables (or predictors) for a learning problem. The probability models with strong assumptions of independence of Naïve Bayes classification using diseased gene classification and is a conditional model.

$$P(C_k|x) = P(C_k) \,.\, P(x|C_k)\, P(x) \tag{5}$$

If there is a genetic instance that has to be classified and is represented by a vector $x = (x_1, x_2 \ldots\ldots x_n)$ will represent n genes with assigned probabilities. $P(C_k, x_1 \ldots\ldots x_n)$ It is for every possible k outcome or class Ck.

G. Support Vector Machine (SVM)

The SVM refers to a new form of supervised learning algorithm for rightly categorizing a target result that uses independent variables which is present inside the dataset. Another new SVM can be the maximum margin classifier, and this will further maximize any separation between the n classes of data. The SVMs are useful, especially when there is a boundary between the groups that are non-linear and owing to this feature, they are normally used for problems of classification where there is a distinction made between groups that are non-linear. The SVM algorithms are used for classifying individuals based on diagnosis, neuroimaging, genes, standardized assessments, and some more measurements [21]. The SVM classification algorithm makes use of training instances to predict other new ones using two different class label −1, 1. As per Fig. 1, a hyperplane is $w^T x + b = 0$, wherein, $w \in R^n$ is orthogonal to a hyperplane and $b \in R^n$is constant. With training data D, equation (6) is derived:

$$D = \{(\vec{x}_i, \vec{y}_i) | \vec{x}_i \in R^m, \vec{y}_i \in \{-1, +1\}\}_{i=1}^{n} \tag{6}$$

In which x_i is the m-dimensional real vector, y_i the input vector class, x_i either -1 or $+1$. The SVM looks for a hyperplane to maximize the margin between two sample classes in D.

$$y_i(\vec{w}^T\vec{x}+b) \geq 1 \tag{7}$$

The aim of the SVM is to increase the distance between two different hyperplanes. One will compute the distance between both hyperplanes $\frac{1}{||\vec{w}||}$. The SVM training in a non-separable case will be solved with a problem of quadratic optimization as in Equation (8):

$$min\,i\,mize\text{: P}(\vec{w},\text{b},\xi) = \frac{1}{2}||\vec{w}|| + C\sum_{i=1}^{n}\xi_i$$

$$subject\ to\text{: y}(\vec{w}.\phi(\vec{x})\text{+b}) \geq 1\text{-}\xi_i\ ,\ \xi_i \geq 0 \tag{8}$$

H. Gradient Boosting Machine (GBM)

An ensemble technique, boosting will help to reduce the bias which is dependent and also generalized error in an ensemble. Another technique of boosting which will repeatedly combine 30 base (weak) learners that have low variance and high bias like the stumps in the decision tree. These base learners are combined to ensure ensemble bias, and this will reduce the variance remaining the same, thus reducing its net ensemble error. For every boosting step or iteration, the GBM will construct yet another new base learner to a negative gradient of the loss function with the observed data in order to ensure the focus of the new base learner is on the model and its weakness. This means a functional approximation for the model has been made by bringing about a consecutive improvement with the negative direction to the loss function [22].

Normally, the GBM algorithm will have better results in the case, for every iterative step, there can be an added decision tree and its contribution, which is reduced by using a parameter for shrinkage α which is known as the rate of learning. The main idea behind this method of shrinkage in the GBM will be that it has more steps, and these small ones will result in better accuracy compared to a less number of larger steps. This parameter of learning α will fall between 0 and 1, and the smaller the value, the more its accuracy.

For each iterative step, as opposed to making use of a complete training dataset, a randomly chosen (that does not have a replacement) subsample that will fit a decision tree will be used. If there are many observations a default fraction of this data will ½=0.5. This means about half i.e.,50%) of the dataset can be used. Additionally, algorithms computation cost will be reduced by subsampling to the means of a factor that is equal to the subsampling factor [23] and improves the accuracy of Gradient boost machine model. The algorithm used for gradient boosting is as given below:

- Initialize the predictions with one simple decision tree.
- Calculate the residual – and this will be its (actual-prediction) value.
- Build a shallow decision tree to predict the residual based on independent values.
- Update its original prediction with another one multiplied by its rate of learning.
- Iterate the step 2 to 4 for certain number of times which should be equal to the tree count.

I. Proposed Greedy RFD GBM

The proposed Greedy RDF is used to optimize the GBM's hyperparameters. The GBM model hyperparameter of number of trees, tree depth, learning rate, Minimum number of observations in terminal nodes is optimized using the Greedy RFD optimization algorithm. In the proposed Greedy RFD GBM, the initial solutions are created randamly. The range specified for number of trees is 1000 to 5000, tree depth 1 to 10, learning rate 0.05–0.3, Minimum number of observations in terminal nodes 5-15. The Root Mean Square Error (RMSE) is used as the objective function. On iterations of the Greedy RBF, the optimal set of GBM hyperparameters to classify ASD is obtained.

IV. Result and Discussion

The techniques were evaluated for two scenarios, using features from EEG data only and using features from both EEG and behavioural data. In the latter, the features of EEG and the behavioural data is fused. In this section, the RFD feature methods with classifiers such as KNN, NB, SVM and GBM are evaluated. Section 4.1 presents the results for EEG data without feature fusion and section 4.2 the results for feature fusion.

A. Without Feature Fusion

Tables I to IV and Fig. 2 to 5 shows accuracy of the classification, precision, recall and F measure for both normal and ASD features.

TABLE I. Classification Accuracy for Greedy RFD-Greedy RFD GBM without Feature Fusion

Techniques Used	Classification Accuracy
RFD-KNN	92.31
RFD-NB	92.69
RFD-SVM	93.27
RFD-GBM	95.19
Greedy-RFD-GBM	95.58
Greedy RFD-Greedy RFD GBM	96.35

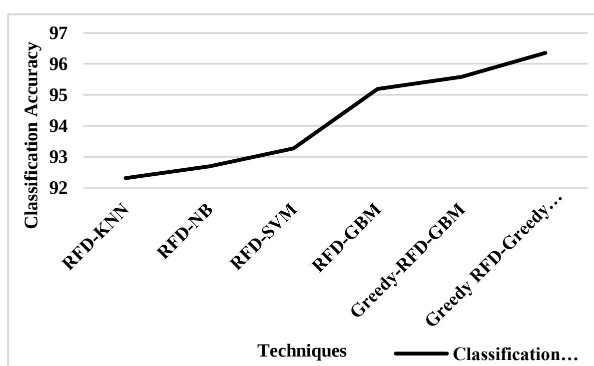


Fig. 2. Classification Accuracy for Greedy RFD-Greedy RFD GBM without Feature Fusion.

From Fig. 2, it can be observed that the Greedy RFD-Greedy RFD GBM has higher classification accuracy by 4.3%, by 3.9%, by 3.25%, by 1.21% and by 0.8% for RFD-KNN, RFD-NB, RFD-SVM, RFD- GBM and Greedy-RFD-GBM, respectively. The selection of optimal hyperparameters of GBM shows improved performance.

TABLE II. RECALL FOR GREEDY RFD-GREEDY RFD GBM WITHOUT FEATURE FUSION

Techniques Used	Recall for normal	Recall for ASD
RFD-KNN	0.9214	0.9237
RFD-NB	0.9286	0.9263
RFD-SVM	0.9357	0.9316
RFD-GBM	0.95	0.9526
Greedy-RFD-GBM	0.95	0.9579
Greedy RFD-Greedy RFD GBM	0.9571	0.9658

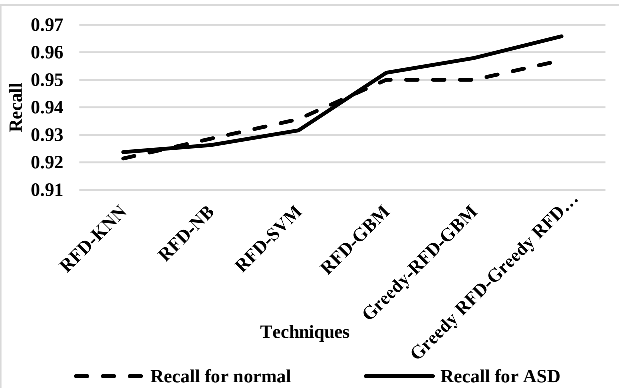


Fig. 3. Recall for Greedy RFD-Greedy RFD GBM without Feature Fusion.

From Fig. 3, it can be observed that the Greedy RFD-Greedy RFD GBM has higher recall by 3.8%, by 3.02%, by 2.26%, by 0.74% and by 0.74% for RFD-KNN, RFD-NB, RFD-SVM, RFD- GBM and Greedy-RFD-GBM, respectively, for normal. Similarly, the Greedy RFD-Greedy RFD GBM has higher recall by 4.46%, by 4.18%, by 3.6%, by 1.38% and by 0.82% for RFD-KNN RFD-NB, RFD-SVM, RFD- GBM and Greedy-RFD-GBM, respectively for ASD. It is observed that the feature selection significantly improves the classification of the ASD. The proposed Greedy RFD feature selections achieves the best performance.

TABLE III. PRECISION FOR GREEDY RFD-GREEDY RFD GBM WITHOUT FEATURE FUSION

Techniques Used	Precision for normal	Precision for ASD
RFD-KNN	0.8165	0.9696
RFD-NB	0.8228	0.9724
RFD-SVM	0.8344	0.9752
RFD-GBM	0.8808	0.981
Greedy-RFD-GBM	0.8926	0.9811
Greedy RFD-Greedy RFD GBM	0.9116	0.9839

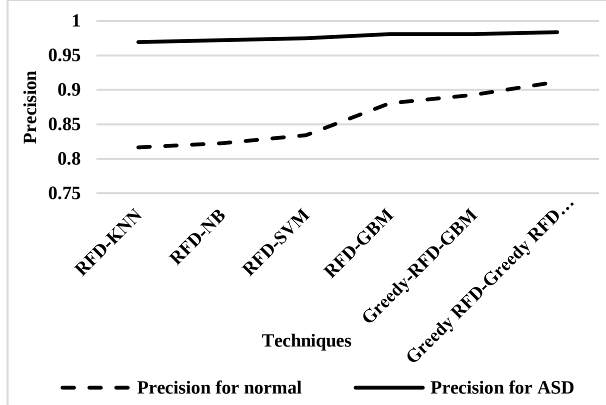


Fig. 4. Precision for Greedy RFD-Greedy RFD GBM without Feature Fusion.

From Fig. 4, it can be observed that the Greedy RFD-Greedy RFD GBM has higher Precision by 11%, by 10.24%, by 8.84%, by 3.44% and by 2.1% for RFD-KNN, RFD-NB, RFD-SVM, RFD- GBM and Greedy-RFD-GBM, respectively for normal. Similarly, the Greedy RFD-Greedy RFD GBM has higher Precision by 1.46%, by 1.2%, by 0.89%, by 0.3% and by 0.29% for RFD-KNN RFD-NB, RFD-SVM, RFD- GBM and Greedy-RFD-GBM, respectively for ASD. The optimization of the feature selection and classifier hyperparameters has significant improvement in the performance. The proposed Greedy RFD is effective when compared to RFD.

TABLE IV. F MEASURE FOR GREEDY RFD-GREEDY RFD GBM WITHOUT FEATURE FUSION

Techniques Used	F measure Normal	F Measure ASD
RFD-KNN	0.8658	0.9461
RFD-NB	0.8725	0.9488
RFD-SVM	0.8822	0.9529
RFD-GBM	0.9141	0.9666
Greedy-RFD-GBM	0.9204	0.9694
Greedy RFD-Greedy RFD GBM	0.9338	0.9748

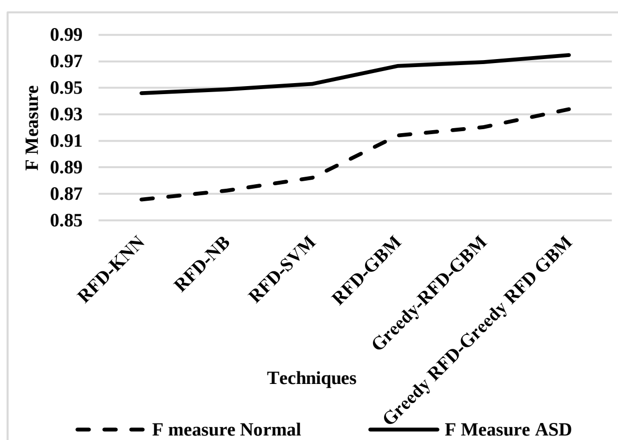


Fig. 5. F measure for Greedy RFD-Greedy RFD GBM without Feature Fusion.

From Fig. 5, it can be observed that the Greedy RFD-Greedy RFD GBM has a higher F measure by 7.6%, by 6.79%, by 5.68%, by 2.13% and by 1.45% for RFD-KNN, RFD-NB, RFD-SVM, RFD- GBM and Greedy-RFD-GBM respectively for normal. Similarly, the Greedy RFD-Greedy RFD GBM has higher F measure by 2.99%, by 2.7%, by 2.3%, by 0.84% and by 0.56% for RFD-KNN, RFD-NB, RFD-SVM, RFD- GBM and Greedy-RFD-GBM, respectively for ASD.

B. After Feature Fusion

Headings, The classification accuracy, recall, precision and F Measure for both normal and ASD features as shown in Table V to VIII and Fig. 6 to 9.

TABLE V. CLASSIFICATION ACCURACY FOR GREEDY RFD FEATURE SELECTION-GREEDY RFD GBM

Techniques	Classification accuracy
RFD-Feature Selection - KNN	94.08
RFD-Feature Selection -Naïve Bayes	94.64
RFD-Feature Selection - SVM	95.21
RFD Feature Selection - GBM	95.78
Greedy RFD feature selection GBM	96.56
Greedy RFD feature selection - Greedy RFD GBM	97.15

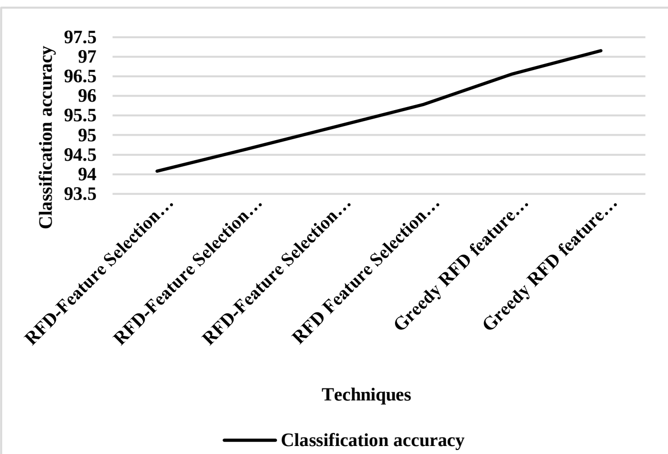


Fig. 6. Classification Accuracy for Greedy RFD Feature Selection-Greedy RFD GBM.

From Fig. 6, it can be observed that the Greedy RFD feature selection-Greedy RFD GBM has higher classification accuracy by 3.21% for RFD-feature selection-KNN, by 2.62% for RFD-feature selection-Naive Bayes, by 2.01% for RFD-feature selection-SVM, by 1.42% for RFD-feature selection-GBM and by 0.61% for Greedy RFD feature selection GBM, respectively.

From Fig. 7, it can be observed that the Greedy RFD feature selection-Greedy RFD GBM has higher recall for normal by 2.94% for RFD-feature selection-KNN, by 2.22% for RFD-feature selection-Naive Bayes, by 1.48% for RFD-feature selection-SVM, by 1.48% for RFD-feature selection-GBM and by 0.74% for Greedy RFD feature selection GBM respectively. The Greedy RFD feature selection-Greedy RFD GBM has higher recall for ASD by 3.3% for RFD-feature selection-KNN, by 2.77% for RFD-feature selection-Naive Bayes, by 2.22% for RFD-feature selection-SVM, by 1.39% for RFD-feature selection-GBM and by 0.55% for Greedy RFD feature selection GBM, respectively.

From Fig. 8, it can be observed that the Greedy RFD feature selection-Greedy RFD GBM has higher precision for normal by 8.33% for RFD-feature selection-KNN, by 6.99% for RFD-feature selection-Naive Bayes, by 5.63% for RFD-feature selection-SVM, by 3.56% for RFD-feature selection-GBM and by 1.44% for Greedy RFD feature selection GBM, respectively. The Greedy RFD feature selection-Greedy RFD GBM has higher precision for ASD by 1.12% for RFD-feature selection-KNN, by 0.84% for RFD-feature selection-Naive Bayes, by 0.56% for RFD-feature selection-SVM, by 0.56% for RFD-feature selection-GBM and by 0.27% for Greedy RFD feature selection GBM, respectively.

TABLE VI. RECALL FOR GREEDY RFD FEATURE SELECTION-GREEDY RFD GBM

Techniques	Recall for Normal	Recall for ASD
RFD-Feature Selection – KNN	0.9348	0.9431
RFD-Feature Selection -Naïve Bayes	0.9416	0.9482
RFD-Feature Selection - SVM	0.9485	0.9534
RFD Feature Selection - GBM	0.9485	0.9613
Greedy RFD feature selection GBM	0.9556	0.9694
Greedy RFD feature selection - Greedy RFD GBM	0.9627	0.9748

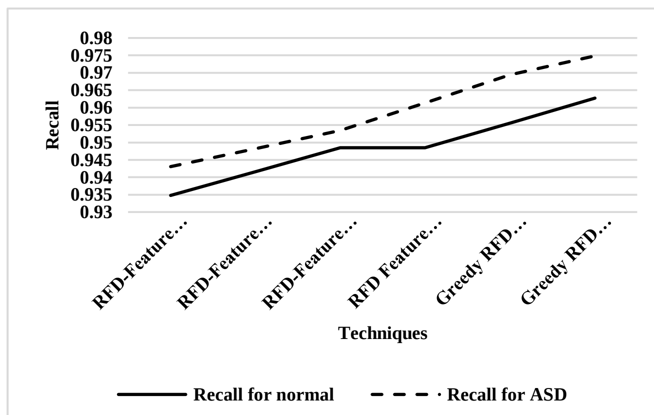


Fig. 7. Recall for Greedy RFD Feature Selection-Greedy RFD GBM.

TABLE VII. PRECISION FOR GREEDY RFD FEATURE SELECTION-GREEDY RFD GBM

Techniques	Precision for Normal	Precision for ASD
RFD-Feature Selection - KNN	0.86	0.9748
RFD-Feature Selection -Naïve Bayes	0.8716	0.9775
RFD-Feature Selection - SVM	0.8836	0.9803
RFD Feature Selection - GBM	0.9021	0.9803
Greedy RFD feature selection GBM	0.9214	0.9831
Greedy RFD feature selection - Greedy RFD GBM	0.9348	0.9858

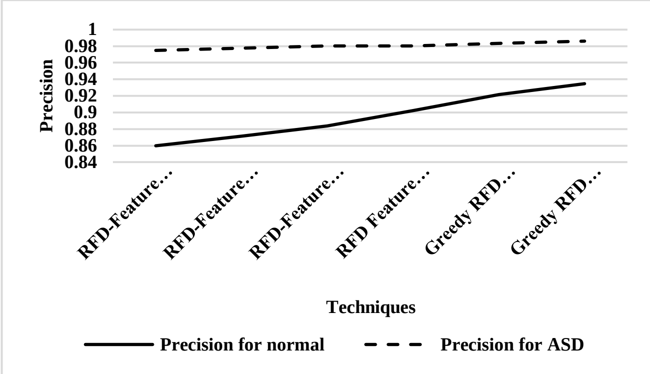


Fig. 8. Precision for Greedy RFD Feature Selection-Greedy RFD GBM.

TABLE VIII. F MEASURE FOR GREEDY RFD FEATURE SELECTION-GREEDY RFD GBM

Techniques	F Measure for Normal	F Measure for ASD
RFD-Feature Selection - KNN	0.8958	0.9587
RFD-Feature Selection -Naïve Bayes	0.9052	0.9626
RFD-Feature Selection - SVM	0.9149	0.9666
RFD Feature Selection - GBM	0.9247	0.9707
Greedy RFD feature selection GBM	0.9381	0.9762
Greedy RFD feature selection - Greedy RFD GBM	0.9485	0.9802

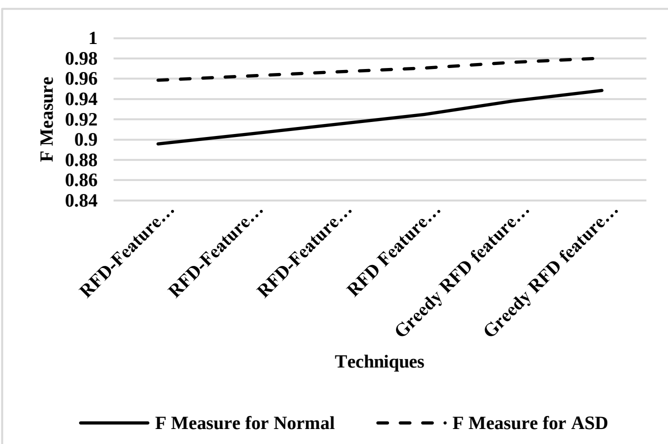


Fig. 9. F Measure for Greedy RFD Feature Selection-Greedy RFD GBM.

From Fig. 9, it can be observed that the Greedy RFD feature selection-Greedy RFD GBM has higher f measure for normal by 5.71% for RFD-feature selection-KNN, by 4.67% for RFD-feature selection-Naive Bayes, by 3.61% for RFD-feature selection-SVM, by 2.54% for RFD-feature selection-GBM and by 1.09% for Greedy RFD feature selection GBM respectively. The Greedy RFD feature selection-Greedy RFD GBM has higher f measure for ASD by 2.22% for RFD-feature selection-KNN, by 1.82% for RFD-feature selection-Naive Bayes, by 1.39% for RFD-feature selection-SVM, by 0.98% for RFD-feature selection-GBM and by 0.41% for Greedy RFD feature selection GBM, respectively.

It is observed that the feature fusion using EEG and behavioural data has higher classification of ASD of 97.15% when compared to 96.35% accuracy with only EEG data. Similarly, precision, recall and f measure achieved is better when feature fusion is used. Though the classification is enhanced, the sampling used is small. The proposed methods needs to be evaluated with larger dataset.

V. CONCLUSION

Autism Spectrum Disorder (ASD) can be a lifelong condition that is quite serious and characterized by restricted but repetitive behaviour along with deficits in terms of communication or its reciprocal social interaction. The EEG and behavioural data is used for identifying the ASD. In this work, the wavelet transform is used to extract features from EEG. Features of the EEG and behavioural data is fused. Feature selection is achieved using proposed RFD and Greedy RFD algorithms. SVM can take a collection of input data to make a prediction where the binary class input will make the SVM a two-class linear classifier that is non-probabilistic. GBM iteratively builds by combining prediction from several weak learners. The hyperparameters of the GBM is optimized using Greedy RFD to improve the classification of the ASD. The results have proved that Greedy RFD-Greedy RFD GBM with fused features from both EEG and behavioural data had a better accuracy of classification than the RFD-KNN, RFD-NB, RFD-SVM, RFD-GBM, and the Greedy-RFD-GBM. The findings show that fusing of EEG and behavioural data does improve the classification of ASD. Future research can focus on evaluating the proposed methods using larger dataset.

REFERENCES

[1] Daniel P. Howsmon , Uwe Kruger, Stepan Melnyk, S. Jill James, Juergen Hahn , “Classification and adaptive behavior prediction of children with autism spectrum disorder based upon multivariate data analysis of markers of oxidative stress and DNA methylation”, PLOS Computational Biology, March 2017.

[2] Bone, D., Goodwin, M. S., Black, M. P., Lee, C. C., Audhkhasi, K., & Narayanan, S. (2015). Applying machine learning to facilitate autism diagnostics: pitfalls and promises. *Journal of autism and developmental disorders*, *45*(5), 1121-1136.

[3] Nogay, H. S., & Adeli, H. (2020). Machine learning (ML) for the diagnosis of autism spectrum disorder (ASD) using brain imaging. *Reviews in the Neurosciences*, *31*(8), 825-841.

[4] Brodu, N., Lotte, F., & Lécuyer, A. (2011, April). Comparative study of band-power extraction techniques for motor imagery classification. In *2011 IEEE symposium on computational intelligence, cognitive algorithms, mind, and brain (CCMB)* (pp. 1-6). IEEE.

[5] Ibrahim, S., Djemal, R., & Alsuwailem, A. (2018). Electroencephalography (EEG) signal processing for epilepsy and autism spectrum disorder diagnosis. *Biocybernetics and Biomedical Engineering*, *38*(1), 16-26.

[6] Raj, S., & Masood, S. (2020). Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques. *Procedia Computer Science*, *167*, 994-1004.

[7] Thabtah, F., & Peebles, D. (2020). A new machine learning model based on induction of rules for autism detection. *Health informatics journal*, *26*(1), 264-286.

[8] Hyde, K. K., Novack, M. N., LaHaye, N., Parlett-Pelleriti, C., Anden, R., Dixon, D. R., & Linstead, E. (2019). Applications of supervised machine learning in autism spectrum disorder research: a review. *Review Journal of Autism and Developmental Disorders*, *6*(2), 128-146.

[9] Kang, J., Han, X., Song, J., Niu, Z., & Li, X. (2020). The identification of children with autism spectrum disorder by SVM approach on EEG and eye-tracking data. *Computers in biology and medicine*, *120*, 103722.

[10] Katuwal, G. J., Cahill, N. D., Baum, S. A., & Michael, A. M. (2015, August). The predictive power of structural MRI in Autism diagnosis. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 4270-4273). IEEE.

[11] Sudirman, R., Hussin, S. S., Airij, A. G., & Hai, C. Z. (2018, July). Profile Indicator for Autistic Children using EEG Biosignal Potential of Sensory Tasks. In *2018 2nd International Conference on BioSignal Analysis, Processing and Systems (ICBAPS)* (pp. 136-141). IEEE.

[12] Herman, P., Prasad, G., McGinnity, T. M., & Coyle, D. (2008). Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *16*(4), 317-326.

[13] Cheong, L. C., Sudirman, R., & Hussin, S. S. (2015). Feature extraction of EEG signal using wavelet transform for autism classification. *ARPN Journal of Engineering and Applied Sciences*, *10*(19), 19.

[14] Fan, B., Kong, Q., Trzcinski, T., Wang, Z., Pan, C., & Fua, P. (2014). Receptive fields selection for binary feature description. *IEEE Transactions on Image Processing*, *23*(6), 2583-2595.

[15] Reeta, R., Pavithra, G., Priyanka, V., & Raghul, J. S. (2018, April). Predicting Autism Using Naive Bayesian Classification Approach. In *2018 International Conference on Communication and Signal Processing (ICCSP)* (pp. 0109-0113). IEEE.

[16] Vaishali, R., & Sasikala, R. (2018). A machine learning based approach to classify Autism with optimum behaviour sets. *International Journal of Engineering & Technology*, *7*(4), 18.

[17] Alzubi, R., Ramzan, N., & Alzoubi, H. (2017, August). Hybrid feature selection method for autism spectrum disorder SNPs. In *2017 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)* (pp. 1-7). IEEE.

[18] Chaudhari, Y. P., & Mahajan, P. M. (2017). Image denoising of various images using wavelet transform and thresholding techniques. *Int. Res. J. Eng. Technol.(IRJET)*, *4*(2).

[19] Redlarski, G., Pałkowski, A., & Dąbkowski, M. (2013). Using river formation dynamics algorithm in mobile robot navigation. In *Solid State Phenomena* (Vol. 198, pp. 138-143). Trans Tech Publications Ltd.

[20] Pałkowski A., Use of swarm intelligence methods in mobile robot navigation [in Polish: Wykorzystanie metod inteligencji rojowej w nawigacji robotów mobilnych], MSc thesis, Gdansk University of Technology, Gdansk, 2012.

[21] Sharma., A, Nagpal., D & Kaur., S, "Comparative Study of Machine Learning Techniques Used in Autism Spectrum Disorder", International Journal for Research in Engineering Application & Management (IJREAM) ISSN : 2454-9150 Vol-05, Issue-10, Jan 2020.

[22] Das, K., & Behera, R. N. (2017). A survey on machine learning: concept, algorithms and applications. *International Journal of Innovative Research in Computer and Communication Engineering*, *5*(2), 1301-1309.

[23] Touzani, S., Granderson, J., & Fernandes, S. (2018). Gradient boosting machine for modeling the energy consumption of commercial buildings. *Energy and Buildings*, *158*, 1533-1543.

An Optimized Hybrid Fuzzy Weighted k-Nearest Neighbor with the Presence of Data Imbalance

Soha A. Bahanshal[1], Rebhi S. Baraka[2], Bayong Kim[3], Vaibhav Verdhan[4]
Department of Computer Science, University of Massachusetts Lowell, Lowell, USA[1, 3]
Department of Computer Science, Islamic University of Gaza, P. O. Box 108, Gaza, Palestine[2]
Analytics Leader, AstraZeneca, London, UK[4]

***Abstract*—We present an optimized hybrid fuzzy Weighted k-Nearest Neighbor classification model in the presence of imbalanced data. More attention is placed on data points in the boundary area between two classes. Finding greater results in the general classification of imbalanced data for both the minority and the majority classes. The fuzzy weighted approach assigns large weights to small classes and small weights to large classes. It improves the classification performance for the minority class. Experimental results show a higher average performance than other relevant algorithms, e.g., the variants of kNN with SMOTE such as Weighted kNN alone and Fuzzy kNN alone. The results also signify that the proposed approach makes the overall solution more robust. At the same time, the overall classification performance on the complete dataset is also increased, thereby improving the overall solution.**

Keywords*—Imbalanced data; fuzzy weighted kNN; SMOTE; classification model; optimized hybrid kNN

I. Introduction

In supervised learning, labeled training data is used to prepare certain classifiers and find the class name of the test data using that classifier [1]. The performance of such classifiers on balanced datasets is generally better than on imbalanced datasets. Hence, there is an increasing need to tackle the issue of class imbalance [2, 3]. The problem of class imbalance states that the number of instances in one class is slightly lower in these datasets than in the other classes [4]. On imbalanced datasets of the binary class, only one positive and one negative class is present. The positive and negative classes are the minor and the major classes, respectively.

In many classification problems, however, the more useful are the instances of the minor with lower instances [5]. Therefore, imbalance occurs whenever the class of interest is relatively rare and has a small number of instances compared to the majority class. In addition, relative to the cost of misclassifying the majority class, for example, the cost of misclassifying the minority class is very high; consider cancer versus non-cancer or fraud versus un-fraud [6]. Since the majority class is over-represented, it impacts the training of the classifier and hence the majority class has better accuracy than the minority class(es).

Although a variety of solutions to data imbalance have been developed, in some ways they have shortcomings. Some solutions consider adding, deleting or weighting the data in order to closely balance the data. Other solutions attempt to find some good pre-processing measures for solving such particular problems in the training dataset that may restore balance between the majority and minority classes before performing classification [7, 8].

In general, existing class-imbalanced classification methods are divided into four categories: either manipulating or modifying the distribution of data by under- or over-sampling (data sampling), modifying in traditional classification existing algorithms to suit class imbalance (algorithmic modification), ensemble approaches [9], or cost-sensitive learning [10]. Therefore, data imbalance can be resolved by over-sampling the under-represented class or under-sampling the over-represented class. But both of these methods are not much scientific and suffer from various drawbacks. Therefore, we adopt the first and the second categories of sampling a dataset in the pre-processing phase and the modification of traditional classification algorithm. The data sampling methods focus on balancing the data, and the common strategies are to reduce the majority class examples (under-sampling) or to add new minority class examples to the data (oversampling) [8, 11].

Synthetic Minority Oversampling Technique (SMOTE) is one of the techniques used to balance imbalanced datasets. Researchers have widely adopted SMOTE due to its versatility and added value with respect to random over-sampling [8]. It reduces the possibilities of over-fitting by randomly resampling the data and generating new samples of the minority class by interpolating multiple samples of the minority class that lie together. Nearest neighbour rule can be used over here. The over-fitting dilemma is thereby eliminated and the decision space is more widespread for the minority class; meanwhile the decision space for the majority class is reduced. By operating in feature space, synthetic instances are created. Some drawbacks of SMOTE, however, are unavoidable because it synthesizes new instances without taking the majority class into account [8], which could lead to fuzzy boundaries between the positive and negative classes. Therefore, SMOTE technique has been proposed with various improvements and extensions that aim to eliminate its drawbacks.

In this paper, we combine a method of class weighting with SMOTE's over-sampling of the minority class to improve the classification accuracy of the minority class without sacrificing the accuracy of the majority class. The combination is performed in a simple classification model based on kNN algorithm [12, 13] which has the ability to accommodate enhancements and extensions [14]. It is the Hybrid Fuzzy weighted k Nearest Neighbor (HFWkNN) classifier introduced in [15]. It is a weighted and fuzzy extension to kNN based on fuzzy set theory.

We optimize HFWkNN to deal with imbalanced datasets and find greater results in the general classification of imbalanced data for both the minority and the majority classes. It determines the fuzzy membership function in favor of the minority class and creates a fuzzy equivalent relationship between the unlabeled instance and its k closest neighbors. In other words, it takes into account the fuzziness of an instance's closest neighbors, which can decrease the disturbance of the majority class to the minority class. The advantages of the neighbor weighted K nearest neighbor method are combined with fuzzy logic, i.e., the assignment of large weights to small classes and small weights to large classes. Fuzzy classification tends to more adequately classify objects as it defines how much of an object belongs to a class.

As presented in [15], the hyperparameters γ, ε and ε_{min} are introduced in the membership function of HFWkNN to increase the accuracy score, improve the performance and handle the class-imbalance. Weight-assignment technique is developed and combined with SMOTE for the class membership function of the HFWkNN of each neighbor, which learns the class weight for each training sample, to process imbalanced data. The minority class samples are given a higher weight to let the classifier concentrate on them.

The rest of the paper is organized as follows. Section II briefly discusses some related work. Section III presents the classification model with the class-imbalanced data. Section IV presents the experimental results, analysis and evaluation of the model. Finally, Section V concludes the paper and suggests future work.

II. Related Work

There are quite a few innovative solutions and methods which have been proposed by researchers to tackle the problem of class imbalance in classification problems. These methods either oversample the minor class or under-sample the majority class [16]. That is why these methods are sometimes called as sampling techniques. Although these methods are popular, they suffer from the problem of impacting the original distribution of the data. Nevertheless, there are approaches which deal with the issue of class imbalance while not impacting the original structure of the data. such approaches can be utilized by many classifiers such as those based on kNN.

kNN is one of the most utilized and quite popular classification algorithms. It is used in various classification problems and is considered as one of the top 10 algorithms in data mining [14]. But the classic kNN algorithm is not equipped enough if the dataset is imbalanced. Hence, to tackle the issue of imbalanced dataset for kNN algorithm, researchers have proposed quite a few distance or similarity based classification algorithms like kENN [25] and CCW-kNN [17]. But these methods are good for numerical data points.

Weighted kNN proved to performs well on imbalanced datasets. Dubey et al. [18] proposed class based weighted approach for performing classification on imbalanced dataset. In this approach, the distribution of the nearest neighbour was analysed and used to calculate the weights. Classic kNN is used to perform the initial classification and is used to get the respective weights for each of the classes in the classification problem. A hybrid approach was proposed by Patel et al. [19]. It tackles the class imbalance by assigning small weights to the majority class and large weights to the minority class. Tomasev and Mladeni_c [20] explored the hubness effect which is related to kNN in high-dimensional datasets, where minority class instances lead to higher misclassification errors. With low or medium dimensional datasets, majority class instances lead to misclassification.

Fuzzy solutions have been used for dealing with imbalanced dataset problems. However, not much work has been done in this area. Liu et al. [21] proposed a fuzzy kNN approach for unequally distributed dataset. The dataset had strong relationships between attributes, instances and classes. The approach utilized assigning sized memberships, similarity calculations and integration as the main methods. Sometimes, addressing the problem of data leaks in a classification model may result in data imbalance [22]. Ramentol et al. [23] have dealt with imbalanced dataset in a fuzzy-rough ordered weighted average nearest neighbour algorithm for binary classification. Six weight vectors and some indiscernibility relations are used with these weight vectors. Han and Mao [24] proposed an approach which utilizes fuzzy and rough properties of nearest neighbours data. The approach minimized the biasness owing to a membership function resulting in an advantage to the minority class.

In our case, we deal with data imbalance using easy to compute neighbour weighted with fuzzy kNN. We use the fuzzy kNN algorithm [25] to keep some of the nearest neighbours and utilizes their respective distances as key values. These distances are important as they help in finding the respective membership of the data instances into classes. This approach is further enhanced and refined by utilizing weights of different classes which are based on their respective sizes. The proposed solution strives to resolve the class imbalance by finding the membership function of the imbalanced data. This membership function uses the hyperparameters γ, ε and ε_{min} which we introduced in the proposed solution to tune up the classifier as they determine fuzzy membership and therefore lead the classifier to handle data imbalance. The fuzzy membership function was originally proposed by Keller et al. [25]. In our model, in addition to hypermeterizing the membership function, we use SMOTE with weight assignment function to get the membership of instances into all of the respective classes. Next, we present the proposed classification model that deals with the class-imbalanced datasets.

III. Optimizing Hybrid Fuzzy Weighted KNN with Imbalanced Data

The Optimized HFWkNN for imbalanced data handles the classification problem on the class-imbalanced mixed type datasets. It is an improved version of kNN and it combines fuzzy logic with weights to give more optimal results of prediction. We have presented the Optimized HFWkNN in full detail in [15]. Next, we summarize it for the benefit of the optimization process with the presence of data imbalance.

HFWkNN, as presented in [15], has two stages. In the first stage, the k nearest neighbors of the train set are calculated against itself. Once the neighbors are calculated, then the class memberships are calculated with the training set using Equation (1).

$$U_c(x_i) = \begin{cases} 0.51 + 0.49 \times \frac{n_c}{k} \times \varepsilon + \gamma & if\ x \in C \\ 0.49 \times \frac{n_c}{k} \times \varepsilon_{min} + \gamma & if\ x \notin C \end{cases} \quad (1)$$

In the membership assignment, we introduce the hyperparameters γ, ε and ε_{min} in a fuzzy membership to handle the class-imbalanced issue.

In the second stage, for each instance of the test set the k closest in the train set is calculated, based on the values of k. The resulting class is decided using Equation (2) instead of majority voting performed by kNN algorithm.

$$U_c(x_i) = \frac{\sum_{j=1}^{k}(U_{cj}) \times 1/\|x_i - x_j\|\verb|^|(2/(m-1))}{\sum_{j=1}^{k} 1/\|x_i - x_j\|\verb|^|(2/(m-1))} \quad (2)$$

The final class is obtained as the class with the greatest combined votes as a result of Equation (3).

$$C(x) = \text{argmax}_i\big(U_i(x)\big) \quad (3)$$

To increase the accuracy score, improve the performance and handle the class-imbalance issue in the data, the hyperparameters γ, ε and ε_{min} are introduced in the membership function of the proposed HFWkNN model [15]. These hyperparameters are optimized using two different methods which are grid search and random search. These two methods are turned into user defined parameterized callable functions to obtain the values of the three hyperparameters. They are based on the optimization process of the class weight parameter to find the weight for each class.

For instance, by using grid search, large weights are assigned to small classes and small weights are assigned to large classes to minimize the bias of the Optimized HFWkNN towards the majority class and avoid minority class. The following is the pseudocode of hyperparameter optimization procedure of HFWkNN using grid search method as we stated it in [27].

Step 1: **Initialize** the different parameters γ, ε and ε_{min}, with cv=5

Step 2: **Creating** the search space. we input the domain and the algorithm selects the next value for each hyperparameter in an ordered sequence.

Step 3: **Generate** a model using grid search. (grid search technique will construct many versions of HFWkNN with all the possible combinations of hyperparameter γ, ε and ε_{min} values that are defined)

Step 4: **Train** the Model.

Step 5: **Train phase:** Once the neighbors are calculated, give each neighbor a weight as the inverse distance of its Euclidean distance from that training data, then find memberships of training data into each class using Equation (1).
The parameters **γ, ε** and $\boldsymbol{\varepsilon_{min}}$ are introduced in the membership function to give it a weight in all classes. This is to minimize the bias of the classifier towards majority class and avoid minority class.

Step 6: **Test phase:** Once the neighbors are calculated, the predicted class is decided as shown in Equation (2) to define the degree of membership of x in each class c.
i = 1, 2, 3, ...C
j = 1, 2, 3, …k
C is the number of classes, k is the number of nearest neighbors and m is the parameter of fuzzy strength.

Step 7: The final class is obtained as the class with the greatest combined vote.
Classifier assigns x, using Equation (3), as belonging to the class label whose fuzzy membership for x_i is maximum.

Step 8: **Calculate** the model accuracy and save the model configuration and accuracy.

Step 9: **Check** if stopping criteria is not complete (no. of iterations end) **Updating** parameter values and **Return** back to step 3.

Step 10: **Get** and **report** the optimal value of the parameters and position of the model with high accuracy. Output the settings that achieved the highest score in the validation procedure.

To treat class imbalance, we reduce the bias inherent in the learning procedure and increase the sampling weights for the minority class. The weight-assignment technique is introduced for the class membership function of each neighbor in HFWkNN. Therefore, it learns the class weight for each training sample to process imbalanced data. Weights are assigned to the selected samples according to their importance in the data. The minority class samples are given a higher weight to let the classifier concentrate on them. The minority class decision space is expanded to allow HFWkNN to have a higher prediction on unknown samples of minority class. It also avoids the overfitting problem. Furthermore, combining class weighting with over-sampling of the minority class using SMOTE improves the classification accuracy of minority data without sacrificing the accuracy of the majority class. SMOTE is used to pre-process the dataset.

This over sampling, SMOTE + Weighting_assignment, strategy, can tune HFWkNN towards a certain performance measure of interest with only moderate computational overhead. Each observation is weighted based on the class to which it belongs. The effect of minority class observations is increased simply by a larger weight of these instances and vice versa for majority class observations. This is similar to sampling-based approaches. It takes advantage of two efficient techniques: SMOTE as it is used at the pre-processing phase and class weighting assignment which is used to adjust the class distributions of the imbalanced datasets and respectively weight the base classifiers. This proposed strategy (SMOTE with Weight Assignment) is shown in Fig. 1 and is summarized as follows:

1) SMOTE as a common general-purpose approach handles data imbalance of the dataset at the pre-processing phase of data processing.

2) We get the weights of the class weight parameter. It is based on the optimization process where we use grid search. Grid search assigns large weights to small classes and small weights to large classes. Such an assignment minimizes the bias of the classifier towards majority class and avoid minority class. The assignments of weights in such a manner assures a better classification performance and tackling of the imbalanced dataset.

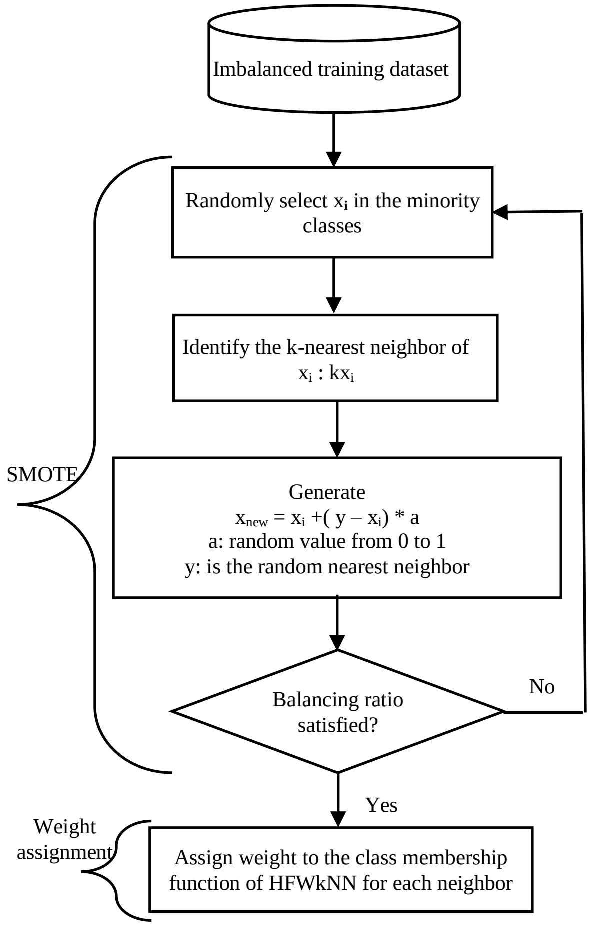


Fig. 1. SMOTE with Weight Assignment.

This strategy employs oversampling, weighting and strengthening. Notice that this improves the minority class samples in boundary region (or uncertain area). It extends the coverage space of minority class samples in boundary region and improves the confidence degree of decision rules without having much impact on the decision space of majority class. Hence the accuracy of HFWkNN is improved. Next, we present the conducted experiments on the optimized HFWkNN based on this strategy and their results.

IV. Experimental Results

In order to verify the performance of the proposed model on benchmark datasets, we have used five datasets from UCI [26]. These datasets are known to be imbalanced by the unequal distribution of instances into classes. Haberman dataset describes the five year or greater survival of breast cancer patients and mostly contains patients who survive. Pima dataset collected originally from the National Institute of Diabetes and Digestive and Kidney Diseases.

The objective of the dataset is to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset. Ionosphere is a radar data collected by a system in Goose Bay, Labrador. Breast Cancer dataset is a classic and binary classification dataset. Readmission dataset represents 10 years of clinical care for diabetic patients at 130 US hospitals and integrated delivery networks. It includes over 50 features representing patient and hospital outcomes.

Table I provides a summary of these datasets. They vary in the number of instances, number of attributes, class label ratio, and minority class percentage. Based on the minority class percentage, we can see that the datasets are highly imbalanced. For Haberman dataset only 26.5% of patients did not survive. For Pima dataset only 34.9% of patient has diabetes. While for Ionosphere dataset only 35.9% of records labeled as bad radar. For Breast Cancer dataset only 37.3% of patients belong to the malignant class. For Readmission dataset with respect to readmissions only 11% of patients have been readmitted within 30-days.

We compare our approach, the Optimized HFWkNN for imbalanced dataset with kNN, Weighted kNN (WkNN) and Fuzzy kNN (FkNN). We conducted a 5-fold cross validation for each dataset to evaluate the performance of all the 4 algorithms. We obtained and compared the results of Recall and F-value of minority class and the Area Under the ROC Curve (AUC) for each experiment. AUC indicates the overall classification performance and the AUC of a perfect classifier equals to 1, a bad one is less than 0.5. A good classification algorithm usually has a higher AUC.

In all our experiments, we set k to 3 since all the 4 classifiers are based on kNN. The cleaned data was randomized to avoid any selection bias and divided into two parts: 80:20 Training and Test data. This allowed us to train the model on 80% of the data and use an additional 20% to test the performance of the model. A 5-fold cross validation is used to avoid over-fitting on the training data in the evaluated models. The models are trained and evaluated using the same data to ensure fair performance comparison.

Tables II, III, and IV show the results of Recall, F-value of minority class and the Area Under the ROC Curve (AUC) measurements for all of the algorithms with the different datasets. Recall results in Table II indicate that the Optimized HFWkNN outperformed kNN, FkNN. It only outperformed WkNN in the case of the Readmission dataset.

This is due to two factors; first, the improvement occurred in the case of the Readmission dataset because it is larger and richer in terms of attributes than the other 4 datasets. Second, WkNN as it considered weights has shown better performance than kNN and FkNN since they do not consider weights. Nevertheless, the 3 algorithms outperformed the basic kNN since they are fuzzy and weighted.

TABLE I. Short Description of the Datasets

Datasets	Number of instances	Number of attributes	Class labels (Minority: Majority)	Percent of the minority class
Haberman	306	3	2:1	26.5%
Pima	768	8	1:0	34.9%
Ionosphere	351	34	1:0	35.9%
Breast Cancer	570	8	1:0	37.3%
Readmission dataset	101,766	50	1:0	11.16%

TABLE II. RECALL RESULTS FOR THE OPTIMIZED HFWKNN IN COMPARISON WITH OTHER ALGORITHMS

Dataset	Recall			
	kNN	FkNN	WkNN	Optimized HFWkNN
Haberman	69%	73%	74%	74%
Pima	72%	75%	77%	77%
Ionosphere	82%	82%	83%	83%
Breast Cancer	87%	89%	89%	89%
Readmission dataset	75%	76%	74%	80%

The F-value results in Table III indicate that the Optimized HFWkNN outperformed kNN, FkNN and marginally outperformed them in the case of the Ionosphere dataset. This is due to the limited size and number of features in this dataset. Again, it outperformed WkNN in the case of the Readmission dataset due to two factors.

First, the improvement occurred in the case of the Readmission dataset because it is larger and richer in terms of classification-considered features than the other 4 datasets.

Second, WkNN, as it considered weights, has shown better performance than kNN and FkNN since they do not consider weights.

The AUC results in Table IV indicate that the Optimized HFWkNN has a high degree of class separability, i.e., it has high probability of distinguishing between classes. The results also indicate that HFWkNN outperformed kNN, FkNN and WkNN in its ability of class separability especially in the case of Ionosphere and Breast Cancer datasets. This is due to the limited number of classes in these two datasets.

TABLE III. F-VALUE RESULTS FOR THE OPTIMIZED HFWKNN IN COMPARISON WITH OTHER ALGORITHMS

Dataset	F-value			
	kNN	FkNN	WkNN	Optimized HFWkNN
Haberman	71%	76%	77%	77%
Pima	72%	75%	77%	77%
Ionosphere	83%	83%	84%	84%
Breast Cancer	87%	89%	89%	90%
Readmission dataset	75%	76%	74%	80%

TABLE IV. AUC RESULTS FOR THE OPTIMIZED HFWKNN IN COMPARISON WITH OTHER ALGORITHMS

Dataset	AUC			
	kNN	FkNN	WkNN	Optimized HFWkNN
Haberman	63%	67%	70%	70%
Pima	67%	70%	72%	72%
Ionosphere	85%	88%	88%	89%
Breast Cancer	87%	89%	89%	90%
Readmission dataset	76%	76%	74%	79%

The overall results of Recall, F-value and AUC show running the 4 models on the 5 datasets achieved better performance for the proposed Optimized HFWkNN than kNN, WkNN, and FkNN. It increases the classification performance of the minority class compared to the other 3 models. Its performance on the entire datasets is better than the other 3 models. Furthermore, these results indicate that kNN, in all cases has the lowest performance compared to FkNN, WkNN and our Optimized HFWkNN. Therefore, these extensions to kNN are justified and necessary.

It is worth mentioning again that the improvements of the Optimized HFWkNN over the other 3 models are due to combining SMOTE and class weighting assignment. SMOTE has succeeded in improving the accuracy of minority classes. Therefore, the Optimized HFWkNN was able to better model the minority class in the dataset by presenting not only the minority class instances, but also a broader representation of such instances. Such a representation resulted in improving the overall accuracy of the Optimized HFWkNN by concentrating on the minimal cases of the minority, positive classes as well as by properly modelling such classes.

Finally, since the Optimized HFWkNN is based on kNN, its complexity does not differ much from that of kNN which is O(n). Taking the time of computing fuzzy membership grade of training and test samples using Equation (2) above, the time for testing fuzzy membership degree of all classes, and the time of SMOTE (although it is consumed once at the processing stage) and the time of class weighting assignment. These collective times add extra overhead but do not change the overall complexity of the model from that of kNN. Nevertheless, there still a need to perform a complete complexity analysis of HFWkNN.

V. CONCLUSION

We have presented the Optimized HFWkNN classification model dealing with imbalanced datasets. The model has used three hyperparameters γ, ε and εmin that are introduced in the membership function to give it more general character and are tuned to give appropriate membership values for each class and help to balance the dataset. The model also has combined the method of class weighting with over-sampling of the minority class, SMOTE, to improve the classification accuracy of minority class without sacrificing the accuracy of the majority class. This has led to better results in the general classification of imbalanced data for both the minority and the majority.

Experimental results have shown higher average performance for the Optimized HFWkNN than kNN, Fuzzy kNN, and Weighted kNN. Results of Recall, F-value and AUC measurements for the different datasets are higher with the optimized HFWkNN model than the other 3 models. For example, Recall, F-value and AUC measurements with the Readmission dataset are 80%, 80% and 79% respectively, which are higher than those of the other 3 algorithms. These results also prove that the proposed model lead to better overall classification performance on the complete datasets than the other 3 algorithms.

The proposed model can be extended for multiclass and large size datasets with different strategies to construct the fuzzy membership function in addition to the three hyperparameters γ, ε and ε_{min}. Although we have mentioned that the performance of the model is in line with that of kNN, the induced overhead due to computing and ranking fuzzy membership as well as the overhead due SMOTE and class weighting assignment need to be investigated. Finally, the model can be extended to other classification algorithms such SVM based on fuzzy similarities, weights and distances.

References

[1] Kantardzic, M., Data mining: concepts, models, methods, and algorithms. John Wiley & Sons, 2011.

[2] Han, J., J. Pei, and M. Kamber, Data mining: concepts and techniques. Elsevier, 2011.

[3] Patel, H., et al., A review on classification of imbalanced data for wireless sensor networks. 2020. 16(4).

[4] Fernández, A., M.J. del Jesus, and F. Herrera, Hierarchical fuzzy rule based classification systems with genetic rule selection for imbalanced data-sets. 2009. 50(3): p. 561-577.

[5] Kotsiantis, S., et al., Handling imbalanced datasets: A review. GESTS International Transactions on Computer Science and Engineering. 2006. 30(1): p. 25-36.

[6] Abd Elrahman, S.M., A. Abraham, and I. Computing, A review of class imbalance problem. Journal of Network and Innovative Computing. 2013. 1(2013): p. 332-340.

[7] Weiss, G., H. He, and Y. Ma, Foundations of Imbalanced Learning. Imbalanced Learning: Foundations, Algorithms, and Applications. Hoboken. 2013, NJ, USA: John Wiley & Sons.

[8] Chawla, N.V., et al., SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research 2002. 16: p. 321-357.

[9] Barandela, R., et al., Strategies for learning in class imbalance problems. Pattern Recognition. 2003. 36(3): p. 849-851.

[10] López, V., et al., An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. Information sciences. 2013. 250: p. 113-141.

[11] Estabrooks, A., Jo, T., and Japkowicz, N. A multiple resampling method for learning from imbalanced data sets. Computational intelligence. 2004. 20(1): p. 18-36.

[12] Cover, T. and P. Hart, Nearest neighbor pattern classification. IEEE transactions on information theory. IEEE transactions on information theory. 1967. 13(1): p. 21-27.

[13] Verdhan, V., Supervised Learning for Classification Problems, in Supervised Learning with Python. 2020, Springer. p. 117-190.

[14] Wu, X., et al., Top 10 algorithms in data mining. Knowledge and information systems. 2008. 14(1): p. 1-37.

[15] Bahanshal, S. and Kim B. Hybrid Fuzzy Weighted K-Nearest Neighbor to Predict Hospital Readmission for Diabetic Patients. in 2020 IEEE Symposium Series on Computational Intelligence (SSCI). 2020. IEEE.

[16] García, S. and Herrera, F. Evolutionary undersampling for classification with imbalanced datasets: Proposals and taxonomy. Evolutionary computation. 2009. 17(3): p. 275-306.

[17] Liu, W. and S. Chawla. Class confidence weighted kNN algorithms for imbalanced data sets. in Pacific-Asia conference on knowledge discovery and data mining. 2011. Springer.

[18] Dubey, H. and V. Pudi. Class based weighted k-nearest neighbor over imbalance dataset. in Pacific-Asia conference on knowledge discovery and data mining. 2013. Springer.

[19] Patel, H. and G. Thakur. A hybrid weighted nearest neighbor approach to mine imbalanced data. in Proceedings of the International Conference on Data Science (ICDATA). 2016. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).

[20] Tomašev, N. and Mladenić, D. Class imbalance and the curse of minority hubs. Knowledge-Based Systems. 2013. 53: p. 157-172.

[21] Liu, C., L. Cao, and S.Y. Philip. Coupled fuzzy k-nearest neighbors classification of imbalanced non-IID categorical data. in 2014 international joint conference on neural networks (IJCNN). 2014. IEEE.

[22] Rezqa, E. Y. and Baraka, R. S. Document classification based on metadata and keywords extraction," 2021 Palestinian International Conference on Information and Communication Technology (PICICT), 2021. p. 18-24.

[23] Ramentol, E., et al., IFROWANN: imbalanced fuzzy-rough ordered weighted average nearest neighbor classification. IEEE Transactions on Fuzzy Systems. 2014. 23(5): p. 1622-1637.

[24] Han, H. and B. Mao. Fuzzy-rough k-nearest neighbor algorithm for imbalanced data sets learning. in 2010 seventh international conference on fuzzy systems and knowledge discovery. 2010. IEEE.

[25] Keller, J.M., et al., A fuzzy k-nearest neighbor algorithm. IEEE transactions on systems, man, and cybernetics 1985. 4: p. 580-585.

[26] Dua, D. and C. Graff, UCI machine learning repository. School of Information and Computer Science, University of California, Irvine, CA. 2019.

[27] Bahanshal, S. and Kim, B. An optimized hybrid fuzzy weighted k-Nearest Neighbor to predict hospital readmission for diabetic patients. 2021 IEEE 13th International Conference on Computer Research and Development (ICCRD), 2021, pp. 115-120.

SPKP: A Web-based Application System for Managing Mental Health in Higher Institution

Mohamad Fadli Zolkipli, Zahidah Mohamad Said, Massudi Mahmuddin
School of Computing, Universiti Utara Malaysia
Kedah, Malaysia

***Abstract*—Client psychology profile is one of the methods used by UUM's Counselling Centre to analyze client psychological health conditions. This psychological profile is in a physical form which means the questions that need to be answered are on the paper. This psychology profile consists of three types of psychological modules whereby each module consists of questions related to psychology. The estimated time taken to answer this psychological profile is around 10 minutes. This physical form method may cause data to be lost and it will create issues when counselors want to retrieve the data back. This paper is aimed to develop Sistem Profil Kesejahteraan Psikologi (SPKP) for Counselling Centre of Universiti Utara Malaysia (UUM) in helping to know the client's psychological health before meeting up with the counselor. By focusing on analyzing Counselling Centre client's data, the basis of this web application system is to create a new space that will help Counselling Centre to improve the way they collect data, store data, and also improve their quality in time management whereby all those data can be collect, store, retrieve and analyze in a single click. By developing this system, Counselling Centre can monitor clients' psychological health before they meet up with the counselor.**

Keywords—Psychology tests; psychology health; counselling; counselor

I. Introduction

Conducting a psychological test is not something that people like to do. Further, the task of getting mental health checkups is quite time-consuming and sometimes people with mental health problems would not like to seek help from others. Even though counselors are like to accept people with open hands, there is still some gap between the client and the counselor. To make things harder, people with low self-esteem and anxiety would think multiple times before asking for people help. This is because, they might feel scared, ashamed, or even afraid of what other people would think about them after getting the mental health checkup. Psychological testing and assessment are a tools use by psychologist or counselor to test and measure a client's behavior to come out with a specific diagnosis and guide treatment. This test and assessment are being conduct for a wide variety of reasons. For example, children who is experiencing difficulty in school. The children may undergo aptitude testing or test for learning disabilities. This test can help a psychologist to understand more about the children condition. The test involves the use of questionnaires or checklist [1].

As time goes by, nowadays everything is just under the fingertip. What do you want to do, what do you want to search for, what do you want to buy is all over the internet? There is no need for people to travel from one place to another just to buy their food and etc. For example, person A wants to buy cloth. He or she just needs to access any website that sells cloth and just buy it from there; Simple and easy, and also time-saving.

It is not that different from psychology. Before this, counselors conduct psychological tests through pen and paper tests. This method is good in terms of meeting the client directly face to face. However, as Covid-19 has taken over the world, the higher institution needs to stop operating face-to-face and change the teaching and learning method online. This means that UUM Counselling Centre also needs to stop operating. It also means that students are no longer able to conduct psychological test at UUM Counselling Centre. Moreover, the way UUM Counselling Centre conducts psychological tests is quite time consuming as they have 3 modules and each module consist of 10 to 20 questions which resulting in a lot of time consuming. Not only that, but the clients of UUM Counselling Centre also need to manually calculate the test results, and later on the test result would be kept in file storage rooms which might have resulted in the test reports might be lost and hard to retrieve.

Due to this problem, there is an urgent need for solutions that can manage the psychological tests efficiently. The emergence of digital technology in human life may open an opportunity for providing an effective way of managing the psychological health of people. Moreover, findings reveal that computer-based models outperform humans in a critical social-cognitive task: personality judgment [2]. It can be stated that, computer is more accurate in judging personality of person. Therefore, this paper is going to investigate the potential of computer as a platform to managing psychological health. This paper is aimed to develop Sistem Profil Kesejahteraan Psikologi (SPKP) for Counselling Centre of Universiti Utara Malaysia (UUM) in helping to know the student's psychological health before meeting up with the counselor. As a result, SPKP prototype of a web-based application for managing psychological health was developed and evaluated. The study contributes towards an understanding the system requirements for such web-based application and could be a reference model for developers and researchers to improve and managing psychological health. The next section describes the background and related studies. Next, the section describes the design and development of SPKP. The subsequent section explains the usability evaluation of SPKP. The last section in this paper concludes the study and lists the future works.

II. Background and Related Studies

This section describes the background of psychological tests. Later, this section discusses the role of information technology in facilitating the creation and management of psychological tests. Although some students and staff found that doing psychological tests is easy and nothing to be worried about, in some cases, there are some students and staff who might have anxiety or depression or any related mental health problems that might find it is hard for them to conduct even one psychological test. Social stressors are amongst the most reliable forms of stress in humans and other species [3]. Accessing help, worries about confidentiality and trust, a preference for informal sources of help, and stigma are all barriers to obtaining help [4]. It is because of the stigma that, people who seek help from counselor or people who get medical treatment for mental health is crazy. Because of that, most people tend to just keep the problem to themselves and make their mental health worsen. The modern way of conducting psychological tests could make use of web-based applications where users use computers for managing their mental health. A detailed discussion of the concept and related studies are presented in the following paragraph of this section.

According to the World Health Organization, over 800,000 individuals commit suicide each year [5]. Latent components that are not observable in nature are evaluated using social and personality psychology exams. For instance, a self-report test could be created to assess belief in a just world, right-wing authoritarianism, or personality traits [6]. Sometimes, what we see is not the truth and we cannot differentiate between people with no mental illness and people with mental illness. Sometimes, people with health problems are more exposed on getting mental health problems too. Depression affects one out of every five individuals with coronary artery disease or heart failure, a rate that is at least three times higher than the overall population [7]. Because everyone is the same. The only way to differentiate is by conducting psychological tests. However, some people are not likely to conduct the tests because they feeling of ashamed and afraid would be called someone with mental health problems afterward. However, receiving psychological assistance is critical because it reduces the long-term detrimental impacts of mental illness [8]. Thus, seeking psychological help is very important in order to maintain a healthy mind.

The desire to be alone is a common behavior among people suffering from mental diseases. As a result of their seclusion, people with mental illnesses turn to social media sites like Twitter to freely discuss their illnesses [9]. However, sometimes the response received is not as expected. There are various reasons why online comments may be more detrimental than responses received in person [10]. When the discloser receives the desired response, requesting support in stigmatized situations is beneficial, but it comes with societal harms [11]. As a result, people feel scared and do not always reveal or seek help when they are in need.

Due to the Covid-19, the majority of the higher institution needs to stop operating as well as the UUM Counselling Centre. Currently, UUM Counselling Centre is conducting psychological tests through pen and paper tests as shown in Fig. 1.

Nama : ________________ Tarikh : ______________

SARINGAN STATUS KESIHATAN MENTAL 20 (SSKM 20)

Untuk setiap soalana yang berikut sila pilih satu jawapan yang mengganbarkan **pengalaman atau perasaan** anda dengan paling tepat dan tandakan (√) di dalam kotak yang disediakan.

Di dalam tempoh **satu bulan** yang lepas pernahkah anda mempunyai **pengalaman atau perasaan** yang berikut:

Soalan	0 Tiada	1 Kadang-kadang	2 Kerap	3 Sentiasa
1.Sedih (atau susah hati)				
2. Mudah marah				
3. Tidak minat bercakap				
4. Kerap terjaga dari tidur				
5. Sakit-sakit badan/sendi				
6. Hilang minat pada perkara yang biasa diminati				
7. Menangis atau rasa hendak menangis				
8. Rasa sunyi				
9. Takut terhadap sesuatu objek atau keadaan				
10. Badan lemah, lesu atau tidak bertenaga				
11. Tidak suka bercampur dengan orang lain (menyendiri)				
12. Perasaan ingin mati				
13. Kurang daya ingatan atau mudah terlupa				
14. Otot tegang				
15. Keyakinan pada diri sendiri yang berlebihan				
16. Buah fikiran yang pantas, seolah berlumba-lumba				
17. Kurang yakin pada diri sendiri				
18. Rasa tiada jalan keluar				
19. Perasaan terlalu seronok				
20. Susah hendak menyesuaikan diri dengan keadaan				
JUMLAH				

Fig. 1. An Example of One of the Psychological Tests.

According to a senior psychology officer at UUM Counselling Centre, this method of conducting psychological tests is not efficient and a lot of times are needed just to conduct one test. This is because the client of UUM Counselling Centre needs to manually calculate the test result which required a lot of time to do so. Not only that, the test results would be kept manually in the files storage rooms which results in the test result might be lost and unretrieved. Moreover, this method is also not environmentally friendly as they need a lot of paper to print the psychological tests questions. It is because, to make one ton of paper, twelve trees, 540,000 gallons of water, and a variety of chemicals are required [12]. It means that this method is not environmentally friendly and it needs to be changed.

However, the method of conducting online psychological tests is not been widely discussed by the previous researchers. Therefore, only a few data can be collected to make a comparison between the two psychological tests method. Another method of conducting psychological tests is by using a mobile device as a platform for conducting psychological tests. Conducting psychological tests personally on mobile devices allows people to replay the tests as many times as needed. Not only that but they will also be received immediate feedback after conducting the tests [13]. Moreover, people with limitations such as handicapped and hard to moving from one place to another can easily use the system without worrying to travel to the specific places just to conduct the psychological tests. SPKP is also good in terms of allowing users to conduct psychological tests multiple times. It is also can help students and staff in saving money and saving time where they no longer need to travel to UUM Counseling Centre just to conduct the psychological tests. Moreover, they can get the test results immediately after done answering the psychological tests questions. It shows that any method of psychological tests when being conducted online or computerized-based platform, will help in efficiency, flexibility, mobility, saving money, and saving time.

The future outcome of the mental health technology revolution depends on accelerating innovation and enhancing the path to adoption by challenging the normal ways of thinking [14]. The emergence of digital devices such as computer, tablets, and laptop has shed light on a more flexible and easier way of conducting psychological tests. A web-based psychological test is a new method that is going to be implemented by UUM Counselling Centre. This method is not only more efficient in a matter of time but also environmentally friendly. This means that by using this method people can also help to save the world from tree cutting and global warming. Therefore, there is a need for research and development of a web-based application for conducting psychological tests that could minimize the effort and time, especially for people who are afraid of seeking counselor help for their mental health. The study presented in this paper aims at designing and developing a web-based application that could help the users in conducting and managing psychological tests.

III. Methodology of the Study

The methodology of development for this paper is Waterfall Model. The rationale why creates this paper using the Waterfall Model is because to relies on paper documentation, like technical specifications to define the scope of work clearly before it is started. Furthermore, the use of Waterfall Model is for ensuring that Counselling Centre deliverable meets expectations. Besides that, the Waterfall Model methodology has been chosen as a methodology to complete the system because it contains planning, analyzing, designing, implementation, testing, and maintenance as illustrated in Fig. 2. This system uses waterfall model as the methodology because this model works well for smaller papers where requirements are very well understood and phases are processed and completed one at a time [15].

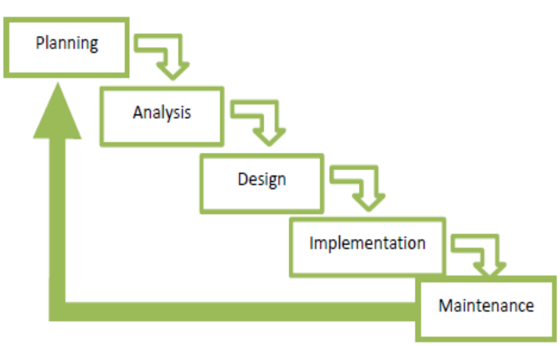


Fig. 2. The Phase of Waterfall Model [14].

- Planning: In the preliminary stage, planning begins with asking about problems that occur at the Counselling Centre. Student collects information about the new system that wanted to be developed by discussing with counselor and Counselling Centre's director. From the results of the discussion, student learn about what problems occur and face by Counselling Centre and to find solution to that problems. The goal of planning is to facilitate this system to be built smoothly. Therefore, the output in first phase is the problems faced by the Counselling Centre will be detected.
- Analysis: Students perform analysis by finding related work with the system to have more information about the current system by making a research at google scholar, to collect more info about similar system that related with SPKP more deeply to gather information and finalizing requirements with supervisor. It requires documentation time for paper technical proposals that need to be prepared. From there it must need to contact the client to ask what the client wants in the system. The output that will be gained from the analysis is student will be able to finalize system requirements and understand what the client wants in the system.
- Design: When requirements are known, a preliminary design or quick design for SPKP will be created. It is not a detailed design and includes only the important aspects of the system. Student will develop a prototype design which is low-fidelity prototype for a system to present for teachers and client. The output for design is to develop low-fidelity prototype for counselor to make sure they know how the system is going to function.
- Implementation: Student will implement it for the last time before officially present to the counselor and Counselling Centre's director. Lastly, all the related documents and final reports are then completed and submitted to the supervisor and the lecturer. The output that will be gained in implementation is the documents already completed and submitted to the supervisor and the lecturer.
- Testing: The paper development included software testing and user testing. A software testing is tested by the student intend to find syntax and logic error. The syntax and logic error on the system will be corrected and rebuild, this process ends until the system is robust. Then, the client can test the system in terms to evaluate usability, functionality, performance, and satisfaction. The evaluation result of the user testing recorded for future improvement and enhancement. The output that will get is the system must be tested until success in terms of usability, functionality, performance, and satisfaction.
- Maintenance: For the purpose of this paper, this maintenance phase will not be implemented. The software is operated to stay effective and relevant. It will be made to the initial software in order that will remains a workable solution that is free from bugs. The

output is the system must always be in a good condition without any failure, security system must be updated from time to time, ensuring high system quality and ease clients to key-in data and get their psychological report.

IV. Design and Development of Sistem Profil Kesejahteraan Psikologi

This section describes the design and development of a web-based application for conducting and managing psychological tests. This section is divided into two sub-sections; (1) the requirements of the web-based application for conducting and managing psychological tests and (2) the prototype development of SPKP.

The Requirements of Sistem Profil Kesejahteraan Psikologi as followed:

A requirement gathering process was carried out using two methods that are interviewing UUM Counselling Centre's counselor and analyzing documents and web-based applications from the Internet that are related to psychological tests. The counselor was asked fewer questions primarily on the feature of the future developed web-based application of psychological tests. The example of the questions is, what are the characteristics of a web-based application of psychological tests that you would like to have, how would the interface design look like, and what are the things that you would like to put in the web-based application. The opinion was recorded, and the requirements were elicited.

For the secondary requirements gathering process, the documents were searched using Google Scholar by providing keywords primarily "psychological test", "mental health", "depression and anxiety", "digital world" and "global warming". The documents were analyzed to elicit the requirements for a web-based application that can conduct and manage psychological tests. Table I lists the requirements and their priority produced from the requirements gathering process. The requirements include new user registration, login to the system, answer submission, report viewing, and report managing.

The requirements presented in Table I were translated into the computer system functionality. The next process is visualizing and modelling the requirements of the web-based application using the appropriate modelling method and tools. In this work, the Unified Modelling Language (UML) was used to visualize and model the requirements. The models used in this work are two behavioral diagrams namely use case and activity diagrams, and a class diagram that represents the structural components of the app. Fig. 3 illustrates the use case diagram and the communications between the use cases and the actor for a web-based application that can be used conduct and manage psychological tests.

TABLE I. List of Requirements for Conducting and Managing Psychological Tests

ID	Requirements Description	Priority
1	**Register**	
1.1	A new user must register before login to the system	Mandatory
1.2	System must display a page that allow user to key in their details: a) Name b) Role c) Email d) Password e) Confirm Password	Mandatory
1.3	The system must save all the information submitted by the user	Mandatory
1.4	User can cancel their registration by clicking the "Return to login" button if they want to	Optional
2	**Login**	
2.1	User must login into SPKP system by key in their email and password	Mandatory
2.2	If user entered incorrect email or password, the system will alert user with error message	Mandatory
2.3	If user forget his/her email or password, the user can click "forget password" button to reset it	Desirable
2.4	User can cancel their log in if they want to	Optional
3	**Submit answer**	
3.1	System must display a page that allow user to answer three psychological test questions	Mandatory
3.2	The system must save all the submitted answers	Mandatory
3.3	User can edit the answers before submit the answers if they want to	Optional
3.4	Once answers submitted, user can no longer edit the answers	Mandatory
4	**View report**	
4.1	The system should allow user to view their report result	Mandatory
4.2	User able to print the report result	Mandatory
5	**Manage report**	
5.1	The system must allow counsellor to view the report result	Mandatory
5.2	The system provide function for the counsellor to search for the specific client's report	Desirable
5.3	The system must allow counsellor to print the report results	Mandatory
5.4	The system should process all submitted answers and come out with the psychological test report results	Mandatory

The structural components of a web-based application for conducting and managing psychological tests are represented in a class diagram as illustrated in Fig. 4. The class diagram in Fig. 4 shows the attributes and operations of the web-based application.

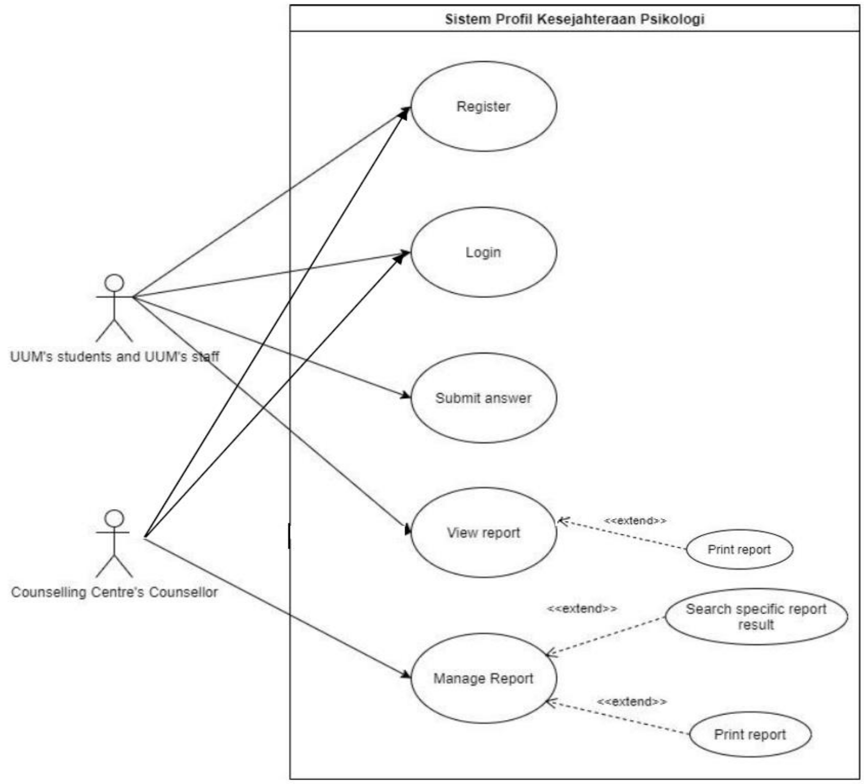


Fig. 3. The use Case Diagram of a Web-based Application for Conducting and Managing Psychological Tests.

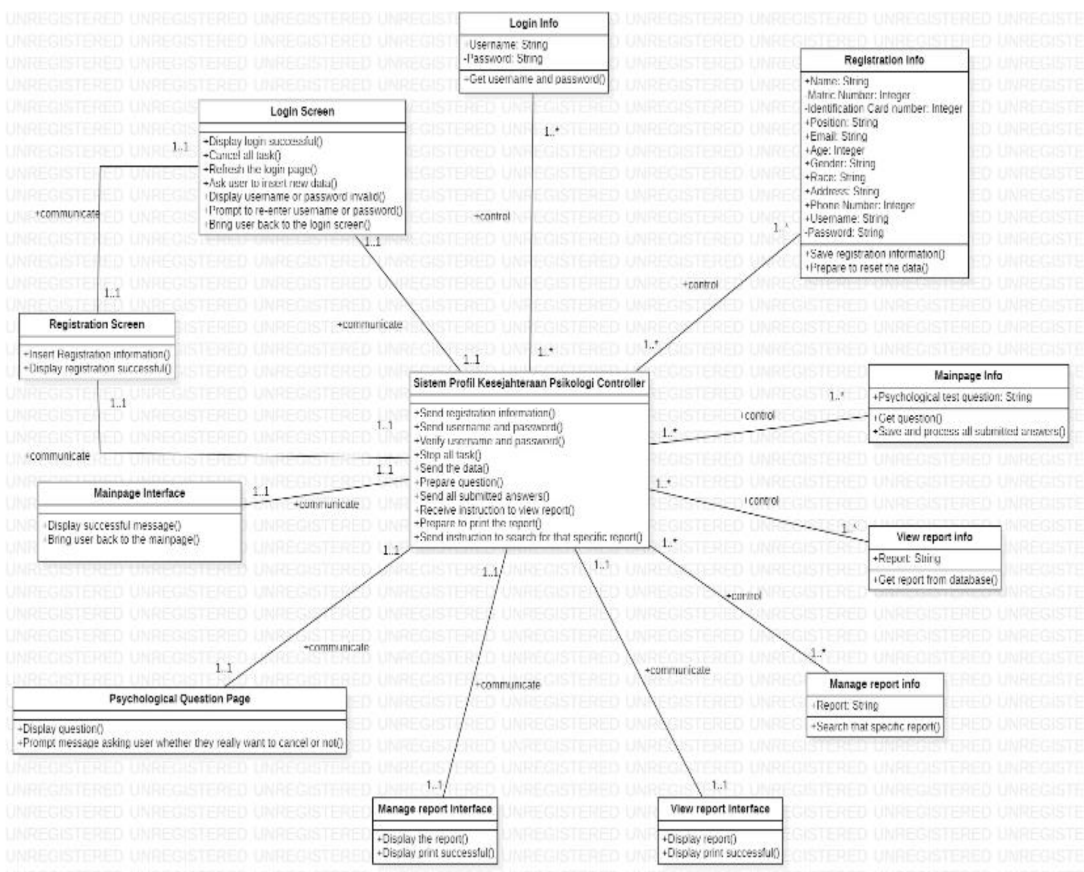


Fig. 4. The Class Diagram of a Web-based Application for Conducting and Managing Psychological Tests.

V. The Sistem Profil Kesejahteraan Psikologi Prototype Development

A prototype of a web-based application for managing mental health named SPKP was developed. It represents the requirements explained in the previous subsection. Software prototyping is a standard way of demonstrating the software requirements so that further comments and suggestions could be obtained from the users based on their experience in interacting with the prototype. The Visual Studio Code was used as the main platform to write the code for this system. Further, XAMPP was used to create the web server and phpMyAdmin as the database for the system. Print screen in Fig. 5 to 13 shows the selected interface of SPKP.

Fig. 5 interface shows the landing page of the system. This is the first interface user will see once they get into the system. This interface consists of four buttons on the left side which is the homepage or the landing page, about us, our counselor, and the login button. The about us button will bring the user to a new section where it will show information about what the counseling center does and its vision and mission. Our counselor button will list all counseling center's staff names, phone numbers, positions, and pictures. The last button is the login button for the user to log in to the system. The interface design is simple and easy for users to understand how to use it. There is no random button on the interface which might cause users to feel uncomfortable using the system.

Fig. 6 shows the login page of the system where users need to key in their email and password. Before login, the user needs to register first by clicking the register now hyperlink. If user forgot their password, they can reset the password back by clicking the click here hyperlink. The hyperlink is in the blue color so that users can easily see it and to make sure that users do not accidentally click it while using the system. The login button is in a red color box with white text so that it will enhance the button's look while the background color is black. This design will help users to recognize the button and know that the button is for login even without reading the text on it.

Fig. 7 shows the registration page of the system. Users need to register into the system by keying in their details such as name, matric no/ staff no, phone no, email address, and password. The interface design of the registration page is similar to the login page to keep the consistent look of the design. Hence, the design will not confuse the user whether the system is the same system or a different system if the interface design keeps on changing. All user needs to register or else they are not allowed to access the system.

Fig. 8 shows the main page of the system by user view. This interface can only be viewed by the students and the staff. The background layout is in pastel color to keep a calm atmosphere for the user while using the system. The interface has three columns representing three psychological tests modules which are the SSOSH test in a blue column, the SSKM test in a green column, and the D.A.S.S test in a red column. The reason why each column is in a different color is to differentiate between the modules. It means that regular users can easily recognize which column is for what psychological tests. Users can choose to answer which modules to get the psychological test results. On the left side of the interface, there are three options in white text in a grey color layout which are home, report, and profile. Once the user finished answering the psychological tests questions, the user can view the results and the history of the results of the previous tests by clicking the report button. User also can edit their details such as passwords by clicking the profile button. The reason why the grey layout is on the left side is to put a systematic look at the interface.

Fig. 5. The Landing Page of the System.

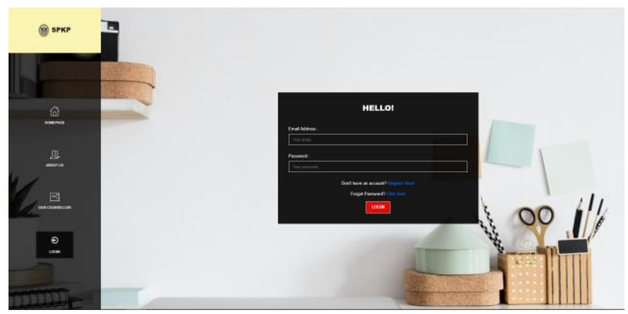


Fig. 6. The Login Page of the System.

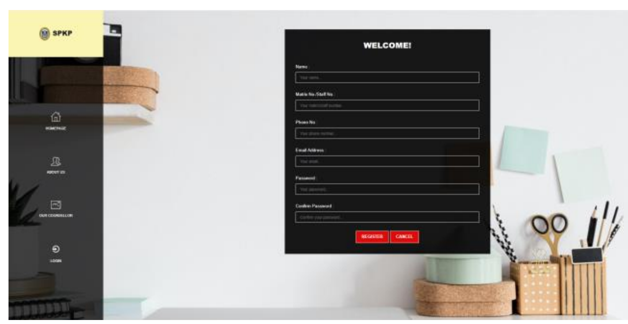


Fig. 7. The Registration Page of the System.

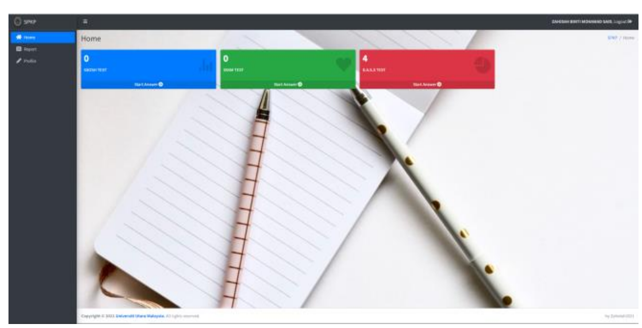


Fig. 8. The User Mainpage of the System.

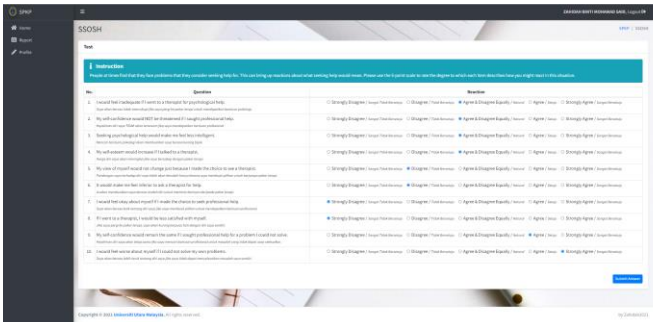


Fig. 9. The Interface for SSOSH Psychological Test Question.

Fig. 9 shows the interface of one of the psychological tests question which is the SSOSH test. This test consists of 10 questions. The questions are aligned in a box to create a systematic and nice look whereas the user can easily see all questions at once. The user does not need to click the next button just to see the next question. All question is compulsory to be answered or else the user cannot click the submit answer button. The reasons are to make sure the result that they will get is reliable and correct.

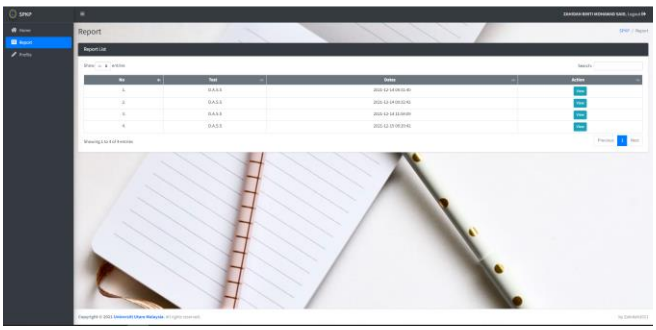


Fig. 10. The Interface for Psychological Test Report Result. (User).

Fig. 10 shows the interface of the test report result. The interface will list down all report results history. Not only that, but the user also can search for the specific report results by inserting particular report results keywords in the search box at the top right of the interface. The search box is at the top right so that users can easily see it.

Fig. 11 shows the main page of the system from the counselor's view. This interface consists of four columns which are the SSOSH test result in a blue column, the SSKM test result in a green column, the D.A.S.S test result in a red column, and user registration in a yellow column. Three columns are just the same as the user view side, only that there are no questions in it but instead there are report results of psychological tests of the users. The user registration column which is the yellow column, that column consists of a list of all registered user info such as their name and registration date. The interface design between the users and counselor view is not that different to keep the consistent looks and image of the system.

Fig. 12 and Fig. 13 show the interface for psychological test results from the counselor's view. The counselor can view all report results of all users by clicking the report button. The counselor also can search for the specific report results of the user by inserting the name or matric no/ staff no of the user in the search box at the top right of the interface. The counselor can also print out the report result by clicking the view button which is in turquoise color and then clicking the print button which is in blue color.

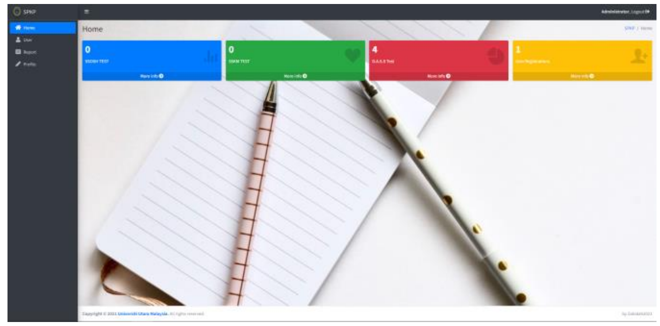


Fig. 11. The Mainpage of the System. (Counselor).

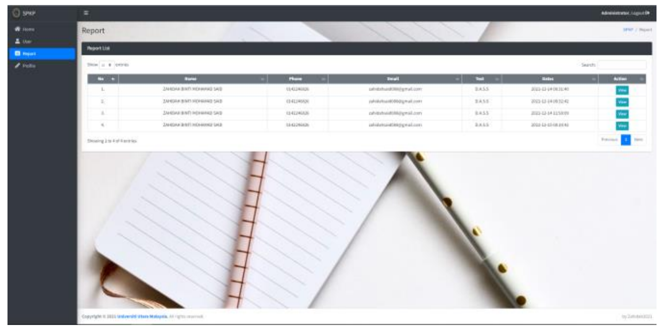


Fig. 12. The Interface for Psychological Test Report Result. (Counselor).

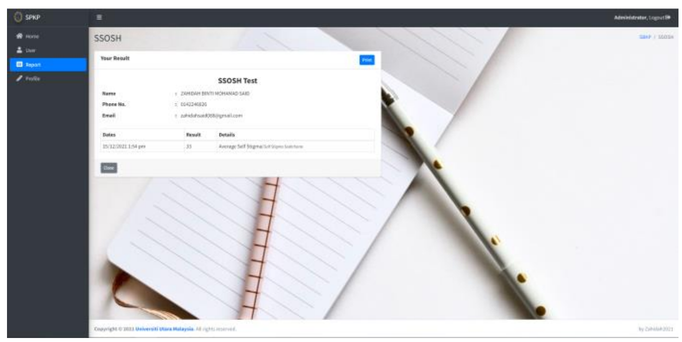


Fig. 13. The Interface for Psychological Test Report Results Print Section. (Counselor).

VI. Evaluation of Sistem Profil Kesejahteraan Psikologi

A. *The Evaluation Setting*

A usability evaluation was conducted on 30 respondents, consisting of students from Universiti Utara Malaysia. The respondents were approached randomly by sending out the Google form consisting of the questionnaire through WhatsApp. However, the respondents can choose to answer or ignore the questionnaire as there is no coercion. The instruments used for the evaluation were the SPKP web-based application and post-task questionnaire. The post-task questionnaire was adapted from [16] which consists of 33 items in four sections. Section A asked the respondents' demographic information while Section B, C, and D asked the respondent's opinions about SPKP web-based application on a five-point Likert scale where one represents strongly disagree, and five represents strongly agree. The respondents performed the following step-by-step procedure for the evaluation: (1) read and signed a consent form, (2) interacted with SPKP web-based application as stated in the experiment procedure, and (3) answered the post-task questionnaire.

B. *The Respondents' Demographic Information*

Analysis of the respondents' demographic information revealed that 70% of them were female and 30% of them were male. 86.7% of them were aged between 21 to 25 years old, 10% of them were 26 to 35 years old and 1% of them were 16 to 20 years old. It is also revealed that 83.3% of them once did the psychological tests while the remaining 16.7% have never done it before.

C. *The Usability of Sistem Profil Kesejahteraan Psikologi*

An analysis was conducted on the respondents' responses in Sections B, C, and D of the post-task questionnaire. The section measures the respondents' perception of SPKP's usefulness and ease of use. It also measured the respondents' satisfaction with SPKP. Tables II, III, and IV reported the frequency and average of the responses. The respondents rated four or five of the post-task scales for the three aspects of usability. None of the respondents rated one or two. Only a few rated neutral.

TABLE II. THE RESPONDENTS' RESPONSE ON THE USEFULNESS OF SYSTEM PROFILE KESEJAHTERAAN PSIKOLOGI

The post-task questionnaire items:	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Average
SISTEM PROFIL KESEJAHTERAAN PSIKOLOGI enhances my effectiveness on managing my mental health condition	0 (0%)	0 (0%)	0 (0%)	2 (6.7%)	28 (93.3%)	4.93
I can register an account in SISTEM PROFIL KESEJAHTERAAN PSIKOLOGI without any error	0 (0%)	0 (0%)	0 (0%)	3 (10%)	27 (90%)	4.90
I can login the system with registered email and password	0 (0%)	0 (0%)	1 (3.3%)	11 (36.7%)	18 (60%)	4.57
The search button can function well	0 (0%)	0 (0%)	0 (0%)	10 (33.3%)	20 (66.7%)	4.67
I can answer all psychological test without any problem	0 (0%)	0 (0%)	2 (6.7%)	7 (23.3%)	21 (70%)	4.63
The report result that I got after answering all psychological test question are as expected	0 (0%)	0 (0%)	1 (3.3%)	8 (26.7%)	21 (70%)	4.67
It saves my time when I use this system to manage my mental health condition	0 (0%)	0 (0%)	1 (3.3%)	8 (26.7%)	21 (70%)	4.67
SISTEM PROFIL KESEJAHTERAAN PSIKOLOGI meets my needs	0 (0%)	0 (0%)	1 (3.3%)	9 (30%)	20 (66.7%)	4.63
SISTEM PROFIL KESEJAHTERAAN PSIKOLOGI is useful in overall	0 (0%)	0 (0%)	0 (0%)	6 (20%)	24 (80%)	4.80

TABLE III. THE RESPONDENTS' RESPONSE ON THE EASE OF USE OF SYSTEM PROFILE KESEJAHTERAAN PSIKOLOGI

The post-task questionnaire items:	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Average
SISTEM PROFIL KESEJAHTERAAN PSIKOLOGI is easy to use	0 (0%)	0 (0%)	0 (0%)	5 (16.7%)	25 (83.3%)	4.83
SISTEM PROFIL KESEJAHTERAAN PSIKOLOGI is user friendly	0 (0%)	0 (0%)	0 (0%)	8 (26.7%)	22 (73.3%)	4.73
SISTEM PROFIL KESEJAHTERAAN PSIKOLOGI is flexible	0 (0%)	0 (0%)	0 (0%)	12 (40%)	18 (60%)	4.60
SISTEM PROFIL KESEJAHTERAAN PSIKOLOGI is easy to learn how to use it	0 (0%)	0 (0%)	1 (3.3%)	3 (10%)	26 (86.7%)	4.83
I can use SISTEM PROFIL KESEJAHTERAAN PSIKOLOGI without written instruction	0 (0%)	0 (0%)	0 (0%)	7 (23.3%)	23 (76.7%)	4.77
I can easily remember how to use SISTEM PROFIL KESEJAHTERAAN PSIKOLOGI	0 (0%)	0 (0%)	1 (3.3%)	2 (6.7%)	27 (90%)	4.87
I didn't notice any inconsistencies as I use SISTEM PROFIL KESEJAHTERAAN PSIKOLOGI	0 (0%)	0 (0%)	0 (0%)	10 (33.3%)	20 (66.7%)	4.67
My interaction with the system would be clear and understandable	0 (0%)	0 (0%)	0 (0%)	9 (30%)	21 (70%)	4.70
I can use SISTEM PROFIL KESEJAHTERAAN PSIKOLOGI successfully every time	0 (0%)	0 (0%)	0 (0%)	7 (23.3%)	23 (76.7%)	4.77

TABLE IV. THE RESPONDENTS' RESPONSE ON THEIR SATISFACTION OF SYSTEM PROFILE KESEJAHTERAAN PSIKOLOGI

The post-task questionnaire items:	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Average
I am satisfied with SISTEM PROFIL KESEJAHTERAAN PSIKOLOGI	0 (0%)	0 (0%)	1 (3.3%)	3 (10%)	26 (86.7%)	4.83
I would recommend SISTEM PROFIL KESEJAHTERAAN PSIKOLOGI to my friends	0 (0%)	0 (0%)	0 (0%)	7 (23.3%)	23 (76.7%)	4.77
SISTEM PROFIL KESEJAHTERAAN PSIKOLOGI works the way I want it to work	0 (0%)	0 (0%)	1 (3.3%)	8 (26.7%)	21 (70%)	4.67
I feel I need to have SISTEM PROFIL KESEJAHTERAAN PSIKOLOGI account in my PC/ Laptop	0 (0%)	0 (0%)	3 (10%)	8 (26.7%)	19 (63.3%)	4.53
SISTEM PROFIL KESEJAHTERAAN PSIKOLOGI is wonderful and pleasant to use	0 (0%)	0 (0%)	0 (0%)	7 (23.3%)	23 (76.7%)	4.77

The outcomes of the evaluation suggested that SPKP is useful and easy to use. Further, 86.7% of the respondents responded that they satisfied with the SPKP. 10% of them agree while 1% choose neutral. Analysis of the respondents' feedback about the specific features offered by SPKP shows that most of the respondents agree that the registration, login, search, and buttons were straightforward and useful with 80% of the respondents strongly agree with the statement that SPKP is useful in overall. Meanwhile, 20% of the respondents choose to agree with the statement. They also perceived that SPKP could help them in save their time in managing their mental health with 70% of the respondents strongly agree while 26.7% agree with the statement. However, there is 3.3% of the respondents unsure about the question. In terms of the user interface, the respondents reported that SPKP was easy to use without the need for written instruction and they can easily remember the way of using the web-based application system with 76.7% and 90% of the respondents respectively choose to strongly agree with the statements. Further, the respondents satisfied with the appearance of the web-based application and intended to recommend the web-based application to other.

VII. DISCUSSION

It has been agreed that the advantages of SPKP are widely acknowledged by UUM students. According to the result that we get from the questionnaire conducted among UUM students, it shows that majority of the students agree that SPKP is useful and easy to use compared to the current method where the students need to go physically to the UUM Counselling Centre just to get their mental health tested. This current method is not efficient and consumes a lot of money and time. As we know, the majority of the students do not have any income and some of them were busy with study life. Hence, most of them tend to ignore any mental health problems like depression and stress. By the time they realize they need help, it was too late. According to a series of researches, suicidal people do not always identify that they have an issue or symptoms that could benefit from expert help. According to a poll of 165 college students at high risk of suicide who had not sought professional help, 66% of these students believed that counseling was unnecessary [17]. Therefore, there is a need for a new method that will benefit them in terms of managing their mental health condition.

Overall, among 30 respondents 80% of them agree that SPKP is useful overall in managing their mental health conditions. This finding was conducted during the Covid-19 pandemic whereby the majority of the students and the staff were staying at home. Thus, conducting psychological tests physically at UUM Counselling Centre is impossible. By developing the system, they no longer need to worry about the Covid-19 pandemic whereby they can easily conduct their psychological tests tested anywhere and anytime they want. This system provides flexibility to them.

In order to attract user attention to using the system, the interface design of the system must need to be attractive, simple, and easy to use. A well-designed user interface can reduce training time, performance speed, error rates, user satisfaction, and the user's long-term retention of operations knowledge [18]. Based on the questionnaire conducted, 86% of the respondents stated that they feel satisfied with the system. However, there are some limitations to be considered while analyzing the study results. First, the sampling technique is only restricted to people with laptops or personal computers only as the system is a web-based application. Second, the research topic itself is not widely discussed and there are only a few related researches that can be used for this paper. Lastly, time is also one of the limitations faced while conducting this research.

Despite the limitations, this study provides the data of a web-based application system for managing mental health in higher institutions for future research. Questionnaires have been found as an effective way of assessing problems related to mental health and this becomes a method of conducting research during the Covid-19 pandemic. Results will make researchers be able to make decisions on what number of rating scale points to use for their survey and questionnaire [19]. Thus, the result of the questionnaire can be used for future research.

However, there is a need for improvements in certain areas such as the web-based application can be changed to a mobile application. Thus, there will be more people to receive the benefits of the system. It is because not every student is coming from a well-off family, so there are a few of them who cannot afford in buying a laptop or personal computer and only depends on a mobile phone. They couldn't afford a laptop, in part because they thought they'd used up all of their lending alternatives and, in part, because they didn't want to take out additional debt [20]. Hence, improving the web-based to the mobile applications can be a good idea.

VIII. Conclusion and Future Work

This paper described the design and development of a web-based application for conducting and managing psychological tests. In the end, the SPKP system is able to be developed, and hope that the system can really perform the work that it is supposed to do such as collecting, storing, retrieving, and calculating the result of the psychological tests for the UUM Counselling Centre. There are many aspects of psychological tests that can be studied. However, there are a few limitations while conducting this research where there is limited access to the data and there are also only a few previous research studies on the topic. Time is also one of the limitations we faced while conducting this research. In the future, we plan to expand the functionality of SPKP by making it available as a mobile application. Past studies suggested that psychological tests are a method conducted to know someone's mental health condition. However, some issues mentioned in the paper show that there is a need for improvement in a specific area so that the psychological tests can be conducted with ease. We also plan on making the web-based application responsive on multiple interfaces so that it will enhance user satisfaction when using the web-based application. It could also be extended into a recommender system which could suggest the counselor that is available at the moment in case the clients need further intervention without the need on making the appointment first.

Acknowledgment

The authors would like to thank to all School of Computing of Universiti Utara Malaysia members who involved in this study. This study was conducted for the purpose of Undergraduate Final Year Project. This work was supported by Universiti Utara Malaysia.

References

[1] M. F. Eabon and D. Abrahamson, "Understanding psychological testing and assessment",. Available: https://www.apa.org/topics/psychological-testing-assessment [Accessed 10 January 2022].

[2] Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. Proceedings of the National Academy of Sciences, 112(4), 1036-1040.

[3] Allen, A. P., Kennedy, P. J., Cryan, J. F., Dinan, T. G., & Clarke, G. (2014). Biological and psychological markers of stress in humans: Focus on the Trier Social Stress Test. Neuroscience & Biobehavioral Reviews, 38, 94-124.

[4] Salaheddin, K., & Mason, B. (2016). Identifying barriers to mental health help-seeking among young adults in the UK: a cross-sectional survey. British Journal of General Practice, 66(651), e686-e692.

[5] Goñi-Sarriés, A., Blanco, M., Azcárate-Jiménez, L., Peinado, R., & López-Goñi, J. J. (2018). Are previous suicide attempts a risk factor for completed suicide?. Psicothema 2018, Vol. 30, No. 1, 33-38.

[6] Hussey, I., & Hughes, S. (2020). Hidden invalidity among 15 commonly used measures in social and personality psychology. Advances in Methods and Practices in Psychological Science, 3(2), 166-184.

[7] Cohen, B. E., Edmondson, D., & Kronish, I. M. (2015). State of the art review: depression, stress, anxiety, and cardiovascular disease. *American journal of hypertension*, *28*(11), 1295-1302.

[8] Topkaya, N. (2015). Factors Influencing Psychological Help Seeking in Adults: A Qualitative Study. *Educational Sciences: Theory and Practice*, *15*(1), 21-31.

[9] Saravia, E., Chang, C. H., De Lorenzo, R. J., & Chen, Y. S. (2016, August). MIDAS: Mental illness detection and analysis via social media. In 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (pp. 1418-1421). IEEE.

[10] Koutamanis, M., Vossen, H. G., & Valkenburg, P. M. (2015). Adolescents’ comments in social media: Why do adolescents receive negative feedback and who is most at risk?. Computers in Human Behavior, 53, 486-494.

[11] Andalibi, N., Haimson, O. L., De Choudhury, M., & Forte, A. (2016, May). Understanding social media disclosures of sexual abuse through the lenses of support seeking and anonymity. In Proceedings of the 2016 CHI conference on human factors in computing systems (pp. 3906-3918).

[12] Shenoy, P., & Aithal, P. S. (2016). A study on history of paper and possible paper free world. *International Journal of Management, IT and Engineering*, *6*(1), 337-355.

[13] Koch, M., Möller, C., & Spinath, F. M. (2021). Are you Swiping, or Just Marking? Exploring the Feasibility of Psychological Testing on Mobile Devices. *Psychological Test and Assessment Modeling*, *63*(4), 507-524.

[14] Mohr, D. C., Weingardt, K. R., Reddy, M., & Schueller, S. M. (2017). Three problems with current digital mental health research... and three things we can do about them. Psychiatric services, 68(5), 427-429.

[15] Alshamrani, A., & Bahattab, A. (2015). A comparison between three SDLC models waterfall model, spiral model, and Incremental/Iterative model. *International Journal of Computer Science Issues (IJCSI)*, *12*(1), 106.

[16] A. M. Lund, "Measuring usability with the use questionnaire," *Usability interface*, vol. 8, pp. 3-6, 2001.

[17] Hom, M. A., Stanley, I. H., & Joiner Jr, T. E. (2015). Evaluating factors and interventions that influence help-seeking and mental health service utilization among suicidal individuals: A review of the literature. *Clinical psychology review*, *40*, 28-39.

[18] Ishak, Z., Alexander, O., Al-Sanjary, O. I., & Yusuf, E. (2020, February). Potential Students Preferences Towards University Website Interface Design: The Methodology. In *2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA)* (pp. 115-119). IEEE.

[19] Taherdoost, H. (2019). What is the best response scale for survey and questionnaire design; review of different lengths of rating scale/attitude scale/Likert scale. *Hamed Taherdoost*, 1-10.

[20] Reisdorf, B. C., Triwibowo, W., & Yankelevich, A. (2020). Laptop or bust: How lack of technology affects student achievement. *American Behavioral Scientist*, *64*(7), 927-949.

Breast Cancer Classification using Decision Tree Algorithms

Omar Tarawneh[1], Mohammed Otair[2], Moath Husni[3]
Hayfa.Y. Abuaddous[4], Monther Tarawneh[5], Malek A Almomani[6]
Software Engineering Department, Amman Arab University, Amman, Jordan[1, 4]
Computer Science Department, Amman Arab University, Amman, Jordan[2]
Software Engineering Department, The World Islamic Sciences and Education University, Amman, Jordan[3, 6]
Computer Science Department, Isra University, Amman, Jordan[5]

Abstract*—Cancer is a major health issue that affects individuals all over the world. This disease has claimed the lives of many people, and will continue to do so in the future. Breast cancer has recently surpassed cervical cancer as the most frequent cancer among women in both industrialized and developing countries and it is now the second leading cause of cancer mortality among women. A high number of women die each year as a result of this disease. Breast cancer is significantly easier to treat if caught early. This paper introduces a decision tree-based data mining technique for breast cancer early detection with highest accuracy, which helps patients to recover. Breast cancers are classed as benign (unable to penetrate surrounding tissue) or malignant (able to infiltrate adjacent tissue) breast growths. Two tests were included in the review. The primary study uses 10 breast cancer samples from the Kaggle archive, whereas the follow-up study uses 286 breast cancer samples from the same pool. The Decision Tree's accuracy in the first trial was 100%, while it was 97.9% in the follow-up inquiry. These findings justify the use of the proposed machine learning-based Decision Tree classifier in pre-evaluating patients for triage and decision-making prior to the availability of data.

***Keywords*—Data mining; decision tree; classifier; breast cancer**

I. Introduction

A huge number of cells make up the human body, each with its own unmistakable reason. Cancer is characterized as the uncontrolled proliferation of any of these cells. Cells partition and create uncontrolled, bringing about a cancer, which is an abnormal mass of tissue. Growth cells increase and infiltrate the stomach related, neurological, and circulatory frameworks, interfering with normal body capacities. Although not all growths are malignant [1].

Cancer is classed based on the sort of cell that is afflicted, and there are more than 200 distinct varieties of cancer. Breast cancer is the subject of this review. Breast cancer is the most continuous type of cancer in ladies and it is considered as a major cause of death from all cancers, especially among ladies [2]. There are no established methods for avoiding breast cancer at this time. Figuring out how to recognize breast cancer at an early stage, then again, will assist with peopling who are affected. Breast cancer fixes chances can be enhanced assuming that the disease is diagnosed early. Provided that breast cancer is diagnosed early can it be avoided and safeguarded. Among the several evaluating strategies for classifying breast cancer, digital mammography is the most generally used [3].

Breast cancer survival rates have worked on considerably lately, regarding to the enhanced screening and treatment decisions. As a result of the advancements in data assortment and storage advances, medical organizations and hospitals may now store massive amounts of data associated with their medical records, including meds and disease indications. Data mining is the method involved with obtaining helpful information from large amounts of data utilizing complex algorithms. Medical data have broadened the applicability and potentials of these strategies [4]. Anticipating the result of a disease is a challenging task. Data mining strategies are regularly used to work on the expectation fragment. Large volumes of medical data may now be gathered and made available to medical research bunches using automated approaches. Thus, data mining apparatuses for finding patterns and correlations among countless variables are turning out to be increasingly normal, allowing for the forecast of ailment results based on recently gathered data [5]. However, most of the previous studies suffer from lack of accuracy in diagnose of this disease.

To fill in this gap, recognize and classify breast cancer at an early stage with high accuracy is very important in order to help patients to recover from this disease. Thus, this paper classified the breast cancer dataset by utilizing Decision Tree classification technique in WEKA. The breast cancer dataset is isolated and classified into categories based on features, performance, and different factors. Fig. 1 shows the entire interaction stream of the suggested paradigm.

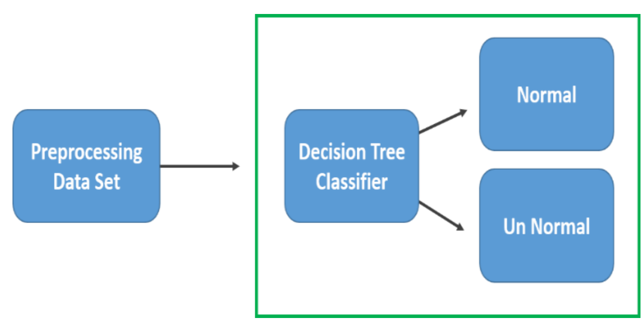


Fig. 1. Block Diagram of the Proposed Approach.

The classifier receives profound characteristics from the completely linked layer. The decision tree classifier makes use of these characteristics. After that, the characterization is complete, and all arrangement properties are defined. The suggested approach might be used to assess images of breast cancer.

The rest of this paper is organized as follows: Section 2 presents the mostly related works. Section 3 explains how decision tree can be used in the proposed paradigm. Section 4 discusses the experimental results. Finally, we conclude the paper in Section 5.

II. Related Work

Poomani and Porkodi conducted a study that aims to get the best classifier, by comparing different learning techniques [6]. The findings of their investigation reveal that the most elevated accuracy is found in the J48 classifier, which returns 0.979 with the least mistake rate of 0.9587 of all the classification methods. Their findings demonstrate that the combination of MLP and J48 classifiers with features determination Prognosis Breast Cancer (PCA) is better than different classifiers. In the Wisconsin Prognosis Breast Cancer WPBC dataset, their discoveries showed that the combination of IBK, J48, SMO, and MLP performs effectively. Saabithet et al. Compare numerous breast cancer data sets in, their research [7]. The extent of exactness with and without features selection techniques for breast cancer data was estimated utilizing classification algorithms like J48, MLP, and Rough. They conclude that the characteristic decision technique is the most reliable signifier process for enhancing the exactness of various categorization techniques, achieving the least Mean Standard Error (MSE) and the most elevated Recipient Operating Characteristics (ROC) to recognize breast cancer disease.

Peter et al. in [8] used two data mining approaches to quantify breast cancer gambles in Nigerian patients using the naïve Bayes algorithm and the J48 decision trees algorithms in their review exertion. The effectiveness of both categorization algorithms was evaluated to figure out which model was the most capable and beneficial. S. Syed Shajahaan, et al. [9] Compare DM approaches to model breast cancer data in a review. They took a gander at how to utilize decision trees to foresee the presence of breast cancer, as well as how to quantify the performance of conservative administered learning algorithms utilizing CART, C4.5, ID3, and Naive Byes. The ID3 random tree was demonstrated to be impressive with the most elevated accuracy in the investigations.

This study [10] compares the performance of the J48, AD Tree, BF Tree, and regression trees (CART) algorithms based on classification accuracy, using several accuracy measurements, for example, FP rate, TP rate, Recall, Precision, ROC Area, and F-measure. Decision trees are common and easy-to-understand structures from which rules may be obtained. Just the numerical values of explicit attributes in the breast cancer data are examined all through the implementation phase. The J48 classifier has the highest accuracy of close to 100%, while the CART algorithm has a 96 percent accuracy, the AD Tree algorithm has a 97 percent accuracy, and the BF Tree algorithm has a 98 percent accuracy. According to the classification findings of the four algorithms, it is obviously shown that J48 beats the other three techniques for the predetermined data set.

In [13], they proposed utilizing gene expression data to classify triple negative and non-triple negative breast cancer patients using a machine learning (ML) technique. The Support Vector Machine approach, out of the four ML algorithms tested, was able to classify breast cancer more accurately into triple negative and non-triple negative breast cancer and had less misclassification errors than the other three.

Epimack et. al [14] presented a computer-aided diagnostic (CAD) system that generates an optimum algorithm automatically which use 13 of the 185 features available to teach machine learning. To distinguish between malignant and benign tumors, researchers deployed five machine learning classifiers. For 10-fold cross-validation, the experimental findings indicated Bayesian optimization with a tree-structured Parzen estimator based on a machine learning classifier.

III. Decision Tree Learning as a Proposed Technique

Breast cancer classification was considered in many recent researches [13][14] using most of machine and deep learning techniques, because such classification is very important to be accurate as much as possible since it is related with most of women lives.

A decision tree is described as a classifier using a recursive split of the instance space. It generates a predictive model that connects node observations to inferences about the desired value of the nodes. The leaf in a tree structure represents the class.

Labels and branches represent the attributes that lead to the class labels. Fig. 2 is an illustrative example of a binary decision tree [11].

The proposed technique comprises constructing a prediction model for detecting whether a tumor is benign or malignant based on various characteristics associated with a given medical record using a decision tree data mining methodology. A decision tree is an excellent tool for classification and prediction in the case of Breast Cancer diagnosis. A number of decision tree approaches are available to categorize the data. Following data pre-processing (in CSV format), WEKA (Java Toolkit for Different Data Mining Techniques) is used to apply the Decision tree algorithm to the dataset, and the data is classed as benign or malignant based on the decision tree's final conclusion. Fig. 3 depicts the flow of the research that was utilized to create the model.

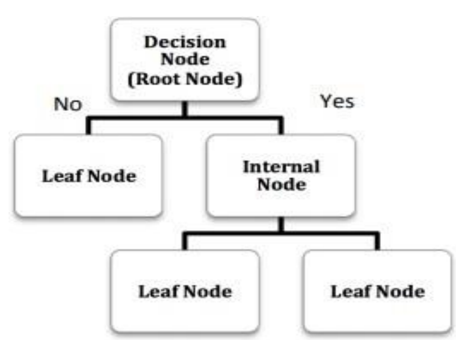


Fig. 2. Illustrated Example of Binary Decision Tree.

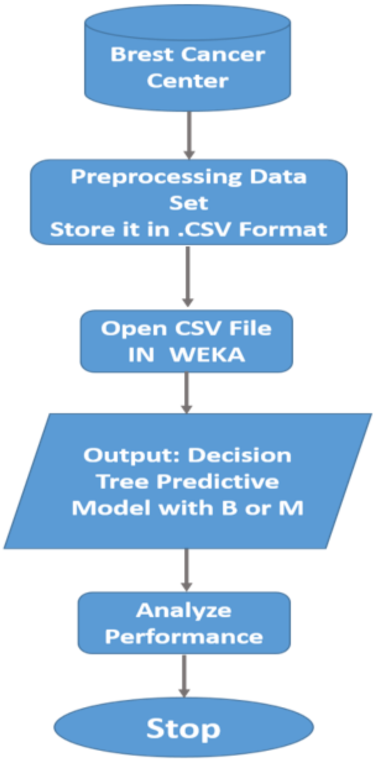


Fig. 3. Example of Binary Decision Tree.

The algorithm is carried out according to the following steps:

1) Collect datasets on breast cancer.

2) Pre-processing data in for performing the Decision Tree data mining approach (Benign, Malignant).

3) Data that has been pre-processed is submitted to the WEKA toolbox for analysis.

4) Implement the Decision Tree algorithm, which generates a decision tree with leaf nodes as class labels (benign and malignant).

5) New patient diagnoses are made by cross-referencing new attribute values in the decision tree and continuing the route until the leaf node is reached, which indicates whether the tumor is benign or malignant.

IV. Experimental Studies

A. Dataset

The CSV (Comma Separated Value) format was utilized in this study. Breast cancer data is supplied in CSV format via sex, age, menopause, tumor- size, inv-nodes, node-caps, deg-malig, breast, and breast-quad, irradiate, Class. These records were prepared in an Excel spreadsheet and saved as a CSV file, then uploaded to WEKA which accepted this format. In breast cancer data two types of properties, benign and malignant. In this study, these data are used and analyzed.

B. Experimental Results

The major goal of this paper is to evaluate the effectiveness of the breast cancer classification algorithms depending on a variety of input factors. They are analyzed using a Decision Tree algorithm. For the performance evaluation, the WEKA application is employed. Two setoff experiments are used to test each classifier. The first experiment has 10 samples. While the second one has 286 samples. This involves categorizing pre-processing based on all of the values of the attributes that have been taken. This paper compares the accuracy of categorization in the Decision Tree algorithm. The analysis of precision, FT Rate, and TP Rate is also performed. The following are the numerous formulas used to calculate various metrics. Precision P is calculated using the Formula (1) where TP = True Positive Rate, which is the proportion of expected positive instances, and FP = False Positive Rate, which is the part of the anticipated false positive cases:

$$Precision\ P = TP/\ (TP+FP) \quad (1)$$

The proportion of positive instances that were accurately detected has been described as True Positive Rate (TPR), Sensitivity, or Recall. Formula (2) will be used to calculate it:

$$\text{Recall} = TP/\ (TP+FN).\ \text{Where FN=False Negative Rate} \quad (2)$$

$$Accuracy = (TP+TN)\ /(TP+TN+FP+FN) \quad (3)$$

The accuracy in (3) will be computed as the fraction of the total number of correct predictions, where TN = True Negative. The calculation of an average of the recall and precision measures for retrieving information, which called as F-Measure as shown in Formula (4).

$$F= (2^*\ \text{Recall}^*\ Precision)\ /Precision+\text{Recall}) \quad (4)$$

1) First experimental results: The first experiment has 10 samples of breast cancer [12]. Tables I to VI show the results of algorithms and compared with them.

Table I shows the summary of accuracy for proposed technique. Where the number of samples is 10.

Table II illustrates all the measures that used to evaluate the proposed method in details. We can see the proposed technique.

TABLE I. Summary Accuracy of Decision Tree Algorithm

Total Number of Instances	**10**	
Correctly Classified Instances	10	100%
Incorrectly Classified Instances	0	0%
Kappa statistic	1	
Mean absolute error	0.1429	
Root mean squared error	0.1429	
Relative absolute error	28.5714 %	
Root relative squared error	28.5714 %	

TABLE II. METRICS RESULTS ACHIEVED BY DECISION TREE ALGORITHM

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1
Weighted Avg.	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	

TABLE III. CONFUSION MATRIX OF DECISION TREE ALGORITHM

a	b	classified as
5	0	a = 0
0	5	b = 1

Table III shows the confusion matrix.

Table IV shows the summary of accuracy in a Random Forest algorithm, where the number of samples is 10.

Table V shows the values of the using measures to evaluate Random Forest algorithm classification.

TABLE IV. SUMMARY ACCURACY OF RANDOM FOREST ALGORITHM

Total Number of Instances	10	
Correctly Classified Instances	5	50 %
Incorrectly Classified Instances	5	50 %
Kappa statistic	0	
Mean absolute error	0.5	
Root mean squared error	0.5	
Relative absolute error	100 %	
Root relative squared error	100 %	

TABLE V. METRICS RESULTS ACHIEVED BY RANDOM FOREST ALGORITHM

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	1.000	0.500	1.000	0.667	0.352	0.500	0.500	0
	0.000	0.000	0.000	0.000	0.200	0.352	0.500	0.500	1
Weighted Avg.	0.500	0.500	0.250	0.500	0.516	0.352	0.500	0.500	

TABLE VI. CONFUSION MATRIX OF RANDOM FOREST ALGORITHM

a	b	classified as
5	0	a = 0
5	0	b = 1

The results that attained from Tables I to VI illustrate the loss and accuracy throughout training and validation. Based on the results, it is clear that the maximum training and validation accuracy is seen in the Decision Tree architecture, with a loss rate of 0%. The Random Forest, on the other hand, achieves the lowest training and validation accuracy of 50% and a loss of 50% at iterations 100. Fig. 4 and Fig. 5 show that the loss values of decision tree equal to zero.

2) Second experimental results: When used large number of samples [12], the accuracy will be change as the Tables VII, VIII and IX. The classification of the attributes (age, menopause, tumor-size, inv-nodes, node- caps, deg-malig, breast, breast-quad, irradiat, Class) can be shown in the Fig. 6.

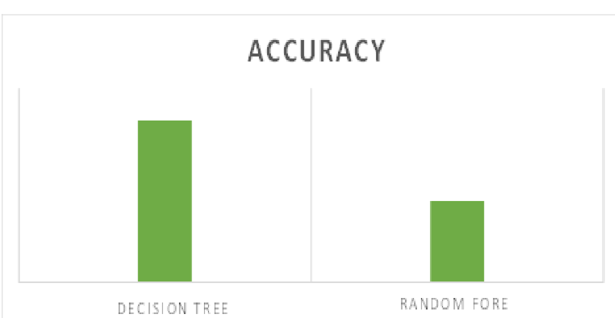


Fig. 4. Accuracy Comparison of Algorithms.

Fig. 5. Loose Comparison of Algorithms.

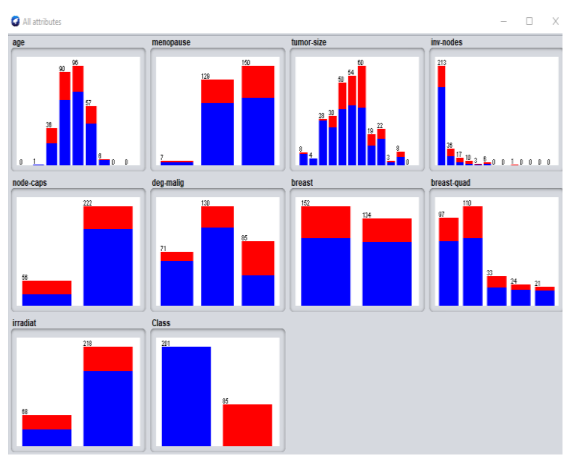


Fig. 6. Attribute Classification.

Table VII shows the accuracy summary in the Decision Tree algorithm, where the number of samples is 286.

Table VIII shows all measures values that were used to evaluate classification in studied methods.

TABLE VII. SUMMARY ACCURACY OF DECISION TREE ALGORITHM

Total Number of Instances	**286**	
Correctly Classified Instances	280	97.9021%
Incorrectly Classified Instances	6	2.0979%
Kappa statistic	0.9491	
Mean absolute error	0.0221	
Root mean squared error	0.1052	
Relative absolute error	5.2937 %	
Root relative squared error	23.0237 %	

TABLE VIII. DETAILED ACCURACY BY CLASS OF DECISION TREE ALGORITHM

	TP Rate	**FP Rate**	**Precision**	**Recall**	**F-Measure**	**MCC**	**ROC Area**	**PRC Area**	**Class**
	0.995	0.059	0.976	0.995	0.985	0.950	0.999	0.999	0
	0.941	0.005	0.988	0.941	0.964	0.950	0.999	0.996	1
Weighted Avg.	0.979	0.043	0.979	0.979	0.979	0.950	0.999	0.998	

TABLE IX. CONFUSION MATRIX OF DECISION TREE ALGORITHM

a	**b**	**classified as**
200	1	a = 0
5	80	b = 1

The results obtained from the Tables VII to IX demonstrate the loss and accuracy during the training and validation stages. The maximum training and validation accuracy are observed for Decision Tree architecture 97.9021% and loss is 2.0979%. These results show that the proposed technique is active in all cases. Fig. 6 illustrates the classification depending on attributes.

V. CONCLUSION

This study demonstrates the use of decision trees to represent actual breast cancer diagnosis for local and systemic treatment, as well as additional strategies that may be employed. The study assesses the performance of the Decision Tree algorithm in terms of classification accuracy, using several accuracy metrics such as the F-measure, ROC Area, Precision, Recall, TP rate, and FP rate. Decision trees are well-known and simple to comprehend structures from that rule may be derived. The studied model's efficacy is demonstrated by experimental findings. For the detection of breast cancer, the effectiveness of the decision tree approach was evaluated and explored. Throughout the implementation phase, only the numerical values of particular breast cancer features are assessed. The experimental findings reveal that the Decision Tree classifier has a 100% accuracy rate, while the Random Forest method only has a 50% accuracy rate. The performance of Decision Tree is superior than the other method for the specified dataset on the basis of the four algorithms' categorization results.

REFERENCES

[1] Syed SS, Shanthi S, Chitra VM. Application of Data Mining techniques to model breast cancer data. International Journal of Emerging Technology and Advanced Engineering. 2018 Nov; 3(11):362–9.

[2] Shiv Shakti S, Sant A, Aharwal RP. An Overview on Data Mining Approach on Breast Cancer data. International Journal of Advanced Computer Research. 2019; 3(13):256–62.

[3] Shweta K. Using data mining techniques for diagnosis and prognosis of cancer disease. International Journal of Computer Science, Engineering and Information Technology. 2019 Apr; 2(2):55–66.

[4] Takiar R, Nadayil D, Nandakumar A. Projections of number of cancer cases in India (2010-2020) by cancer groups. Asian Pac J Cancer Prev. 2017; 11(4):1045– 9.

[5] Vaidehi K, Subashini TS. Breast tissue characterization using combined K-NN classifier. Indian Journal of Science and Technology. 2019 Jan;8(1):23–6.

[6] Salama GI, Abdelhalim MB, Abd-elghany Zei Md. Breast cancer diagnosis on three different datasets using multi-classifiers. International Journal of Computer and Information Technology. 2020 Sep;1(1):36–43.

[7] Gupta S, Kumar D, Sharma A. Data mining classification techniques applied for breast cancer diagnosis and prognosis. Indian Journal of Computer Science and Engineering. 2020 Apr-May; (2):188–95.

[8] [Williams, K., Idowu, P. A., Balogun, J. A., & Oluwaranti, A. I. (2015). Breast Cancer Risk Prediction Using Data Mining Classification Techniques. Transactions on Networks and Communications, 3(2), 01. https://doi.org/10.14738/tnc.32.662.

[9] S. Syed Shajahaan, S. Shanthi, & V. ManoChitra (2013), Application of Data Mining Techniques to Model Breast Cancer Data, International Journal of Emerging Technology and Advanced Engineering 3(11): 362-369.

[10] Rajesh K, Anand S. Analysis of SEER Dataset for Breast Cancer Diagnosis using C4.5 Classification Algorithm. International Journal of Advanced Research in Computer and Communication Engineering. 2019 Apr; 1(2):72–7.

[11] Saravana Kumar K, Arthanariee AM. Evaluate the multiple breast cancer factors and calculate the risk by software tool breast cancer risk evaluator. Indian Journal of Science and Technology. 2018 Apr;8(S7):686–91.

[12] “breast cancer dataset” accessed on: Jan. 10, 2022[online]. Available: https://www.kaggle.com/search?q=breast+cancer+dataset.

[13] Wu J, Hicks C. Breast Cancer Type Classification Using Machine Learning. J Pers Med. 2021; 11(2):61.

[14] Epimack Michael, He Ma, Hong Li, Shouliang Qi, "An Optimized Framework for Breast Cancer Classification Using Machine Learning", BioMed Research International, vol. 2022, Article ID 8482022, 2022.

Development and Validation of e-Books during the Post-Pandemic to Improve Attitude towards Environmental Care in Case of Indonesia

Anggraeni Mashinta Sulistyani[1], Zuhdan Kun Prasetyo[2], Farida Hanum[3], Rizki Noor Prasetyono[4]
Department of Science Education, Yogyakarta State University, Yogyakarta, Indonesia[1, 2]
Department of Philosophy and Sociology of Education, Yogyakarta State University, Yogyakarta, Indonesia[3]
Faculty of Science and Technology, University of Peradaban, Brebes, Jawa Tengah, Indonesia[4]

***Abstract*—Indonesia has natural diversity in the form of tropical rain forests, the sea, flora, and fauna. Information technology adaptation is needed considering its development and use is getting higher, especially during the pandemic and post-pandemic periods. Through information technology in the form of E-books based on local wisdom, an attitude of caring for the environment grows. This research aims to combine information technology, environmental-based science learning, and local wisdom in growing students' environmental care character. Through the use of e-books based on Google's website as information technology in implementing a green science learning model that is oriented to local wisdom. This study develops and validates an e-book based on local wisdom through the ADDIE research design. The results of the analysis and discussion are that the development of an e-book based on Google's website with a green science model oriented to local wisdom is declared valid and feasible to use with an average percentage of 90.74% from the three experts. Then the results of student responses also showed a positive response from students with a percentage of 90.29%. The application of google site-based e-books in the green science model oriented to local wisdom in learning can increase the growth of environmental care characters for junior high school students as evidenced by the results of multiple regression analysis of significance 0.00 below 0.05.**

***Keywords*—e-Book; Google site; green science learning model; local wisdom; environmental care character**

I. Introduction

In the new normal era after being released from the pandemic, Indonesia began to rise and begin to be able to adapt to the pandemic environment. Likewise, education was also affected by the pandemic which eventually adopted technology and information as a solution in the learning process. Adaptation in digital learning is beneficial in pandemic learning because it can reduce the spread of the coronavirus [1]. Layali et al in their research argue that E-learning can be positively accepted in learning during a pandemic [2]. In addition to the pandemic, Indonesia has experienced a decline in culture and care for the environment because people cannot leave their homes so cultural activities do not run effectively. Research conducted by Yanuarita & Haryati [3] took samples in socio-cultural conditions in Indonesia which showed a negative impact during this pandemic. Meanwhile, Indonesia is a country that is rich in cultural diversity. This culture has been passed down from generation to generation by the ancestors until now. If these problems are not addressed, Indonesia can lose its original culture and education can also have an impact.

Local culture or local wisdom is also essentially the foundation of the philosophy of education in Indonesia as a filter in an outside culture that enters through electronic media. This local wisdom can also shape the character of discipline, honesty, cooperation, responsibility, environmental care, and tolerance [4]. This local wisdom can be developed and integrated into the educational curriculum [5]. Even in the research of Ratminingsih et al [6], learning media integrated with the local culture can increase students' basic competence in learning. In social life, local wisdom is ingrained and becomes a guide in determining a decision. The local wisdom can be expressed in the form of learning models or learning tools. Local wisdom is associated with scientific and rational real life. Local wisdom also has a relationship with the environment or nature so that local wisdom in learning is more contextual.

The green science learning model is a learning model that contains the characteristics of the learning model according to Arends [7]. The characteristics of the model include the underlying theory, learning objectives, syntax, and classroom management. The green science learning-oriented model is a learning model that is based on the environment. Environmental-based learning was first initiated by Jan Ligthart in 1859. The next figure who moved environment-based learning was J.J Rousseau. The learning idea is to bring students to real situations within the main learning environment. Piaget [8] supports this by stating that children's thinking in the formal operational period is no longer centered on what is seen, but can think of a hypothesis. Besides being able to have an impact on student character, the green science learning model also requires learning media and, books/student worksheets as an aid in learning.

Previous research on educational technology in general regarding the use of technology in learning has not discussed integration into aspects of local wisdom. For example in research Sindi et al [9] stated that today's education, especially in pandemic conditions, applies technology and information in the classroom because it includes kinesthetic, visual, and auditory aspects. Learning media in education in the digitalization era is now applying technology and information as mandatory. Constraints in digital learning are concentration in learning, interaction in activities, and decreased motivation, but information technology-based learning can still find a place if it is packaged interactively and innovatively to generate enthusiasm for learning [10]. It is indeed needed in the development of learning media in technology because in the digital era there are now many programs that are used. According to Xin & Singh [11], the transformation in educational technology is now available in various platforms that can be used, such as e-learning, e-books, learning management systems, and evaluation systems learning. e-Books are part of the learning media in the form of interactive electronic-based textbooks for access [12], [13], [14].

The difference and the novelty of the research carried out is the development of e-books made into google sites for easy access, e-book content containing local wisdom-based science learning models in growing environmental care characters. The choice of developing an e-book makes it easier for students to access and use it because it can be used with electronic media [15]. The use of the google site is easier in design, development, and implementation, especially free and unpaid aid, which can be used by all groups. The cultural element is local wisdom in green science learning as character building. So to combine the components of technology, teaching, and culture, it is developed, integrated, and implemented to grow the character of caring for the environment.

II. Method

A. The Research Model

The research was conducted using a research and development (R&D) research model. The development procedure used refers to the ADDIE development model. Branch [16] describes the ADDIE model development procedure including analysis, design, development, implementation, and evaluation steps. The procedure for developing the ADDIE model in this study can be explained through the flow chart in Fig. 1 below:

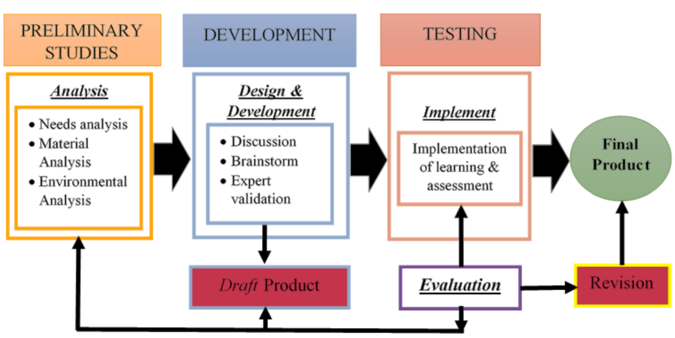


Fig. 1. ADDIE Model Development Flow.

Following Fig. 1 shows the stages of development to implementation. The analysis phase contains an analysis of needs in learning, especially during a pandemic and online learning. Regarding the need for technology and information in learning, these componentexamininginінto learning materials that can be linked to culture and environmentnmin learningngning natural sciences. The Design and Develop stage is a follow-up to the analysis stage, which requires learning media that combines technology, local culture, and green science learning model learning in growing the character of caring for the environment. The selection of e-books through the Google site as a development product with the contents of the green science learning model based on local wisdom foster a caring character for the environment. The developed learning media products are then tested for feasibility and discussed with the Experts in the Group Discussion Forum.

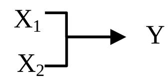


Fig. 2. Product Implementation Paradigm.

Product development results are implemented in learning to be tested using quantitative analysis. According to Fig. 2, X_1 is the independent variable. The first independent variable is the development of e-books based on google sites, X_2 is also the independent variable is a green science learning model based on local wisdom, and Y is the dependent variable is the character of environmental care. This is followed by the evaluation stage of the implementation of the e-book based on the Google site as a development product with the contents of the green science learning model based on local wisdom to cultivate the character of caring for the environment, questionnaires, and student responses.

B. Research Sample and Instrument

To implement and evaluate product development, a sample of 32 students was taken at Bumiayu Junior High School, Brebes Regency, Indonesia to collect response data.

Evaluation of the application of student worksheets was taken using a questionnaire and student responses. The student response instrument using student worksheets also uses a response questionnaire. Both instruments use a Likert scale of 1-4 with a description of a score of 1 is not good, a score of 2 being quite good, a score of 3 being good, and a score of 4 being very good.

C. Data Analysis

The feasibility test data is used to determine the feasibility of the product that has been developed through expert validation, which consists of 3 experts including material experts, technology media, and models. Furthermore, implementation is carried out to students to determine student responses using student worksheets. Taking expert validation sheets and student response questionnaires through qualitative data which is converted into quantitative data with a Likert scale of 1-4. Furthermore, the data is analyzed and considered feasible if it is declared valid using descriptive analysis. Based on the following equation (1):

$$P = \frac{\sum x}{\sum x^i} \times 100\% \tag{1}$$

Info :

P : percentage of expert/practitioner validity.

$\sum x$: total number of respondents' answers.

$\sum xi$: the total number of ideal scores in one item.

The data is then converted based on the validity conversion which can be seen in Table I.

TABLE I. VALIDITY CONVERSION GUIDELINES

Score Range (%)	**Category**
90 – 100	Very valid
75 – 89	Valid
65 – 74	Quite valid
55 – 64	Less valid
0 – 54	Invalid

The data from the implementation of the use of e-books based on google sites in the local wisdom-oriented green science learning model were analyzed through multiple regression tests. Multiple regression test aims to determine the effect of the use of e-books through Google sites in the local wisdom-based green science learning model on the character of environmental care. By using the IBM SPSS 25 application in multiple regression tests with the following hypothesis:

H_0 The variable of e-books based on google sites in the local wisdom-oriented green science learning model affects the character of caring for the environment.

H_a Variable e-books based on google sites in the green science learning model oriented to local wisdom do not affect the character of caring for the environment.

The multiple linear regression equation used for data analysis on Eq. (2).

$$Y = a + b_1X_1 + b_2X_2 \quad (2)$$

According to equation (2), the values of X_1 and X_2 are independent variables of the study with Y as the dependent variable. Where a is a constant coefficient, b_1 is a variable coefficient of X_1 and b_2 is a variable coefficient of X_2 which is obtained from the results of data analysis.

The result of the multiple regression test is interpreted that the significance value of the SPSS calculation results is less than the degree of freedom, which is 5% (Sig. < 0.05), then the hypothesis H_0 is accepted. This means that there is an influence on the use of e-books based on Google sites in the green science learning model oriented to local wisdom on the character of caring for the environment. Conversely, if the significance value is more than 5% (Sig. > 0.05), then there is no effect, meaning that H_0 is rejected.

III. RESULT AND DISCUSSION

The findings and discussions were carried out by developing an e-book of local wisdom oriented towards green science learning through the ADDIE design. The development of this e-book product was validated by several experts to test its feasibility. Then it was tested and evaluated on improving the attitude of caring for the environment in junior high school students. the ADDIE development process is produced as follows:

A. Analysis Stage

The analysis stage is carried out by analyzing needs, materials, and the environment. At this stage, as complete information as possible is extracted to obtain everything needed to proceed to the next stage. The analysis phase is carried out by direct observation of students, teachers, curriculum, and the school environment. The needs, material, and environmental analysis stages are carried out by analyzing essential learning materials and learning media so that they can be used in learning that relates to these myths and fosters an attitude of caring for the environment. The results of observations for the analysis of needs, materials, and the environment in the development of learning media, the following is a description of the analysis:

TABLE II. DESCRIPTION OF THE ANALYSIS STAGE

No.	**Aspects analyzed**	**Analysis Results**
1.	Need analysis	Innovative natural science learning based on technology without losing cultural and environmental values. By using an e-book-based Student Worksheet in which students interact in learning activities.
2.	Material Analysis	Natural science learning has themes and materials that can be related to cultures such as mythical culture and the environment. All of Indonesia has a wide variety of cultures and nature because it is an archipelagic country.
3.	Environmental Analysis	The education sector can contribute to environmental sustainability. Students are taught indirectly through learning in nature.

Table II shows the need for learning that combines learning about natural science, culture, and technology to develop e-books based on Google sites in a green science model oriented to local wisdom to foster an attitude of caring for the environment.

B. Design and Development Stage

The next stage is product development, namely from product drafts according to Fig. 3. The use of google sites is easy to use in the process of making e-books, without using complicated web coding, you can create good website features.

Fig. 3. Product Draft Start Page View.

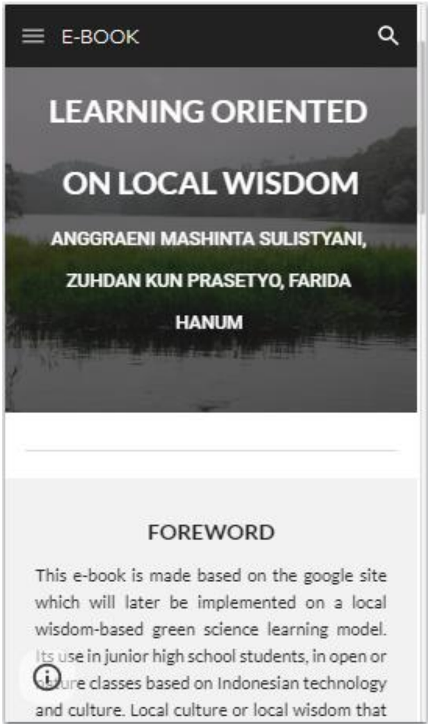


Fig. 4. E-book Display in a Smartphone.

Even by using the Google websites that we have designed, they can be accessed in the form of a smartphone display as shown in Fig. 4. The goal is to facilitate access so that students can use it anywhere and anytime.

The feasibility test of the product development draft from Fig. 3 and Fig. 4 was carried out through a Forum Group Discussion (FGD) between product makers and experts who assessed the feasibility of the product. Each expert who participated in the FGD provided an assessment and suggestion for the product. The results of the e-book validation questionnaire based on Google sites in the local wisdom-oriented green science model from the Learning Model Experts obtained scores according to Table III.

Following Table III, the total score is 31, and the conversion score is 86.11%, so it is included in the Valid category. From the two scores, according to experts, the product model is valid and suitable for use in learning. Suggestions from model experts are to realize the application of e-books in the learning of green science models oriented to local wisdom with student activities in the environment.

The results of the e-book validation questionnaire based on google sites in the local wisdom-oriented green science model from the Material Experts obtained scores according to Table IV.

Based on Table IV, the total score of 32 is calculated, and conversion score is 88.89, so it is included in the Valid category. From the two scores, according to the expert, the product material is valid and feasible to be implemented or applied in learning. Suggestions from material experts for product development, namely the learning theme is made to foster an attitude of caring for the environment.

TABLE III. MODEL EXPERT VALIDATION SCORE

No.	Assessment Indicators	Score	Percentage Conversion Score	Category
1.	Products are presented systematically	4	100	Very Valid
2.	Each technology-based activity presented has a clear purpose	4	100	Very Valid
3.	The activities presented can foster conceptual understanding, environmental care attitudes, and observation skills	4	100	Very Valid
4.	The product provides sufficient answer space	3	75	Valid
5.	Use of google sites	3	75	Valid
6.	The language used is following the level of students' cognitive development	3	75	Valid
7.	The language used is communicative	3	75	Valid
8.	The sentences used are clear and easy to understand	3	75	Valid
9.	Clarity of instructions or directions	4	100	Very Valid
total score		31	86,11	Valid

TABLE IV. MATERIAL EXPERT VALIDATION SCORE

No.	Assessment Indicators	Score	Percentage Conversion Score	Category
1.	Products are presented systematically	4	100	Very Valid
2.	Each technology-based activity presented has a clear purpose	4	100	Very Valid
3.	The activities presented can foster conceptual understanding, environmental care attitudes, and observation skills	3	75	Valid
4.	The product provides sufficient answer space	4	100	Very Valid
5.	Use of google sites	3	75	Valid
6.	The language used is following the level of students' cognitive development	3	75	Valid
7.	The language used is communicative	4	100	Very Valid
8.	The sentences used are clear and easy to understand	3	75	Valid
9.	Clarity of instructions or directions	4	100	Very Valid
total score		32	88,89	Valid

The results of the questionnaire validation assessment of e-books based on google sites in the local wisdom-oriented green science model from Media Technology and Information Experts obtained scores according to Table V.

TABLE V. MEDIA TECHNOLOGY AND INFORMATION EXPERT VALIDATION SCORE

No.	Assessment Indicators	Score	Percentage Conversion Score	Category
1.	Products are presented systematically	4	100	Very Valid
2.	Each technology-based activity presented has a clear purpose	4	100	Very Valid
3.	The activities presented can foster conceptual understanding, environmental care attitudes, and observation skills	3	75	Valid
4.	The product provides sufficient answer space	4	100	Very Valid
5.	Use of google sites	4	100	Very Valid
6.	The language used is following the level of students' cognitive development	4	100	Very Valid
7.	The language used is communicative	4	100	Very Valid
8.	The sentences used are clear and easy to understand	4	100	Very Valid
9.	Clarity of instructions or directions	4	100	Very Valid
total score		35	97.22	Very Valid

Based on Table V, the total score is 35 and the result of the calculation of the conversion score is 97.22, so it is included in the Very Valid category. From the two scores, according to the Media Technology Expert, the product is very valid and feasible to be implemented or applied in learning.

Nine assessment indicators in validating e-books based on Google sites in a local wisdom-oriented green science model, each measuring the feasibility of several aspects. These indicators become the basis for the assessment of experts. Based on the scores obtained from the three experts, then the scores of the three experts on average obtained a score of 90.74% which means that the student worksheets are very valid and can be applied to local wisdom-oriented green science learning models.

The expert's assessment of the product is suitable for use in learning, the comments of the experts support technology as a means of learning media by combining culture. Due to the pandemic conditions that force people not to leave the house, learning is now adapting to online-based learning [17]. Through this e-book, students can be helped in re-learning the material because the content can be accessed anytime and anywhere. Expert comments regarding the development of e-books based on google sites which are implemented in a local wisdom-oriented green science learning model must be applied in learning to find out how it affects students.

C. Implementation and Evaluation Stage

The implementation phase through the use of the product in grade 7 junior high school students as many as 32 students. The implementation is carried out by bringing students to open natural locations, precisely at Telaga Ranjeng in Brebes Regency, Central Java Province, Indonesia according to Fig. 5. This activity is carried out by integrating technology, namely e-books based on Google sites that can be accessed via smartphone applications.

Fig. 5. Students Conduct Learning Activities at Ranjeng Lake.

Students are asked to carry out activities according to the instructions on the student worksheets in the e-book while at Telaga Ranjeng as shown in Fig. 5. Student activities continue to apply the health protocol when interacting with the environment and the community around Telaga Ranjeng. The teacher only directs and guides in activities and explains learning material. After the students carried out the activity, a response questionnaire was distributed to find out the student's responses to the student worksheet. From the results of the student response questionnaires obtained in Table VI.

TABLE VI. STUDENT RESPONSE QUESTIONNAIRE RESULTS

No.	Response Indicator	Total Score of Respondents	Percentage Conversion Score	Category
1	The e-book presented is interesting and based on information technology	120	93.75	Very Valid
2	Each activity presented has a clear purpose	102	79.68	Valid
3	Activities related to environmental care	104	81.25	Valid
4	Clarity of description of local wisdom	122	95.31	Very Valid
5	The clarity of local wisdom, namely the myth of Telaga Ranjeng in science learning	123	96.09	Very Valid
6	Accuracy of presentation of material with learning objectives	120	93.75	Very Valid
7	Environmental awareness in learning	123	96.09	Very Valid
8	The suitability of the material with google sites-based technology media	107	83.59	Valid
9	Use of standard and clear language	100	78.12	Valid
10	The language used is communicative	119	92.96	Very Valid
11	The sentences used are clear and easy to understand	124	96.87	Very Valid
12	Ease of use	123	96.09	Very Valid
TOTAL		1387	90.29	Very Valid

Based on Table VI, the percentage score of student responses to student worksheets is 90.29%, indicating that students after using and carrying out activities on the student worksheets have a positive response. Judging from direct observation, students enjoy and have fun during learning. Enjoy the learning process based on technology, environment, and local wisdom. During the learning activities, students were supervised and observed to collect data on the environmental care characters that grew in junior high school students. The environmental care character data were analyzed through multiple regression tests to determine the effect of the independent variable (X_1 & X_2) on the dependent variable (Y).

The results of the partial variable test in multiple regression analysis are to determine the effect of each/partial of the independent variables (X_1 & X_2) on the dependent variable (Y) obtained following Table VII using SPSS 25 software.

TABLE VII. THE RESULTS OF THE REGRESSION ANALYSIS OF EACH INDEPENDENT VARIABLE ON THE DEPENDENT VARIABLE

Model		**Unstandardized Coefficients**		**Standardized Coefficients**	**t**	**Sig.**
		B	**Std. Error**	**Beta**		
1	(Constant)	21,984	13,142		1,673	0,105
	X1	,430	,118	,496	3,634	0,001
	X2	,368	,123	,409	2,994	0,006

a. Dependent Variable: Y

Based on Table VII in row X_1 obtained the value of Sig. 0,001 is smaller than the value of the degree of freedom, which is 0,05 (0,00 <0,05). Supported by the value of t-count SPSS analysis results in Table VII in row X_1 get a value of 3.634 which is greater than the t-table that is 2,045 (3,634>2,045). The significance interpretation and t-test analysis show that the X_1 variable is the use of e-books based on google sites which affects the Y variable of caring for the environment.

This shows that the independent variable of e-books based on google sites can influence the development of environmental care characters for junior high school students. Through e-books based on Google sites in a green science model oriented to local wisdom, students are allowed to cultivate themselves individually with group collaboration. Research, by Harjono et al [18] stated that e-books by adding interaction features by adding student activities, videos, and animations helped in strengthening students' concepts. Alqahtani et.al [19] also in their research on the effectiveness of e-learning during a pandemic, the current e-learning needed must prioritize different individual learning and provide control to students in learning. Through interactive activities carried out in e-book content based on Google sites in a local wisdom-oriented green science model, students are invited to interact fully in learning activities.

The second independent variable analysis, seen in Table VII rows X_2, obtained the value of Sig. 0,006 is smaller than the value of the degree of freedom, which is 0,05 (0,006<0,05). Supported by the t-count row X_2, the result of the SPSS Table VII analysis is 2,994, the value is greater than the t table is 2,045 (2,994>2,045). The significance interpretation and t-test analysis showed that the variable X_2 was the implementation of the local wisdom-oriented green science model which affected the Y variable for the environmental care character. The green science model itself applies the concept of environmental-based natural science learning, coupled with an Indonesian local wisdom oriented that is obedient and protects the environment in its application. The research of Sera et al [20], explains that local wisdom knowledge is considered very relevant to modern ecological principles so that it can be used in environmental or natural management. Ismawati [21] also stated that through local wisdom, we can instill environmental care with trusted habits and cultures. Christiawan [22] through strong local wisdom in an area can control the exploitation of nature, including forests. So, through learning the green science model oriented to local wisdom, students get to know environmental care by getting to know the culture that is embraced around them.

Then the results of the analysis of two independent variables (X_1 & X_2) on the dependent variable (Y). Data from e-books based on Google sites in a local wisdom-oriented green science model that is carried out to foster an attitude of caring for the environment. Based on multiple regression analysis using SPSS 25 software according to Table VIII.

TABLE VIII. THE RESULTS OF THE ANALYSIS OF MULTIPLE REGRESSION

Model		**Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
1	Regression	194,897	2	97,449	12,928	$0,000^b$
	Residual	218,603	29	7,538		
	Total	413,500	31			

a. Dependent Variable: Y

b. Predictors: (Constant), X2, X1

Based on Table VIII in column Sig. shows a value of 0,000 when compared to the value of the degree of freedom of 0,05, the value is smaller (0,000<0,05). It is confirmed that the calculated F value is 12,928 which is greater than the F table value, which is 3,32 (F_{-count} > F_{-table}). Following the significant value and F value, a decision was made that the accepted hypothesis is that there is a simultaneous influence on the use of e-books based on google sites (X_1) in the green science model oriented to local wisdom (X_2) there is a simultaneous influence on environmental care characteristics (Y).

e-Books based on Google sites in the green science model are oriented towards local wisdom to foster an attitude of caring for the environment, basically instilling in junior high school students to protect the environment, ecosystem management, and conservation. Due to the effects of a pandemic that does not allow students to go out for activities, as a result, they form a character who is indifferent to the environment [23]. Combining technology, and green science learning, namely learning natural sciences based on the environment and local culture can grow the character of caring for the environment. It is evident from the results of the SPSS analysis in Table VIII which shows the simultaneous effect on increasing the character of caring for the environment by using e-books based on google sites in a local wisdom-oriented green science model.

This can also be applied in learning in other schools so that students have a sense of concern for their environment. The values contained in local wisdom are closer and closer because they are part of everyday life so they are easy to understand. According to Glasson et al [24] and Ilhami et al [25] also stated that learning that links local content will deepen the material and increase the community's sense of care for environmental sustainability. According to Uge et al [26], it is explained that local wisdom is rarely used properly, especially in learning it can have an impact on environmental sustainability. Sofyan et al. [27] also support local wisdom that is applied in learning, through modules based on local wisdom students can learn about the culture and the environment.

IV. Conclusion

The conclusion from the data analysis and discussion shows the success of combining information technology, green science learning and local wisdom through the development of e-books. The success of e-book development is shown by the percentage of the three experts, namely 90.74%, indicating that the e-book based on Google sites in a local wisdom-oriented green science model is declared valid and feasible to use. Activities and the use of student worksheets and from direct observation students enjoy and have fun during learning. The implementation also shows the results of multiple regression analysis of Sig. 0.000<0.05, which means that the development of an e-book based on Google sites in a local wisdom-oriented green science model influences and fosters an attitude of caring for the environment.

V. Recommendations for Further Research

Indonesia has a diversity of cultures and rich natural resources so it is necessary to introduce students so that they know about culture and the environment as well as learn. It takes learning to link culture and the environment both in media, models, teaching materials, and other learning materials

Acknowledgment

The author thanks the Directorate of Research and Community Service (Direktorat Riset dan Pengabdian Kepada Masyarakat) Directorate General of Research and Technology Strengthening and Development of the Ministry of Research, Technology, and Higher Education for funding support in the form of Doctoral Dissertation Research (PDD) grants for the 2020 budget year.

References

[1] Spielman and G. Sunavala-Dossabhoy, “Pandemics and education: A historical review,” J. Dent. Educ., vol. 85, no. 6, 2021, DOI: 10.1002/jdd.12615.

[2] K. Layali and A. Al-Shlowiy, “Students� perceptions of e-learning for ESL/EFL in Saudi universities at the time of coronavirus: a literature review,” Indones. EFL J., vol. 6, no. 2, 2020, DOI: 10.25134/ieflj.v6i2.3378.

[3] H. A. Yanuarita and S. Haryati, “Pengaruh Covid-19 Terhadap Kondisi Sosial Budaya Di Kota Malang Dan Konsep Strategis Dalam Penanganannya,” J. Ilm. Widya Sosiopolitika, vol. 2, no. 2, p. 58, 2021, doi: 10.24843/jiwsp.2020.v02.i02.p01.

[4] N. A. Hidayati, H. J. Waluyo, R. Winarni, and Suyitno, “Exploring the implementation of local wisdom-based character education among Indonesian higher education students,” Int. J. Instr., vol. 13, no. 2, 2020, doi: 10.29333/iji.2020.13213a.

[5] W. Yang and H. Li, “The role of culture in early childhood curriculum development: A case study of curriculum innovations in Hong Kong kindergartens,” Contemp. Issues Early Child., 2020, DOI: 10.1177/1463949119900359.

[6] N. M. Ratminingsih, I. G. Budasi, and W. D. A. Kurnia, “Local culture-based storybook and its effect on reading competence,” Int. J. Instr., vol. 13, no. 2, 2020, doi: 10.29333/iji.2020.13218a.

[7] M. Nasir, Madlazim, I. G. M. Sanjaya, and Wahyudi, “Development of physics metacognition learning models; An innovative learning in Indonesia,” Int. J. Innov. Sci. Res. Technol., vol. 5, no. 1, pp. 1002–1005, 2020.

[8] D. H. Schunk, Learning theories: An educational perspective. 2012.

[9] A. Sindi, J. Stanfield, and A. Sheikh, “Technology in Education: Attitudes Towards using Technology in Nutrition Education,” Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 2, 2021, DOI: 10.14569/IJACSA.2021.0120208.

[10] H. H. Razami and R. Ibrahim, “Distance Education during COVID-19 Pandemic: The Perceptions and Preference of University Students in Malaysia Towards Online Learning,” Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 4, 2021, DOI: 10.14569/IJACSA.2021.0120416.

[11] O. K. Xin and D. Singh, “Development of Learning Analytics Dashboard based on Moodle Learning Management System,” Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 7, 2021, DOI: 10.14569/IJACSA.2021.0120793.

[12] D. V. Sigit, E. P. Azrai, E. Heryanti, I. Z. Ichsan, Y. P. Jajomi, and R. Fadrikal, “Development Green Consumerism E-Book for Undergraduate Students (GC-EBUS) as learning media in environmental learning,” Indian J. Public Heal. Res. Dev., vol. 10, no. 8, 2019, DOI: 10.5958/0976-5506.2019.02152.1.

[13] A. S. Adam and N. Suprapto, “One-stop physics E-Book package development for senior high school learning media,” Int. J. Emerg. Technol. Learn., vol. 14, no. 19, 2019, DOI: 10.3991/ijet.v14i1910761.

[14] T. F. Muqarrobin and H. Kuswanto, “Open Access Development of an android-based physics e-book to ease students ’ physics learning And its influence on their learning achievement American Journal of Engineering Research (AJER),” Am. J. Eng. Res., vol. 5, no. 10, 2016.

[15] R. N. Prasetyono and R. C. S. Hariyono, “Development of flipbook using web learning to improve logical thinking ability in the logic gate,” Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 1, 2020, DOI: 10.14569/ijacsa.2020.0110143.

[16] R. M. Branch, Instructional design: The ADDIE approach. 2010.

[17] Q. Kharma, K. Nairoukh, A. R. Hussein, M. Abualhaj, and Q. Shambour, “Online Learning Acceptance Model during Covid-19: An Integrated Conceptual Model,” Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 5, 2021, DOI: 10.14569/IJACSA.2021.0120561.

[18] A. Harjono, G. Gunawan, R. Adawiyah, and L. Herayanti, “An interactive e-book for physics to improve students’ conceptual mastery,” Int. J. Emerg. Technol. Learn., vol. 15, no. 5, 2020, DOI: 10.3991/IJET.V15I05.10967.

[19] N. Alqahtani and F. Nadeem, “Improving the Effectiveness of e-Learning Processes through Dynamic Programming: A Survey,” Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 5, 2021, DOI: 10.14569/IJACSA.2021.0120595.

[20] K. S. H. Udayani and S. Sriyati, “The Development of Teaching Materials Based on Bali Local Wisdom in Environment Conservation,” 2020, DOI: 10.2991/assehr.k.200130.100.

[21] R. Ismawati, E. Juliyanto, S. Prajoko, and E. Rahayuningsih, “Instilling The Conservation Character Through Reconstruction of Scientific Knowledge Based on Local Wisdom,” Indones. J. Sci. Educ., vol. 4, no. 1, 2020, doi: 10.31002/ijose.v4i1.1434.

[22] P. I. Christiawan, “The role of local wisdom in controlling deforestation,” 2017.

[23] A. M. Sulistyani, Z. K. Prasetyo, R. N. Prasetyono, and F. Hanum, “The Effect of Student Worksheets Oriented to The Myth of ‘Telaga Ranjeng’ to Foster an Attitude of Caring For The Environment,” 2020, DOI: 10.1145/3452144.3452159.

[24] G. E. Glasson, N. Mhango, A. Phiri, and M. Lanier, “Sustainability science education in Africa: Negotiating indigenous ways of living with

nature in the third space," Int. J. Sci. Educ., vol. 32, no. 1, 2010, DOI: 10.1080/09500690902981269.

[25] A. Ilhami, R. Riandi, and S. Sriyati, "Implementation of science learning with local wisdom approach toward environmental literacy," J. Phys. Conf. Ser., vol. 1157, no. 2, 2019, DOI: 10.1088/1742-6596/1157/2/022030.

[26] S. Uge, A. Neolaka, and M. Yasin, "Development of social studies learning model based on local wisdom in improving students' knowledge and social attitude," Int. J. Instr., vol. 12, no. 3, 2019, doi: 10.29333/iji.2019.12323a.

[27] H. Sofyan, E. Anggereini, and J. Saadiah, "Development of E-Modules Based on Local Wisdom in Central Learning Model at Kindergartens in Jambi City," Eur. J. Educ. Res., vol. 8, no. 4, pp. 1137–1143, 2019, DOI: 10.12973/EU-jer.8.4.1137.

A Two-Stage Assessment Approach for QoS in Internet of Things based on Fuzzy Logic

Mutasim Elsadig Adam[1]
Computer Sciences Department, Sudan University of Sciences and Technology, Khartoum, Sudan

Dr. Yasir Abdalgadir Ahmed Hamid[2]
College of Business. MIS
Alburaimi Univesity, Alburaimi, Oman

***Abstract*—In the sphere of IoT, one of the most significant issues is quality of service (QoS), which is critical for both developers and customers. As a result, IoT platform developers are working to enhance models that will meet consumer expectations in terms of IoT services meeting their expected specified levels of quality. The multidimensional architecture of the IoT platform, combined with the ambiguous mindset of consumers' thinking, makes QoS evaluation a difficult process. As a result, this study seeks to solve these issues and proposes a new paradigm for assessing QoS in IoT ecosystems. The proposed approach evaluates QoS in two steps, with the goal of assessing QoS at all levels. To address the issue of uncertainty, the metric values and QoS were represented using a fuzzy logic method. The model correctly estimated the QoS for 50 services in the dataset, and the results show that 16 services are classed as high quality, while 25 are rated as medium quality and the rest rated as low quality.**

Keywords*—IoT; internet of things; QoS; quality of services; fuzzy logic; evaluate; assess

I. Introduction

The IoT service is a key component in IoT ecosystems and is actually considered to be the backbone of it. Obviously, service plays a crucial role in any IoT platform as it is always the center of attention for both providers and consumers, for after all, it is the source from which both ends of the equation are potentially going to benefit. In general, the concept of services as defined by [1] is the commercial transaction between two parties in which one party allows another to access specific resources. Recent years have witnessed a significant development in IoT market, in turn, the number of IoT services have increased rapidly. This rapid increase in IoT services has been on the rise due to the dramatic increase in the number of devices, that are considered the main source of IoT services. There is no unanimous opinion on the number of things that will connect to the Internet [2]. According to most surveys undertaken by reputable companies such as Gartner, HIS Markit, and Ericson, the number of internet-connected devices will reach nearly 30 to 50 billion in 2021. (excluding smartphones, tablets, and computers) [3]. However, according to a recent estimate which conducted by [4], the number of linked devices worldwide will approach 75 billion by 2025. This reported big jump in the number of devices gives an indication of the massive number of IoT services that we will have and underpins the projection of a significant increase in the number of IoT services in the near future.

This consequential growth in IoT market motivates huge number of consumers to adopt IoT services. However, one of the big challenges that face most of the consumers is how to find the services fulfilling their specific requirements among massive numbers of services that have similar functionality but different in non-functional attributes. Therefore, the IoT developers endeavor to develop IoT platforms that would allow consumers to have access to services meeting their requirements. One of the most requirements most consumers seek and are concerned with is the quality of Services (QoS). According to [5], QoS is the overall performance of a telecommunication system, computer network, or IoT platform, explicitly, the performance perceived by network or platform users, and there are numerous metrics used to measure performance. As QoS is a key metric in measuring the performance of IoT platforms, many researches have been conducted with the view to tackling the issue of QoS in terms of QoS metrics and assessment methods. Researcher in [6] addressed the challenge of how to meet the consumers QoS-based demand whereas [7] discussed the problem of how to calculate the aggregate QoS of a composite service.

As we have noticed, QoS has significant impact on IoT performance. Hence, many researches have been developed in order to address the all issues related to the QoS in IoT platforms. QoS assessment is one of the most challenging issues that must be taken into consideration in any IoT platform, for more explanation, to develop a convenient IoT platform that would allow consumers to get the services with the desirable level of quality, it is necessary first to have a convenient approach for assessment of the QoS. However, developing an assessment approach for QoS in IoT is a rather challenging task for several reasons including: initially, the architecture of IoT platforms is made up of many layers and each layer has a conspicuous impact on the QoS. Thus, firstly, we assume that assessment the QoS for one layer and ignore the other layers will give a result that doesn't express the QoS for all IoT platform, secondly, calculate the QoS for all layers but deal with each result of layer individually also will affect the final result of QoS for IoT platform. On the other side, the non-functional attributes (QoS metrics) are represented in numerical values, in contrast, humans always think in inexact ways in their daily life, and used linguistic terms such as Short/tall, close/far, and hot/cold. Therefore, we assume that the way of representing the metrics of QoS has a great impact in the final results.

The goal of this work is to address the issues raised above by offering a new method for assessing QoS in IoT ecosystems. The suggested model examines QoS in two steps, with the goal of computing QoS at the all layers. To address the uncertainty issue, we employed the fuzzy logic approach to describe metric values and compute QoS.

II. Background

A. IoT Service Profile

The IoT service profile consists mainly of three models: Service profile, service model and service grounding, Fig. 1.

1) Service model: This model addresses the functional component of services at a high level; it specifies what a service does and is mostly used for service discovery [8].

2) Service profile: Mainly used for service selection. The goal of this section is to aid service selection by semantically describing the service's non-functionality. There are several QoS attributes that used to specify the non-functional of services such as: accuracy, reliability, security, reputation etc..

3) Service grounding: This section contains information about the message format and transfer protocol that are required for service invocation. Some models, for example, employ WSDL to deliver messages over common network protocols [9].

B. IoT Architecture from QoS Prospective

Form QoS prospective, IoT architecture made up of three layers, the perception layer, the network layer and the application layer, Fig. 2. In the perception layer we find the IoT devices from which services publish. The network layer represents the medium communication used to transfer services such as WiFi, ZigBee. The application layer is the interface where the end users access the services. Each layer has its own QoS metrics and based on these metrics the quality of services will be computed.

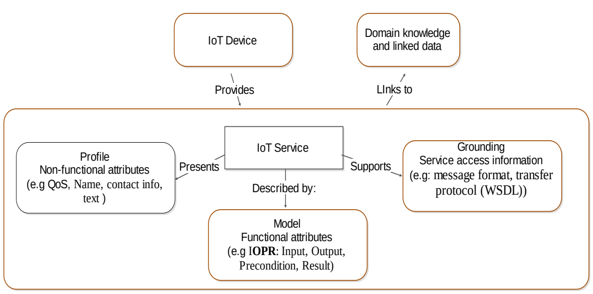


Fig. 1. Shows IoT Service Profile.

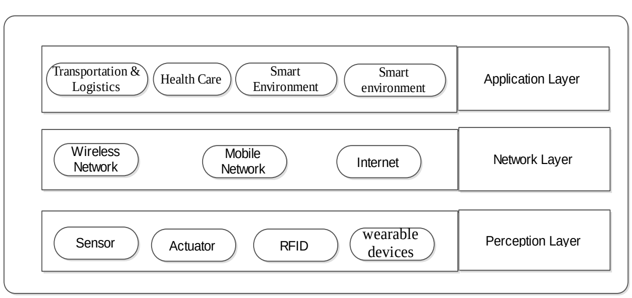


Fig. 2. Architecture in IoT from QoS Prospective.

1) QoS of Perception Layer: This layer is the backbone of the IoT and it is described as the five sense organs of loT ecosystems [10]. In this layer we find the IoT devices. These can be the edge devices such as sensors, actuators, RFID, cameras, GPS, wearable devices that interact with their environment. Several QoS metrics can be defined under this layer and they include:

a) Accuracy: Accuracy is one of the parameters that play vital role in QoS. Accuracy of sensor as defined by [11] is the greatest uncertainty between a sensor's actual value and the standard value defined at output settings.

b) Precision: Precision refers to the sensors' capacity to measure the deviation in the output obtained when the same signal is measured repeatedly under the same conditions [11].

c) Reliability: Reliability provides a high degree of trust and trustworthiness during the operation of IoT systems [12]. The reliability also reflects the ability of systems to recover and self-configure in the changing environment.

d) Sensor Time Constant (τ): A sensor constant time is defined as the time required for the sensor reading/output to reach to 63.2% of its total step change in measurement" [13].

e) Response Time: It is defined as the time taken by the sensors to change its output state with the change in input parameters. The Sensors that have low response time are more desirable for any application [11].

2) QoS of Network Layer: Network layer connects the IoT devices to other smart objects, servers, and network devices, it is considered as an intermediate between the application layer and perception layer [14] [15]. Several QoS metrics can be defined under this layer and they include:

a) Bandwidth: Represetns the measured amount of data that transmitted over netwrok at a given period of time. For accessing, the ones with a high available bandwidth are preferred for accessing [11].

b) Reliability: If the backet arrives at its destination without any loss or security breach, the service is considered reliable. [15].

c) Availability: Availability represented as the percentage of time, in a specific time interval, during which network components such as (a server, cloud service, or other machine) can be used for the task that it was mainly designed and created for [16] [17].

d) Security: There is an increasing demand for security among the IoT users. Most of the users want to ensure that the network used to provide IoT services is secure and the information provided will be treated confidentially [18].

3) QoS of Application (Cloud) Layer: It is responsible of providing the user with application-specific services. or what the user interacts with [19]. Several QoS metrics which can be defined under this layer and they include:

a) Price: The price of getting services in IoT is one of the metrics that reflects the QoS.

b) Reputation: Reputation of IoT services is measured based on the individual's experience and reviews. Many services selection approaches are customized based on the preferences [20].

c) Availability: As in sensor layer the availability also used as metric of QoS in application layer. It can be defined as a probability ratio that the service is operational and accessible when selected [21].

III. Related Work

The non-functional or QoS attributes are the key criteria when IoT end users need to select specific service to perform specific task. Therefore, it is important to develop an approach that allows providers of services to evaluate the QoS before making them available for the end users. Recently, many researches have been developed in order to improve tools to calculate or assess the QoS.

A study conducted by [22] investigated the problem of how to find the best service providers that offer smart parking services. The researchers identified 25 parameters that play a vital role in identifying the QoS in such a service. In [23], investigated the most suitable QoS criteria for optimal services selection problem in composition and classify them to the negative and positive criteria. In order to develop a good services evaluation model for IoT environment, a model based on multi-objective decision making (MODM) is proposed by [24]. The parameters that used to assess the QoS are: battery energy cost RE, CPU cost RC, memory usage RM, user-friendly RU and network bandwidth usage RB. The result obtained showed that this model successfully evaluated the QoS. A model based on fuzzy logic is developed by [25] in order to propose an assessment method to evaluate the level of QoS in IoT. The authors used execution time and reliability as the measurement metrics to the QoS. Fuzzy Logic Estimating Level (FLEL) estimates the QoS level by performing many tests. To evaluate the model, the researchers compares the proposed model with the existing model such as a Randomization Test (RT) and the estimating method by a Single Test in Steps (STS), and the result revealed that the efficiency of the FLEL is close to the STS and RT or even higher through the comparison of Average Paased Times APTs. in [26], a model developed to assess the QoS in IoT network layer. The parameters used in this study are end to end delay, energy consumption, energy fairness, jitter, throughput, routing load, packet delivery ratio and normalized routing overhead. The QoS of each parameter is evaluated by testing the performance of each parameter by increasing and decreasing the number of nodes in IoT network. The result obtained showed that, as the number of nodes grows, the performance of QoS measures such as end-to-end delay, energy consumption, energy fairness, routing load, and normalized routing overhead improves.

In [27], a model for evaluating QoS is developed based on five attributes. The attributes are classified in two categories: positive attributes (reliability, reputation, security) and negative attributes (cost, response time). To compute the QoS, the weight for each attribute is identified based on users' preferences, then the QoS is aggregated for negative and positive attribute separately using tow aggregation formulas.

To improve the services selection, [28] developed QoS model based on three components of IoT (things, communication and computing). The model provides description to the services in all three components. The aim of this study is to improve the services selection by considering the QoS in the three mentioned components. Evaluation of the quality of services in the proposed model passes through three steps: first the QoS parameters will be identified; secondly, the weight of each QoS parameter is calculated using Analytic Hierarchy Process (AHP) method. In the last steps the IoT services are ranked by using Order Preference by Similarity to Ideal Solution (TOPSIS). To validate the model, it compared with AHP, the evaluation showed that execution time taken by AHP- TOPSIS is less that than the time taken AHP.

A QoS model was established by [21] in order to compute QoS in the application layer. The main goal of this study is to tackle the issues of selecting the ideal service among similar functionally identical services and varied non-functionality needs. They classified the QoS factors into two categories: the first one is Business Quality Type (BQT), which assesses the quality of services from a business view point, the factors in this type like Reputation and Execution Price. The second type is System Quality Type (SQT). This type relates to the processing time of system and used factors such as Reliability, Availability and Response Time.

As we have seen, though the assessment of QoS in IoT ecosystems has been addressed by many researchers, however, some of the adopted solution developed based one specific layer and ignore the other layers. Moreover, most of the proposed solution did not consider the fuzziness of the QoS attributes. Therefore, in this paper is going to address all these issues by proposing new approach for computing the QoS in IoT ecosystems.

IV. Assessment QoS Model

In this research our assessment QoS model is based on a new approach. The model is based on three-layers IoT architecture as shown in Fig. 2. We used the fuzzy logic system to assess the QoS in each layer. The assessment of QoS performed in two stages (Fig. 3).

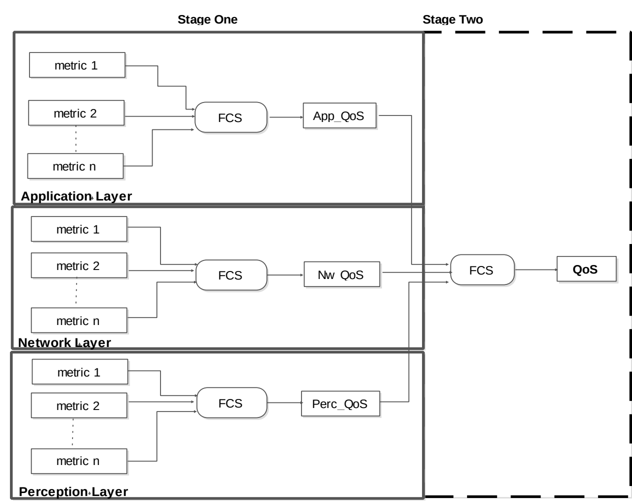


Fig. 3. Show the Two Fuzzy Stages of Assessment QoS for IoT Services.

In the first stage we have three FCSs which assess the QoS in perception layer, network layer and application layer. The metrics of each layer represent the inputs of the FCS. The result of this stage is three outputs named Perc_QoS, Nw_QoS and App_QoS, which represent the QoS of perception layer, network layer and application layer, respectively.

In the second stage the Perc_QoS, Nw_QoS and App_QoS obtained from first stage will be inputs of the FCS of the second stage. The result of this stage is the QoS which represents the final output of the system. The methodology that we used to perform the two stages mentioned is fuzzy control system which based on Lotfi A. Zadeh's fuzzy logic theory, which was developed in the 1960s to offer mathematical rules and functions that allowed natural language queries. The general idea behind the Fuzzy Logic is to mimic the way humans think [29], which tends to think in approximate rather than precise terms [30]. In other word, fuzzy logic is a method for describing and processing vague information that is commonly used by humans in their daily lives. Unlike traditional propositional logic, fuzzy logic assigns numeric values between 0 and 1 to each proposition in order to reflects uncertainty [31]. It is possible to calculate the degree to which an item belongs using fuzzy sets. For example, if a person is.83 tall, they are considered "rather tall." Fuzzy logic determines the shades of gray between black and white and true and false. The fuzzy logic control system typically consists of three major components, which are as follows:

(1) fuzzifier (2) inference rules and (3) defuzzifier. These components represent the sequences processes of the fuzzy logic controller (Fig. 4).

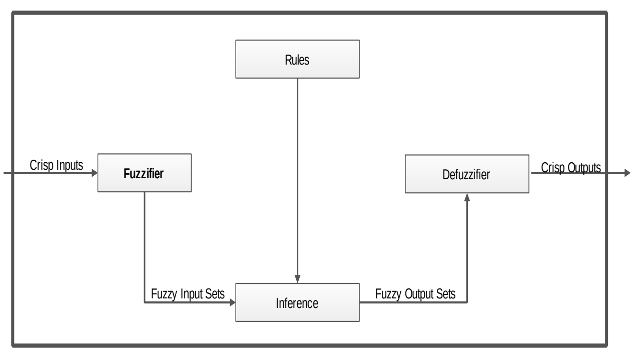


Fig. 4. Shows Fuzzy Logic Control System Architecture [32].

A. Fuzzification

This is the first step in developing FCS which explains the process of converting a set of crisp data into a set of linguistic variables using the membership functions (fuzzy sets) [33]. Fuzzification is a useful tool for dealing with vague and uncertain information, which can be objective or subjective. The membership function chart can be drawn in a variety of ways, including triangles, trapezoids, bell curves, or any other shape that accurately represents the distribution of information inside the system. A triangular fuzzy number is represented by a triplet (a, b, c), as shown in Fig. 5, and the membership function is calculated by equation (1):

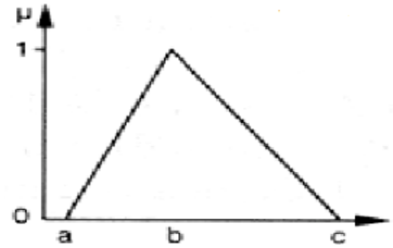


Fig. 5. Shows a Triangular Fuzzy Shape [34].

$$\mu(x) = \begin{cases} \frac{x-a}{b-a} a \le x \le b \\ \frac{c-x}{c-b} b \le x \le c \\ 0 \quad otherwise \end{cases} \tag{1}$$

A trapezoids fuzzy number is defined by four variables (a, b, c, d), as shown in Fig. 6, and the membership function is calculated by Equation (2):

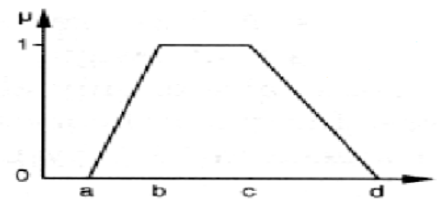


Fig. 6. Shows a Trapezoid Fuzzy Shape [35].

$$\mu(x) = \begin{cases} \frac{x-a}{b-a} a \le \mathrm{x} \le b \\ 1 \quad b \le \mathrm{x} \le c \\ \frac{d-x}{d-c} c \le \mathrm{x} \le d \end{cases} \tag{2}$$

B. Inference Mechanism

This is the main intelligent control of this system. Forming the rules base doesn’t has systemic tools that can be used as standard in developing the fuzzy logic controller [30]. As a controller designed as expert system, thus, developing the rule base depend on intuitive knowledge and the experience.

C. Defuzzification

After examining the fuzzy rules, the defuzzification component of fuzzy logic converts the fuzzy data values into real-world data values, and these real-world data values are determined by the defuzzification method. Different methods are used for defuzzification process such Center of Gravity (SOG), Weighted Average method, Mean of Maxima (MOM) and Smallest of Maxima (SOM). These methods are set by the controller designer.

V. Implementation and Results

Based on the above definition and steps of developing FCS, we used Fuzzy Logic Designer in MATLAB to develop the proposed model. Our purpose of this research is to assess the QoS of IoT service in all layers. As shown in Fig. 3, the calculation of QoS will be performed in two stages:

A. Stage One

In this stage we develop three FCSs for the three layers. For straightforwardness we selected tow metrics in each layer to represent inputs for FCS. Upcoming in this we will describe in details the steps of building the FCS for the three layers:

1) Fuzzification: For each layer we choose two QoS metrics to represent the inputs of the FCS. For example, we choose price and reputation for application layer, bandwidth and reliability for network layer and accuracy and response time for perception layer. Then we identified the fuzzy set of each input and outputs in each layer as showed in Table I and Table II.

We used triangular and trapezoid shapes to represent the fuzzy set for all inputs and outputs in the three layers.

2) The rule base: The inputs identified in the fuzzification step will be applied to a set of IF/then control rules in this step. The results of this step are combined to produce a set of fuzzy outputs. In our proposed model we have 2 inputs for each layer, and each input has 3 fuzzy sets, as showed in the Table I. Therefore, the number of control rules that we need will be based on the equation (3) below:

$$\text{Number of control rules} = n^m \quad (3)$$

where n = the number of fuzzy sets and m = the number of inputs.

By applying the above rule, number of control rules = 3^2 = 9 rules (for each FCS).

In this paper we used (And) as conjunction operator to form the rules as showed below:

If (metric1 is membership function1) and (metric2 is membership function2) then (output is output membership function).

TABLE I. SHOWS THE FUZZY SETS FOR INPUTS

Layer	**Input Parameters**	**Fuzzy set**	**Universe of Discourse**
Application Layer	Price	Cheap, medium, Expensive	0 – 15
	Reputation	Low, Medium, High	0 – 100
Network Layer	Bandwidth	Low, Medium, High	0-100
	Reliability	Low, Medium, High	0-100
Perception Layer	Accuracy	Low, Medium, High	0-100
	Response time	Fast, Medium, Slow	0-1

TABLE II. SHOWS THE FUZZY SETS FOR OUTPUTS

Output Parameters	**Fuzzy set**	**Universe of Discourse**
App_QoS	Low, Medium, High	0 -10
Nw_QoS	Low, Medium, High	0 -10
Perc_QoS	Low, Medium, High	0 -10

3) The defuzzification: Since the outputs from the previous step still fuzzy and uncertain, a defuzzification process is needed. The function of the defuzzification is to convert back the linguistic variable obtained from the previous step into numerical value in order to make the fuzzy output suitable to use in real application [36]. we used the MOM fuzzification method in order to get the results. From this step we received three outputs which represent the QoS for IoT services in perception layer, network layer and application layer, these outputs are Perc_QoS, Nw_QoS and App_QoS respectively.

B. Result of Stage One

To test the new model, we used the random function in MATLAB to generate a data set of 50 services. By applying the model to this stage, the model successfully calculates the QoS for all 50 services in perception layer, network layer and application layer. The Table III shows random samples of results obtained in this stage.

TABLE III. SHOWS SAMPLE OF RESULTS OF STAGE ONE

Service_ID	App_QoS	Nw_QoS	Perc_QoS
1	5.00	5.10	5.10
5	5.00	5.10	3.50
8	8.50	8.50	4.20
9	5.00	8.50	5.10
10	1.10	5.10	0.10
11	5.00	1.40	7.50
12	5.00	8.50	5.10
13	1.30	8.50	0.10
16	1.10	1.40	7.60
17	5.00	5.10	0.10
21	1.40	1.10	3.20
24	1.10	8.50	5.10
31	5.00	1.50	3.90
33	8.50	0.80	7.40
35	5.00	5.10	7.90
50	8.50	8.50	0.10

C. Stage Two

The FCS for this stage is based on the outputs of the FCSs of stage one. As we mentioned before, we received three outputs from the FCSs of stage one, which are: App_QoS, Nw_QoS and Perc_QoS. These outputs will be the inputs of the FCS for this stage. To develop the FCS, we follow the same steps as in the first stage.

1) Fuzzification: In this step we identified the fuzzy set of the inputs parameters as shows in Table IV.

The result of this stage is one output which represents the QoS. And as the purpose of this research is to assess the Quality of IoT service and put it into three categories (high quality, medium quality and low quality), we identified the fuzzy set of the output based on these categories, Table V.

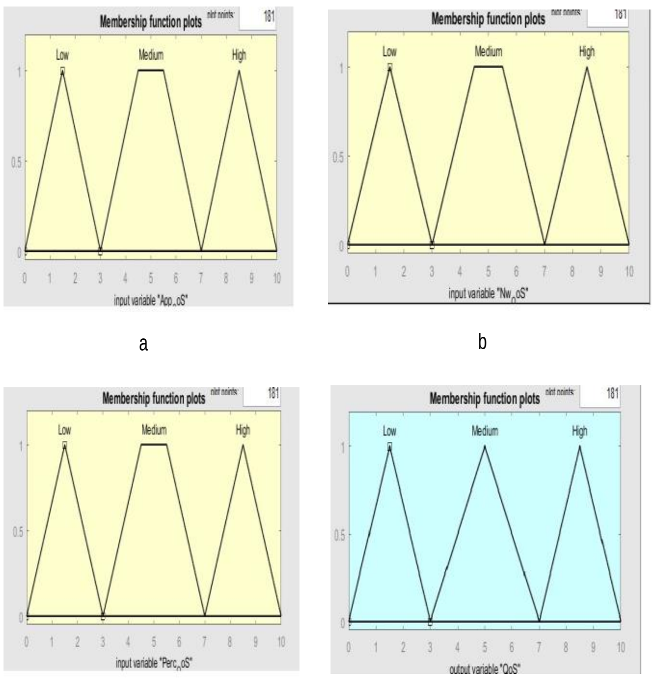
TABLE IV. SHOWS THE MEMBERSHIP FUNCTION FOR INPUTS

Inputs Parameters	Fuzzy set	Universe of Discourse
App_QoS	Low, Medium, High	0 –100
Nw_QoS	Low, Medium, High	0 –100
Perc_QoS	Low, Medium, High	0-100

TABLE V. SHOWS THE MEMBERSHIP FUNCTION FOR OUTPUTS OF FCS FOR STAGE TWO

Outputs Parameters	Fuzzy set	Universe of Discourse
QoS	Low, Medium, High	0 –10

We used triangular and trapezoid shapes to represent the fuzzy set for all inputs and output in FCS Fig. 7(a-b-c-d).



a b

c d

Fig. 7. Shows the Representation Fuzzy Set for Inputs and Output.

2) The rule base: As we have 3 inputs and 3 fuzzy set for each input, and referred to the equation (3), the number of the rules will be = 3^3 = 27 rules.

3) The defuzzification: We received from the previous step the output which represents the QoS. To convert the fuzzy output to crisp form, we used MOM defuzzification method.

D. Result of Stage Two

The result obtained from stage one has become the dataset for stage two. By applying the dataset in this stage, the model successfully calculates the QoS for all 50 services which are based on the three layers. Table VI shows the final result we obtained and which represents QoS of some of the targeted services.

TABLE VI. SHOWS QOS OF SOME OF SERVICES

Service_ID	App_QoS	Nw_QoS	Perc_QoS	QoS
1	5.00	5.10	5.10	5.00
5	5.00	5.10	3.50	5.00
8	8.50	8.50	4.20	8.50
9	5.00	8.50	5.10	8.50
10	1.10	5.10	0.10	1.50
11	5.00	1.40	7.50	5.00
12	5.00	8.50	5.10	8.50
13	1.30	8.50	0.10	1.50
16	1.10	1.40	7.60	1.50
17	5.00	5.10	0.10	5.00
21	1.40	1.10	3.20	1.50
24	1.10	8.50	5.10	5.00
31	5.00	1.50	3.90	5.00
33	8.50	0.80	7.40	8.50
35	5.00	5.10	7.90	8.50
50	8.50	8.50	0.10	8.50

VI. DISCUSSION

Through two stages our model calculated the QoS for 50 services. In the first stage the model calculated the QoS for all services in perception layer, network layer and application layer. Then in the next stage the model used the results of the first stage to calculate the QoS for all targeted services. Unlike other approaches which calculate QoS by focusing on some layers and ignore the other layers, or calculate the QoS in all layers but deal with each layer separately, our model evaluated the QoS in three layers and combined the results of all layers in order to get the final result which represents the QoS for all targeted services.

Based on the final result obtained and reference to the range of fuzzy set of QoS parameters that we identified earlier in the FLC Fig. 7(d), we have successfully classified the all services into three categories: high quality services, medium quality services and low quality services. Based on this classification: the number of high quality services is 16, the number of the medium quality services is 25 and the number of the low quality services is 9.

Our findings show that our model was able to calculate the QoS for all services in the dataset. As a result, the model will be of great assistance to service providers by providing them with convenient tools that allow them to accurately assess QoS before delivering it to end users. In addition, the model enables end users to request services using linguistic terms rather than numerical values. When compared to existing models such as AHP-TOPSIS, our developed model is more flexible while AHP-TOPSIS is very restricted because the end user is required to use numerical values to identify the weight of each metric. To explain this point, for example suppose that we have a user need an IoT services with a medium quality. In AHP-TOPSIS the user must use numeric value as the model does not support fuzziness. The user has to say ("I need a service with quality = 55”), but if there is no service with exact number in registry, the system will replies “null”, maybe in registry there are services with medium quality = 56 or 54, But because the AHP- TOPSIS model is very restricted, the other services will not be recommended to the end user. In our model we have solved this problem by using fuzzy logic technique. Instead of use numbers, end user just has to say (“I need a service with a medium quality”), immediately the model will recommend 25 services that match the user requirement.

To validate the model, we compare its execution time with AHP-TOPSIS. AHP-TOPSIS took 0.0386 second to rank 50 services, whereas our model took 0.2274 second to assess QoS for 50 services. This slight increase in execution time in our developed model is due to the model calculating QoS across all layers and then aggregating the results to obtain the final results.

VII. CONCLUSION

Delivering a service that meets certain standards of quality demanded by end users in the realm of IoT has clearly piqued the interest of many academics. As a result, it is essential to first build a model that allows IoT service providers to measure QoS before making them available to users. We solved this issue in this research by building a novel model that allows us to calculate QoS for all targeted services. The main contribution of this model lies in the evaluation of QoS in IoT platforms at the level of three layers. Moreover, the model used the fuzzy logic system in order to support the uncertainty thinking of humans. Over all, the model provides IoT services providers with a convenient tool which allows them to evaluate QoS for all available services. In future work, firstly we will try to improve the performance of the model, then we will use the results obtained from this model to develop a selection model that would allow IoT service consumers to select the kind of services meeting the levels of Quality they anticipate and look for.

REFERENCES

[1] Barros and D. Oberle, “Handbook of service description: USDL and its methods,” Handb. Serv. Descr. USDL Its Methods, no. February, pp. 1–538, 2012, doi: 10.1007/978-1-4614-1864-1.

[2] M. E. Adam, “Usages of Semantic Web Services Technologies in IoT Ecosystems and its Impact in Services Delivery : A survey,” vol. 4523.

[3] A. Nordrum, “Popular Internet of Things Forecast of 50 Billion Devices by 2020 Is Outdated,” 2016. https://spectrum.ieee.org/tech-talk/telecom/internet/popular-internet-of-things-forecast-of-50-billion-devices-by-2020-is-outdated.

[4] Statista, “Internet of Things (IoT) connected devices installed base worldwide from 2015 to 2025,” Stat. Res. Dep., 2022, [Online]. Available: https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/.

[5] Https://www.iotone.com/, “Quality of Service:,” 2004. https://www.iotone.com/term/quality-of-service/t622 (accessed Sep. 19, 2021).

[6] C.-L. Fok, C. Julien, G.-C. Roman, and C. Lu, “Challenges of satisfying multiple stakeholders,” p. 55, 2011, doi: 10.1145/1988051.1988062.

[7] Y. Yang, M. Dumas, L. García-Bañuelos, A. Polyvyanyy, and L. Zhang, “Generalized aggregate Quality of Service computation for composite services,” J. Syst. Softw., vol. 85, no. 8, pp. 1818–1830, 2012, doi: 10.1016/j.jss.2012.03.005.

[8] D. Martin et al., “Bringing semantics to web services with OWL-S,” World Wide Web, vol. 10, no. 3, pp. 243–277, 2007, doi: 10.1007/s11280-007-0033-x.

[9] U. Bellur, H. Vadodaria, and A. Gupt, “Semantic Matchmaking Algorithms,” Greedy Algorithms, no. November, 2008, doi: 10.5772/5884.

[10] W. D. Fang, W. He, W. Chen, L. H. Shan, and F. Y. Ma, “Research on the application-driven architecture in internet of things,” Front. Artif. Intell. Appl., vol. 293, pp. 458–465, 2016, doi: 10.3233/978-1-61499-722-1-458.

[11] M. Singh and G. Baranwal, “Quality of Service (QoS) in Internet of Things,” Proc. - 2018 3rd Int. Conf. Internet Things Smart Innov. Usages, IoT-SIU 2018, pp. 1–6, 2018, doi: 10.1109/IoT-SIU.2018.8519862.

[12] A. Mavrogiorgou, A. Kiourtis, C. Symvoulidis, and D. Kyriazis, “Capturing the Reliability of Unknown Devices in the IoT World,” 2018 Fifth Int. Conf. Internet Things Syst. Manag. Secur., pp. 62–69, 2018.

[13] B. D. Technologies, “Difference between sensor response time and sensor time constant τ (tau) 63.2%.,” 2019. https://www.baranidesign.com/faq-articles/2019/5/6/difference-between-sensor-response-time-and-sensor-time-constant-tau#:~:text=Sensor Response Time %3D 5*τ,five times the time constant. (accessed Aug. 28, 2020).

[14] Y. Li, Y. Huang, M. Zhang, and L. Rajabion, “Service selection mechanisms in the Internet of Things (IoT): a systematic and comprehensive study,” Cluster Comput., vol. 23, no. 2, pp. 1163–1183, 2020, doi: 10.1007/s10586-019-02984-4.

[15] L. A. K. Subash, D. Janet Ramya and S. Awarded, “Quality of Service in the Internet of Things (IoT) – A Survey,” Res. Teach. Learn. Lett., vol. 21, no. June, 2019.

[16] S. Gridelli, “How to calculate network availability?,” NetBeez, 2014. https://netbeez.net/blog/how-to-calculate-network-availability/ (accessed Nov. 08, 2020).

[17] N. M. M. Alhag and Y. A. Mohamed, “An Enhancement of Data Encryption Standards Algorithm (DES),” 2018 Int. Conf. Comput. Control. Electr. Electron. Eng. ICCCEEE 2018, vol. 1, pp. 1–6, 2018, doi: 10.1109/ICCCEEE.2018.8515843.

[18] R. A. A. Farah and Y. A. Mohamed, “Adaptive Immune-Based System for Network Security,” 2018 Int. Conf. Comput. Control. Electr. Electron. Eng. ICCCEEE 2018, pp. 1–7, 2018, doi: 10.1109/ICCCEEE.2018.8515827.

[19] A. K. Sikder, G. Petracca, H. Aksu, T. Jaeger, and A. S. Uluagac, “A Survey on Sensor-based Threats to Internet-of-Things (IoT) Devices and Applications,” arXiv, 2018, [Online]. Available: http://arxiv.org/abs/1802.02041.

[20] G. S. M. Sathya, M. Swarnamugi, P. Dhavachelvan, “Evaluation of QoS based Web- Service Selection Techniques for Service Composition,” JInternational J. Softw. Eng., vol. 110, no. 9, pp. 73–90, 2017.

[21] A. N. Abu-safe and S. E. Elrofai, “An Efficient QoS-aware Services Selection in IoT using A Reputation Improved- Social Spider Optimization Algorithm,” Res. Sq., 2020, doi: 10.21203/rs.3.rs-38596/v1.

[22] S. Singh and R. K. Tiwari, “Quality of Services in Assets of Smart Parking,” Proc. - 2018 3rd Int. Conf. Internet Things Smart Innov. Usages, IoT-SIU 2018, pp. 1–5, 2018, doi: 10.1109/IoT-SIU.2018.8519885.

[23] Y. Dongre and R. Ingle, “An Investigation of QoS Criteria for Optimal Services Selection in Composition,” 2nd Int. Conf. Innov. Mech. Ind. Appl. ICIMIA 2020 - Conf. Proc., no. Icimia, pp. 705–710, 2020, doi: 10.1109/ICIMIA48430.2020.9074950.

[24] S. Fan, W. Shi, N. Wang, and Y. Liu, “MODM-based evaluation model of service quality in the Internet of Things,” Procedia Environ. Sci., vol. 11, no. PART A, pp. 63–69, 2011, doi: 10.1016/j.proenv.2011.12.011.

[25] B. Jia, L. Hao, C. Zhang, and D. Chen, “A dynamic estimation of service level based on fuzzy logic for robustness in the internet of things,” Sensors (Switzerland), vol. 18, no. 7, pp. 1–13, 2018, doi: 10.3390/s18072190.

[26] & K. Sheikh, A., Ambhaikar, A., “analysis of QoS parameters in IoT network 2021.pdf,” Open J. Sci. Technol., 2021.

[27] X. Zhang, J. Geng, J. Ma, H. Liu, and S. Niu, “A QoS-driven Service Selection Optimization Algorithm for Internet of Things,” Eur. PMC, 2020.

[28] M. S. • G. B. • A. K. Tripathi1, "QoS-Aware Selection of IoT-Based Service.pdf," Arab. J. Sci. Eng., vol. 45(12), no. 10033-10050., 2020.

[29] D. Shannaq, B., & Thakkar, "On The Development of Neural Network Models Using Data Mining Tools," Asian J. Comput. Sci. Inf. Technol., vol. 2, no. 7, 2012.

[30] L. A. Zadeh, "Fuzzy sets," nformation Control, vol. 8, no. 3, pp. 338–353, 1965, [Online]. Available: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=L.A.+Zadeh%2C+1965%2C+Fuzzy+Sets.&btnG=.

[31] Charles Elkan, "fuzzy logic," 1999, [Online]. Available: https://www.scientificamerican.com/article/what-is-fuzzy-logic-are-t/.

[32] M. Norsuzila Ya'acob et al., "Enhancing Fuzzy Controllers Using Generalized Orthogonality Principle," Intech, vol. 32, pp. 137–144, 1989, [Online]. Available: http://www.intechopen.com/books/trends-in-telecommunications-technologies/gps-total-electron-content-tec-prediction-at-ionosphere-layer-over-the-equatorial-region%0AInTec.

[33] T. Sai, D. Nakhaeinia, and B. Karasfi, "Application of Fuzzy Logic in Mobile Robot Navigation," Fuzzy Log. - Control. Concepts, Theor. Appl., no. March 2012, 2012, doi: 10.5772/36358.

[34] F. Liu, S. Chen, M. Heiner, and H. Song, "Modeling biological systems with uncertain kinetic data using fuzzy continuous Petri nets," BMC Syst. Biol., vol. 12, no. Suppl 4, 2018, doi: 10.1186/s12918-018-0568-8.

[35] H. Fakhrurroja, S. A. Mardhotillah, O. Mahendra, A. Munandar, M. I. Rizqyawan, and R. P. Pratama, "Automatic pH and Humidity Control System for Hydroponics Using Fuzzy Logic," 2019 Int. Conf. Comput. Control. Informatics its Appl. Emerg. Trends Big Data Artif. Intell. IC3INA 2019, no. July 2020, pp. 156–161, 2019, doi: 10.1109/IC3INA48034.2019.8949590.

[36] Y. Bai and D. Wang, "Fundamentals of fuzzy logic control — fuzzy sets, fuzzy rules and defuzzifications," Adv. Ind. Control, no. 9781846284687, pp. 17–36, 2006, doi: 10.1007/978-1-84628-469-4_2.

Design and Usability Study of Hypertension Management Guideline Mobile Application with Hypertension and Non-hypertension Patients

Nor Azman Ismail*, Nor Atiqah Mohd Fua'ad, Muhammad Syahmi Zulkifli, Farhat Embarak
Nur Zuhairah Afiqah Husni, Su Elya Mohamed, Puteri Syaza Kamarina Megat Mohd Zainon
School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, Skudai, Malaysia

***Abstract*—Hypertension is currently rising steadily among the world population. The first level of screening to know whether one is suffering from hypertension is essential as this lays the foundation for the actual diagnosis. This research details the user interface design and usability evaluation of the hypertension management guideline. The proposed mobile application prototype assists people in screening themselves with regards to hypertension based on symptoms. This prototype also acts as a sharing platform for hypertension patients to help them share their concerns and advice within the related online community. The eye-tracker experiment was used to support the visual strategy of the prototype design. In studying the usability of the mobile application, an experiment carried out with two groups of people, of which one group of people have hypertension. In contrast, the other group of people do not have hypertension. An independent-samples t-test conducted to compare the user performance scores using the proposed prototype. Based on the usability study, both user groups understood and used the applications with ease. However, the findings revealed there was a significant difference in overall scores for hypertension patients and non-hypertension patients. The findings of this study could help software developer design an effective application for hypertension guideline tool for monitoring health and well-being.**

Keywords—Hypertension management guideline; hypertension patient; hypertension symptoms; user interface; user experience

I. Introduction

High blood pressure (hypertension) is not usually something that one can feel or notice. It does not give any symptoms ahead or apparent signs. The only way to know that one is suffering from high blood pressure is to have it measured [1-2]. Recently in 2015, a survey was conducted by National Health and Morbidity concerning hypertension. Based on the overall prevalence of hypertension, 30.3% hypertension-related disease is among adults of 18 years and above in Malaysia. Apart from age, hypertension patients are also categorized based on the grouping of areas. In rural areas, the prevalence of people suffering from hypertension is 33.5% compared to an urban area with a percentage of 29.3%. Other findings from the survey are that the overall percentage prevalence of people suffering from hypertension according to their gender is lower among females with 29.7% whereas higher among males with 30.8%. The survey result provides the observation that hypertension diseases can attack anyone regardless of their age, gender, and even their place [3]. This evidence directs that hypertension is a severe and dangerous disease and should be firmly dealt with as it can also lead to death.

The information gathered helps to conduct an in-depth study which in turn helps in creating an application that can help hypertension patients [4]. Various difference methods have used to build a related mobile app [5-9]. However, based on previous studies, it is found that most related applications provide guidance that is too general and less helpful for patients with different symptoms. In this study, a user-centred approach is adopted in designing Hypertension Management Guideline mobile app for helping the hypertensive patients. Interviews were conducted with hypertension patients as used in Jolles et al. [10] to get to know more about the hypertension disease and how they handled the condition. Also, two different groups of people were invited, one group that suffers from hypertension disease and another group that does not suffer from hypertension disease. This study also used the visual strategy with an eye tracker in order to find out the pattern behaviour of the selected users looking at the proposed mobile app design. The proposed framework also supports a simpler user interaction through the user interface. In the following section, the brief description of Hypertension Management Guideline mobile app prototype is developed as MYHpGuide and the methods used to complete this study are discussed.

The rest of the article is structured as follows. Section 2 details the recent literatures of the prototypes used for self-diagnosis of hypertension; Section 3 illustrates the materials and methods used for the development of proposed prototype design; Section 4 describes the prototype design and user experimental results; and finally, Section 5 concludes the paper.

II. Related Work

In this technology era, people are quite advanced and prefer to self-diagnose their ailments with the help of the World Wide Web before visiting the doctors. In the case of hypertension, self-screening helps the patients to measure their blood pressure, symptoms, and diagnose their disease [11].

Carrera et al. [12] offer a user-friendly high blood pressure (HBP) app to monitor hypertensive patients. The HBP app called BP control is an Android and IOS app which allows the hypertensive patients to communicate their blood pressure measurements with their clinician, thus providing monitoring

*Corresponding Author.

and diagnosis. Cruz et al., [13] propose a mobile application called HeartBeat+ for recording and monitoring blood pressure readings for hypertensive patients. It provides an easy way for self-monitoring. The application is a combination of a mobile application for patients and web applications for the clinician to monitor the status of their patients. The web application keeps track of the vital records of the patient, charts of the previous readings, medical history, allergies, habits, and medications. The mobile application can be used by hypertensive patients to feed and monitor their vital signs, daily activities, medications, and food intake. In turn, it helps in identifying what triggers the patients' high blood pressure every time.

Alvarez et al., [14] provide a hypertension prevention method in the lower economic region of Peru with information and communication technology. In this paper, a mobile application called Wondershare-MobileGo is used, which provides a record of the blood pressure of the patients in real-time. Some of the papers also focus on using smartwatch along with smartphone and mobile applications for blood pressure monitoring [15]. Schaeffer et al., [16] present a prototype in his paper that combines an Android application called Lifestyle App, a smartphone, and a smartwatch which all integrated with a Server Lifestyle. The data input from the smartwatch and the smartphone application integrates with the server which lets to several abstract concepts of connection and communication between different devices, which facilitates the development of better lifestyle for patients based on the blood pressure measurement and daily activities.

Meanwhile, Pulgarin et al., [17] study presents the integration of a physical arterial blood pressure monitor with a web server with the help of WIFI interconnection module. They use a wrist blood pressure monitor that measures the heart rate, systolic, and diastolic pressure data, which in turn stores the measured data in an EEPROM storage memory with I2C connectivity, integrates with a Wi-Fi interconnection module to enable communication with the server [18]. The major drawback of conventional hypertension management system is its higher cost of purchasing and system maintenance.

III. Materials and Methods

The overview of the design, development and evaluation phase of the proposed model is illustrated in Fig. 1.

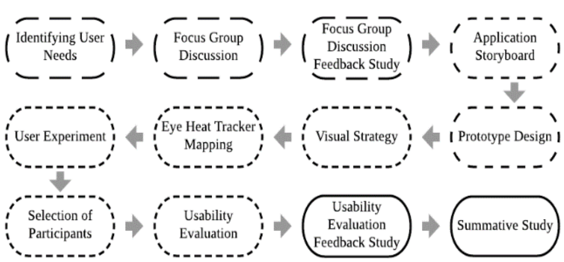


Fig. 1. Overview of the Design, Development and Evaluation Process. Milestones: Application Discussion Study (Rough Dotted), Concept Development (Medium Dotted), Visual Strategy and usability Evaluation (Dense Dotted), and Summative Evaluation (Solid Line).

A. Identifying User Needs

An experiment was carried out to identify the user needs on the proposed prototype applications. Two main user groups were involved in this experiment. The first group comprises of hypertension patients who have experience with other hypertension applications previously, and the second group includes non-hypertension patients. To gain information and in-depth understanding of hypertension disease, a focus group discussion was conducted from the perspectives of hypertension patients. Each talk was audio recorded. A survey was also conducted with a set of questionnaires with other people on their requirements of the related applications.

B. Prototype Design

Low fidelity prototype has been designed based on the findings of identifying user requirements experiment and Gestalt Principle [19]. Also, the visual strategy is used by conducting eye tracker experiment for the proposed prototype screen and layout design. The proposed prototype design was also tested using an eye tracker device and software called Tobbi Pro. The device detected the movement of the user's eye during the testing of the application's user interface [20]. The device is linked to the software and generates a heat map based on the user's eyes focus during the testing of the user interface.

C. User Experiment

Ten participants are chosen for the usability study of MYHpGuide mobile application. Those who are non-hypertension patients were eager to take part to know more about this application. After the testing, the participants need to give their opinion about the app. Smartphones and laptops were used as a platform to do the testing. The instructions were given to each participant on how to use the proposed prototype mobile application. Below are the tasks that assigned to the participants for the experiment:

Task1: Use the search bar and select headache symptom

Task2: Browse and select headache symptom guidelines

Task3: Bookmark, like and share headache symptom guidelines

Task4: Give feedback on headache symptom guidelines by typing "The guidelines were very helpful" on the feedback box for bulleted list.

IV. Results and Discussion

A. Identifying user Needs

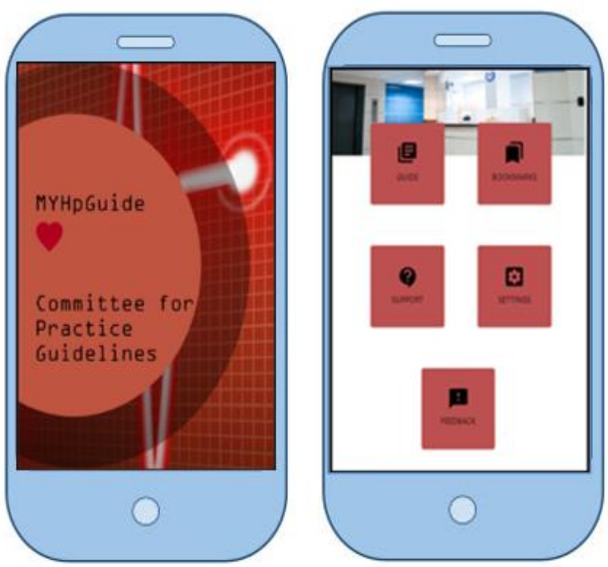
The results from the interview indicate that most hypertension and non-hypertension patients come from various age ranges for both men and women. Thus, it is concluded that the proposed application design and directions should not be complicated and rather be friendly to the user so that the application is accessible among adults and elderly users. The main component is consisting of sets of guidelines to manage hypertension disease. The priority of proposed system feature and functionality of hypertension management guideline application is divided by hypertension and non-hypertension patients. Table I shows the percentages of hypertension and non-hypertension group who have chosen each component.

TABLE I. PRIORITIES OF NON-HYPERTENSION AND HYPERTENSION ON HYPERTENSION MANAGEMENT GUIDELINE APPLICATION COMPONENTS

Component of Self-care Application	Hypertension (%)	Non-Hypertension (%)
Self-Monitoring	100	100
Diet	70	86
Exercises	80	80
Personal Data	75	90
Activities Planning	55	70
Goal Setting	90	97
Online Education	45	88
Forum	39	80
Social Support	98	80
Alert / Reminder	56	88
Coaching	60	70
Meditation	89	77
Reward System	45	56

B. Prototype Design

Our proposed prototype is MYHpGuide application, as shown in Fig. 2(a), a hypertension management guideline application on a smart phone that designed specifically for patients diagnosed with hypertension. It has a database of hypertension guideline management for a patient to refer. Other than that, user can also like, share, save and give feedback for any guideline that they want as illustrated in Fig. 2(b).



(a) (b)

Fig. 2. Front Page (a) and Main Menu of MYHpGuide Application (b).

The task given to participants is to take note of the time in which they complete a task, their understanding of the tasks and tracking their eyes to understand their point of focus. These parameters were noted for both the participating groups.

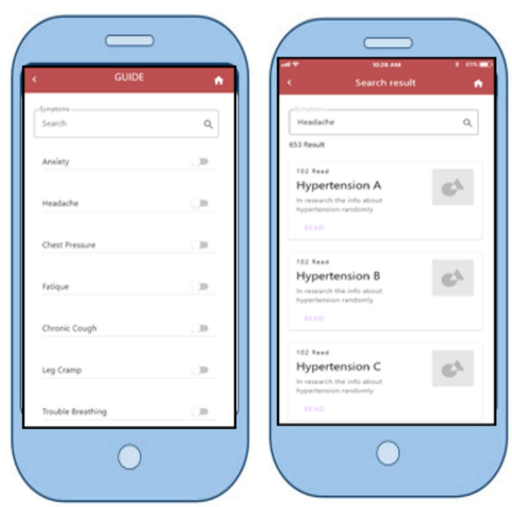
In the first task, participants need to type "headache" as a symptom in the search box, as shown in Fig. 3 below. Participants can also browse through the symptoms already shown in the app and can choose the headache symptom and or choose any other additional symptoms as per their need. However, some participants were confused about attempting the first task, as most thought that they can only choose the symptoms shown in the list and not type in.

In the second task, the participants need to browse the headache symptom guidelines and select the proper guidelines, as shown in Fig. 3 (Task 1). A list of hypertension guidelines regarding the headache symptom is shown in a list (Task 2). The list view of guidelines shows the title and a short description of the guideline, including the number of people who have read it. In the third task, the participants need to bookmark, like or share the headache symptom guidelines. The application allows the participants to like and share the guidelines if the participants think that the guidelines are helpful. A bookmark button provided so the participants can save the guidelines for future use. The final task for the participants was that the users could give feedback for the guideline that they choose. A feedback box was provided for the users to share their thoughts and comments regarding the guidelines.

C. Visual Strategy

The primary purpose of the visual strategy in the usability study was to identify the suitable user interface design. The group conducted the visual strategy by using the eye tracker test. In the Color mapping result, the region where the eye focused longer represented with an increasing degree of a warmer colour. For the front page, Fig. 4 shows the logo of the apps had the warmest tone because it had the most observation from the test by the users.

Other than that, the result for the guideline page shows that the author's name had the warmest tone because the users were inquisitive about the author app, as shown in Fig. 5.



Task 1 Task 2

Fig. 3. Type on the Search or Select Symptom on Task1 and Browse all the Guide shown based on the Symptom Task 2.

Fig. 4. Heat Map for Front Page of MYHpGuide Apps.

Fig. 5. Heat Map for Guideline of MYHpGuide Apps.

Lastly, Fig. 6 shows the result of the guideline system main menu. The results show that the users were more keenly observing the logo of each menu as all the logos had the warmest tone, especially for bookmarks and guideline menu logo.

Fig. 6. Heat Map for Guideline of MYHpGuide Apps.

D. User Experimental Results

An independent-samples t-test conducted to compare the performance scores for hypertension patients and non-hypertension patients. There was a significant difference in scores for hypertension patients (M=3.33, SD=0.08) and non-hypertension patients [M=4.08, SD=0.33); t(8)=-5.41, p<0.01]. The magnitude of the differences in the means is (eta squared= 0.079). Based on the result, all participants successfully managed to complete all the task given. Table II shows the result of user performance. For task 1, 2, 3 and 4, the time difference between the users spending their time on the application was not too big between the two groups. All the data have been recorded in Table II as shown, the time users spend using the application (minutes) with the task given to two groups of hypertension patients and are non-hypertension patients are provided. The result shows that the mean difference between hypertension patients (m=3.3333) and non-hypertension patients (m=4.0750) is approximately 7 second apart. When compared with all the tasks, the value may seem significant, but in real life situation, the 7-second difference is not that big and noticeable.

TABLE II. USER PERFORMANCE RESULTS

TASK	The Time Users Spend Using the Hypertension Management Guideline Mobile Application (minute)									
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
	(HP)	(NHP)	(HP)	(HP)	(NHP)	(HP)	(HP)	(NHP)	(NHP)	(HP)
Task 1	0.8	0.6	0.5	0.7	0.6	0.7	0.4	0.9	0.8	0.5
Task 2	1.0	1.3	0.8	0.7	1..2	0.9	0.8	0.9	1.5	0.5
Task 3	0.5	0.8	0.7	0.8	0.9	0.8	1.0	0.7	0.6	0.9
Task 4	1.1	1.2	1.3	1.2	1.6	1.0	1.0	1.2	1.5	1.4
TOTAL TIME	3.4	3.9	3.3	3.4	4.3	3.4	3.2	3.7	4.4	3.3

RESULT			
Type of Participant	Amount of Participant	Mean	Std. Deviation
Hypertension Patient (HP)	6	3.3333	0.08165
Non-Hypertension Patient (HP)	4	4.0750	0.33040

However, it still shows the difference between the performances of the two-user group to complete the task. The standard deviation of hypertension patients overall time taken to complete all tasks is more prominent (SD: 0.33040) than non-hypertension patients (SD: 0.08165). It proves that, in using the hypertension guideline management application, non-hypertension patients take more time than hypertension patients.

Usability testing result was analyzed; both groups can complete all given tasks successfully. In using the MYHpGuide application, non-hypertension patients take more time than hypertension patients to complete all task. An independent-samples t-test showed that there was a significant difference in time taken to complete all the tasks between hypertension patients and non for non-hypertension patients. This result may be due to the lack of experience of non-hypertension patients with hypertension symptoms and related terminology. Compared to non-hypertension users, patients with hypertension are using the proposed apps efficiently, and they took a shorter time to complete all task given because they already familiar and have some experienced with various hypertension symptoms.

The result from the eye tracker test shows that users focused more on visual content (e.g., icons and graphics) compared to words. It can also be concluded that user's eye is more focus on symbols rather than texts because users maybe need to identify and analyze the logo meaning. Other than that, the eye tracker test results also show that users may be more interested in looking at the name of the author's guideline because of their curiosity (see Fig. 5(B)).

It can also be concluded that the eye heat tracker test used on the visual strategy for the proposed design is sufficient to prove the understanding of the target user on the user interface features and the structure of the proposed application. Other than that, the result also shows the design strength and weakness that have on the user interface design, which can guide us to improve the user interface design of related applications.

V. Conclusion

The ability of user-centered design in considering the user's needs, devise a plan to meet them, and direct the elderly person's perspective to a positive attitude, has made it plays a significant role throughout the design phase of mobile application for hypertension management guideline. The result shows that the app presented in this paper has a higher engagement of user interaction with the features present in the user interface. Based on the heat map, it can be concluded that the visual engagement with icon and graphic has higher interaction and ease of use for the user. It is also observed that the hypertension group users spend less time on each task as they usually use other similar apps and are experienced with this, while the other non-hypertension group users take more time to get to know each feature in the app. The reported usability evaluation demonstrates that hypertension user performs positively with ease. The app can assist them and provide considerable benefit with the interaction with other similar hypertensions user. This study could help in designing an effective application for hypertension guideline and innovative tool for monitoring health and well-being during daily life activities.

Acknowledgment

We want to thank VicubeLab for providing infrastructure for usability testing and eye heat tracker session. The authors also wish to thank the Ministry of Higher Education Malaysia for funding this study under Fundamental Research Grant Scheme (FRGS/1/2019/ICT04/UTM/02/8) with Grant No – 4L809, Universiti Teknologi Malaysia.

References

[1] Reule S, Drawz P. Heart Rate and Blood Pressure: Any Possible Implications for Management of Hypertension? Current Hypertension Reports. 2012;14(6):478-484.

[2] Johnson H, Thorpe C, Bartels C, Schumacher J, Palta M, Pandhi N et al. Undiagnosed hypertension among young adults with regular primary care use. Journal of Hypertension. 2014;32(1):65-74.

[3] Tabrizi J, Sadeghi-Bazargani H, Farahbakhsh M, Nikniaz L, Nikniaz Z. Prevalence and Associated Factors of Prehypertension and Hypertension in Iranian Population: The Lifestyle Promotion Project (LPP). PLOS ONE. 2016;11(10):e0165264.

[4] Kang H, Park H. A Mobile App for Hypertension Management Based on Clinical Practice Guidelines: Development and Deployment. JMIR mHealth and uHealth. 2016;4(1):e12.

[5] Alessa T, Abdi S, Hawley M, de Witte L. Mobile Apps to Support the Self-Management of Hypertension: Systematic Review of Effectiveness, Usability, and User Satisfaction. JMIR mHealth and uHealth. 2018;6(7):e10723.

[6] Anderson K, Burford O, Emmerton L. App Chronic Disease Checklist: Protocol to Evaluate Mobile Apps for Chronic Disease Self-Management. JMIR Research Protocols. 2016;5(4):e204.

[7] Ankem K, Cho S, Simpson D. Nurses' perceptions and problems in the usability of a medication safety app. Informatics for Health and Social Care. 2017;44(1):48-69.

[8] Hui C, Creamer E, Pinnock H, McKinstry B. Apps to Support Self-Management for People With Hypertension: Content Analysis. JMIR mHealth and uHealth. 2019;7(6):e13257.

[9] Jolles E, Padwal R, Clark A, Braam B. A Qualitative Study of Patient Perspectives about Hypertension. ISRN Hypertension. 2013;2013:1-10.

[10] Fleming S, Atherton H, McCartney D, Hodgkinson J, Greenfield S, Hobbs F et al. Self-Screening and Non-Physician Screening for Hypertension in Communities: A Systematic Review. American Journal of Hypertension. 2015;28(11):1316-1324.

[11] Jamaladin H, van de Belt T, Luijpers L, de Graaff F, Bredie S, Roeleveld N et al. Mobile Apps for Blood Pressure Monitoring: Systematic Search in App Stores and Content Analysis. JMIR mHealth and uHealth. 2018;6(11):e187.

[12] Carrera A, Pifarré M, Vilaplana J, Cuadrado J, Solsona S, Mateo J et al. BPcontrol. Applied Clinical Informatics. 2016;07(04):1120-1134.

[13] Cruz B, Encina A, Macale H, Lavarias C, Ramirez E, Jamis M et al. HeartBeat+: A Development of Mobile Application on Recording and Monitoring Blood Pressure Readings for Hypertensive Patients. 2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology,Communication and Control, Environment and Management (HNICEM). 2018.

[14] Marin-Alvarez J, Quispe-Sipan J, Sotomayor-Beltran C. Hypertension Prevention in a Human Settlement of the Peruvian Capital Using Information and Communications Technology. 2019 IEEE International Conference on E-health Networking, Application & Services (HealthCom). 2019.

[15] Cuadrado J, Vilaplana J, Mateo J, Solsona F, Solsona S, Rius J, Alves R, Camafort M, Torres G, Betriu A, Gutierrez JM, Fernández E., (2016), "HBPF: a Home Blood Pressure Framework with SLA guarantees to follow up hypertensive patients," PeerJ Computer Science 2:e69, 2016, https://doi.org/10.7717/peerj-cs.69.

[16] Schaeffer, M. K., Veiga, J. E., Biduski, D., Rebonatto, M. T., & De Marchi, A. C. B., (2017), "Android app lifestyle — Smartphone and smartwatch integred into a cloud computing by web services," 2017 12th Iberian Conference on Information Systems and Technologies (CISTI). doi:10.23919/cisti.2017.7975814.

[17] Bolivar Pulgarin, N. G., Cangrejo Aljure, L. D., & Salcedo Parra, O. J., eHeart-BP, (2019), "Prototype of the Internet of Things to Monitor Blood Pressure," 2019 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE). doi:10.1109/chase48038.2019.00025.

[18] Stojanova A, Koceski S, Koceska N., (2019), "Continuous blood pressure monitoring as a basis for ambient assisted living (AAL)—review of methodologies and devices," J Med Syst. 2019;43(2):24.

[19] Dumitru, ML and Joergensen GH., (2016), "Gestalt Reasoning with Conjunctions and Disjunctions," PLoS ONE 11(3):e0151774, 2016, Interaction Design Foundation, Gestalt Principle.

[20] Asan O, Yang Y., (2015), "Using Eye Trackers for Usability Evaluation of Health Information Technology: A Systematic Literature Review," JMIR Hum Factors 2015;2(1):e5.

A CNN based Approach for Handwritten Character Identification of Telugu Guninthalu using Various Optimizers

B. Soujanya[1]
Assistant Professor, Dept of Computer Science and Engineering
Institute of Technology, GITAM (Deemed to be University, Visakhapatnam, India

Suresh Chittineni[2], T. Sitamahalakshmi[3]
Professor, Dept of Computer Science and Engineering
Institute of Technology, GITAM (Deemed to be University
Visakhapatnam, India

G. Srinivas[4]
Associate Professor, Dept of Computer Science and Engineering, Institute of Technology, GITAM (Deemed to be University, Visakhapatnam, India

***Abstract*—Handwritten character recognition is the most critical and challenging area of research in image processing. A computer's ability to detect handwriting input from various original sources, such as paper documents, images, touch screens, and other online and offline devices, may be classified as this recognition. Identifying handwriting in Indian languages like Hindi, Tamil, Telugu, and Kannada has gotten less attention than in other languages like English and Asian dialects like Japanese and Chinese. Adaptive Moment Estimation (ADAM), Root Mean Square Propagation (RMSProp) and Stochastic Gradient Descent (SGD) optimization methods employed in a Convolution Neural Network (CNN) have produced good recognition, accuracy, and training and classification times for Telugu handwritten character recognition. It's possible to overcome the limitations of classic machine learning methods using CNN. We used numerous handwritten Telugu guninthalu as input to construct our own data set used in our proposed model. Comparatively, the RMSprop optimizer outperforms ADAM and SGD optimizer by 94.26%.**

***Keywords*—Character recognition; Adam; RMSProp; SGD; CNN**

I. Introduction

In today's world, the internet is brimming with images and video representations, providing sufficient opportunity for building numerous research applications for image and video analysis [1] to educate people about more complex material and techniques. With the rise of Artificial Neural Networks, machine learning has advanced significantly in recent years (ANN). These ideas enhance the model's capabilities beyond machine-learning tasks and other domains. Convolutional neural network (CNN) architecture has been considered as one of the most inventive. Using CNN in image processing became clearer and more beneficial as ANN performance deteriorated in object recognition and image classification. As better CNN became accessible, research using CNN in image processing domains grew dramatically [2-4]. CNNs have had a lot of success in various domains, including computer vision, natural language processing, and speech recognition.

One of the most widely used machine learning models is CNN which has been expanded to handle a wide range of visual image applications, item classification, and audio identification challenges by applying mathematical representations. Multi-layer network structure that may be learned and consists of several layers [5]. Raw pixel values may be utilized as input to the network instead of feature vectors, which are often employed in machine learning. Even though there are many different kinds of CNNs (fig.1), they always have the same basic structure: a convolutional layer, a pooling layer, and an entirely connected layer.

1) Convolution layer: Images are filtered using this tool, which identifies characteristics that are used to identify matching spots during testing. Enlarged images need a convolution procedure with minimum parameters. With a filter or kernel, the input data is transformed into a feature map for use by CNN.

2) Pooling layer: This layer receives the extracted characteristics. It reduces bigger images while keeping the most critical data. It keeps the maximum value from each window by preserving the best fit value. This function shrinks the picture spatially to minimize the number of parameters and computations in the model. Max Pooling is a typical strategy in pooling. It selects the greatest element from the feature map covered by the filter.

3) Fully connected layer: High-level filtered images are fed in and categorized using labels in the final layer. Every neuron in this layer is related to the one below it. Layers of convolution and pooling are common in most designs.

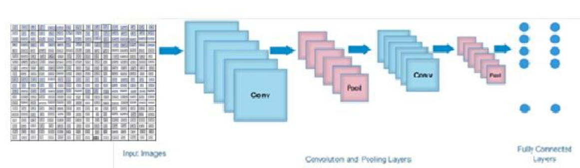


Fig. 1. Layers CNN.

II. Literature Review

CNN performs well in various applications, inspiring academics to work on it in key fields such as natural language processing (NLP), image classification and face recognition, predictive analytics, and so on. P.V. Ramana Murthy et al. [6] built a model that recognizes online handwritten Telugu letters for various domains and companies, yielding a model with 98.3 per cent accuracy, exceeding their expectations. P. Sujatha et al. [7] used CNN architecture to identify a few deep learning strategies for detecting Telugu and Hindi scripts. They also developed a new architecture for identifying low-level textual properties of handwritten characters. Buddharaju Revathi et al. [8] also conducted a survey on Optical Character Recognition (OCR) for the Telugu language, demonstrating the progress of processing the characters stage by stage and performing operations such as segmentation and processing, which resulted in higher accuracy.

B. Hari Kumar et al. [9] used data from several sources to conduct script identification in Telugu. Konkimalla Chandra Prakash et al. [10] used CNN for Telugu script recognition. They supplied a list of Telugu typefaces, a client-server solution for the algorithm's online deployment, and deep learning-based OCR techniques in their study. The segmentation method can be improved so that each character, along with its gunintham and vattu, is segmented. Chirag I Patel et al. [11] developed a technique for identifying characters in a given digitalized text and reading the changing effects of the Models utilizing ANN by employing a back growing neural network to improve script identification accuracy. A.Ram Bharadwaj et al. [12] built a model for Telugu text extraction and recognition using CNN and a recurrent neural network (RNN) with their own data set, achieving an accuracy of 81% by selecting 100 random words from the validation set.

III. Proposed Method

In image-based classification, the CNN architecture is a popular choice. To distinguish features like edges and forms, CNNs apply a range of filters composed of trainable parameters to an image. These high-level filters often utilize weights learned from the spatial attributes of each subsequent level to capture an image's spatial properties. For this design, parameters named as Hyper Parameters are employed. These parameters may influence both the network architecture and training before the training begins.

Hyper Parameters used for building the network are:

- Count the number of hidden layers: the layers between input and output.
- To prevent overfitting, dropouts are employed.
- The activation function is used to add some nonlinearity.
- Learning Rate: It specifies the rate at which the network's parameters are updated.
- Momentum: It keeps things from oscillations.
- One complete prediction cycle of a neural network is one epoch.

Batch Size: The number of subsamples sent to the network.

Fig. 2 [13] depicts the suggested model's implementation stages.

To speed up CNN model training, all of the images in the build data set were categorized, reduced to 70 by 70 pixels, transformed to greyscale, and saved in .npz format. Telugu Guninthalu dataset was trained using three hidden layers of a CNN and tested on it. The network consists of three convolutional layers, two Maxpooling layers, and a fully connected layer. The first hidden layer is a Convolution2D layer. A Maxpooling layer with a pool size of 22 was then employed, along with 256 filters, each with a kernel size of 33 and a stride of 1. A layer of padding has been added to the design to keep the original input size.

Each of the 256 convolution filters in the second convolution layer has a kernel size of 33, followed by a Maxpooling layer of 22. New features have been added to the algorithm, including an additional 256-filter convolution layer and a Maxpooling layer of pool size 22. The two-dimensional matrix data is first transformed into a one-dimensional vector using a Flatten layer before the fully connected layers are generated. Next, we used a totally corresponding layer with the activation of ReLu in it. Dropout, a regularization layer, is set to randomly eliminate 20% of the layer neurons to prevent overfitting.

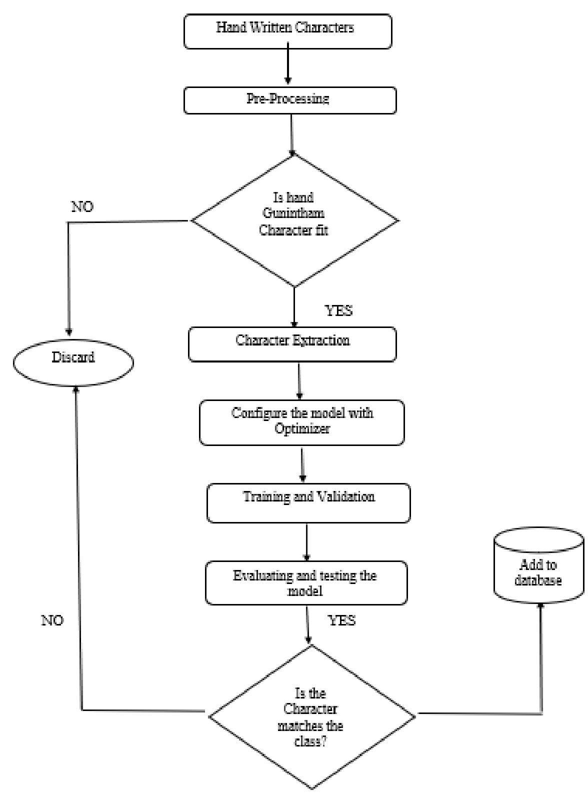


Fig. 2. Process Flow.

At long last, a sigmoid activation function is added to the 16-neuron output layer. An essential part of CNN is the activation function; it determines the output of a neuron concerning a set of inputs. The activation function is used to introduce nonlinearity into the model. A CNN model's performance is improved by selecting the appropriate activation function. The ReLu and Sigmoid activation functions are used in the suggested model.

1) Rectified Linear Unit (ReLu): It's a biologically and mathematically sound activation function [14]. If the input is negative, ReLu returns 0, if the input is positive; it returns the value itself (Eq 1). ReLu's max operation calculates more quickly than other activation functions. For many kinds of neural networks, it is the default activation function.

$$f\left(x\right)=max\left(0,x\right) \tag{1}$$

If we are doing a matrix-vector product, this is often done element by element to get the desired outcome. Nonlinearity is vital for CNN since every layer in the network contributes to it.

2) Sigmoid: Eq 2 indicates that it is a probabilistic strategy for decision-making with a range of 0 to 1. We used this activation function to forecast an output since it would be more accurate.

$$f\left(x\right) \ = \ 1/\left(1+e^{-x}\right) \tag{2}$$

It's beneficial for models that forecast probability as a result. Because the probability of anything spans between 0 and 1, a differentiable sigmoid function is the best option. As a result, this function yields the curve's slope at any two locations.

Training and validation errors were calculated after each cycle. The training is completed when the number of epochs in training and validation hasn't changed considerably. On the test set, faults in training and validation were discovered, and the network was approximated. It is vital to utilize optimizers in order to improve accuracy and decrease mistakes. Optimizers alter the weight settings in order to minimize the loss function. To get the greatest possible design, we considered a set of priorities or limitations. Adaptive Moment Estimation (Adam), Root Mean Square Propagation (RMSprop) and Stochastic Gradient Descent (SGD) were used in the proposed model.

3) Adaptive Moment Estimation (Adam): Using Adam, an adaptive learning rate technique, individual learning rates are calculated. The exponentially declining average of the preceding squared gradient is stored, and the value of previous historical gradients is preserved, much as momentum. It may also be used to replace stochastic gradient descent systems when updating the network weights in training data. Equations 3 and 4 are used to derive the gradients, which are then used to estimate the moments using exponentially moving averages. Where m_t and v_t are moving averages, g is gradient, β1, β2, and second moment of gradients are gradient forgetting features, and index t is the current training iteration.

$$m_t=\beta_1 m_{t-1}+\left(1-\beta_1\right)g_t \tag{3}$$

$$v_t=\beta_2 v_{t-1}+\left(1-\beta_2\right)g_t^2 \tag{4}$$

The recommended Adam configuration parameters [15] of β_1=0.9 and β_2=0.999.

4) Root Mean Square Propagation (RMSprop): This approach is similar to a gradient descent with the momentum that confines oscillations to the vertical plane. Thus, increasing the learning rate enables the algorithm to make considerable horizontal jumps toward rapid convergence. The magnitude of recent gradient descents is used to normalize the gradient. Selecting alternative learning rates for each parameter causes the pace of learning to be automatically changed. The parameters are updated using Eq. 5, 6, where gt is the gradient at time t, vt is the exponential average of squares of gradients, and η is the learning rate, which is set at 0.001.

$$v_t=\rho v_{t-1}+\left(1-\upsilon\right)g_t^2 \tag{5}$$

$$\Delta\omega_t=-\frac{\eta}{\sqrt{v_t+\varepsilon}}g_t \tag{6}$$

5) Stochastic gradient descent: Stochastic Gradient Descent (SGD) is a simple yet efficient optimization algorithm used to find the parameters/coefficients of functions that minimize a cost function. In other words, it is used for discriminative learning of linear classifiers under convex loss functions such as SVM and Logistic regression.

$$E\left(w,b\right)=\frac{1}{n}\sum_{i=1}^{n}L\left(yi,\,f\left(xi\right)\right)+\alpha R\left(w\right) \tag{7}$$

IV. Experimental Results

A. Data Set

Due to the lack of publicly available training data for Telugu characters, we had to construct our own dataset. There is an initial gathering of people's handwritten Telugu guninthalu in various formats and scanning them into precise symbols. Fig. 3 shows a scanned copy of a handwritten guninthalu both online and offline. There are 16 characters in each of the 21 guninthalu, for a total of 275520 handwritten characters.

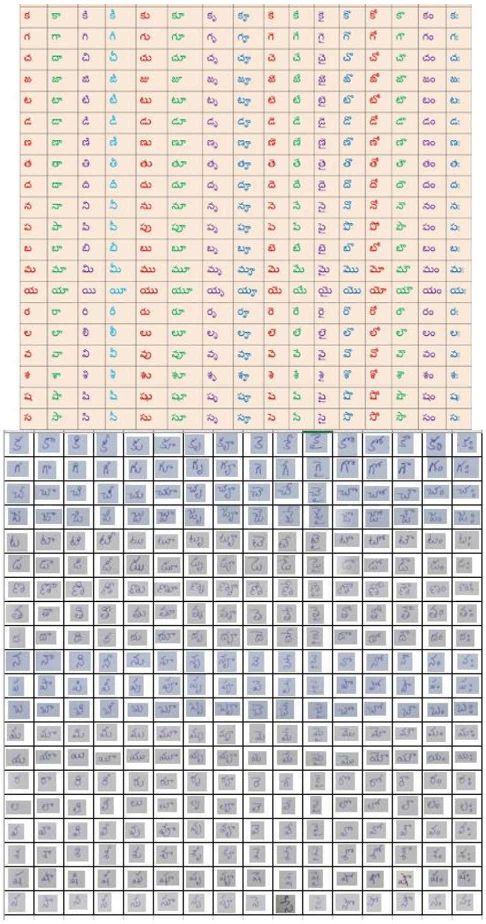


Fig. 3. Notations used in the Model (Both Online and Offline).

V. Requirements

The backend engine for the recommended model is Tensor Flow 2.2. CNN was implemented using the built-in dataset. These tests were run on a 64-bit operating system with an Intel i7-4770 CPU running at 4.00GHz and 16GB of RAM.

A. *Results*

The suggested model is applied to the pre-existing data set in the system. This research aims to develop a system for recognizing handwritten characters in Telugu, which makes extensive use of classification. Table I displays the model's layer-by-layer attributes. The proposed network will handle a total of 1,986,640 parameters.

Metrics including recall, precision, and accuracy [16] are used to assess the proposed system.

- Positive patterns that can be correctly anticipated from the totality of other positive patterns are "predicted with precision".

$$\mathrm{Pr}\,ecision = \frac{true\ positives}{true\ positives + false\ positives}$$

- Recall computes the fraction of positive patterns that are appropriately classified.

$$\mathrm{Re}\,call = \frac{true\ positives}{true\ positives + false\,negatives}$$

- In order to measure the model's classification accuracy, the percentage of accurate predictions is calculated. The formula provided may be used to figure this out.

$$Accuracy = \frac{true\ positives + true\,negatives}{true\,examples}$$

Tables II to V illustrate the results of different optimizers such as Adam, SGD and RMSprop with all the performance metrics. When compared to Adam, and SGD the RMSprop optimizer generated superior results. When RMSprop is employed, the learning rate is enhanced since it takes huge horizontal steps and converges faster. The loss function quantifies how close the network's predicted output and given ground truth labels are through forward propagation. The confusion matrices of different optimizers are shown in Fig. 4 to 6.

TABLE I. Parameters of CNN Model

Layer	Output Shape	Parameters
conv2d_16 (Conv2D)	68, 68, 256	2560
activation_21 (Activation)	68, 68, 256	0
max_pooling2d_	34, 34, 256	0
conv2d_17 (Conv2D)	32, 32, 256	590080
activation_22 (Activation)	32, 32, 256	0
max_pooling2d_	16, 16, 256	0
conv2d_18 (Conv2D)	14, 14, 256	590080
activation_23(Activation)	14, 14, 256	0
max_pooling2d_	7, 7, 256	0
flatten_6 (Flatten)	12544	0
dense_11 (Dense)	64	802880
dropout_6 (Dropout)	64	0
dense_12 (Dense)	16	1040
activation_24(Activation)	16	0

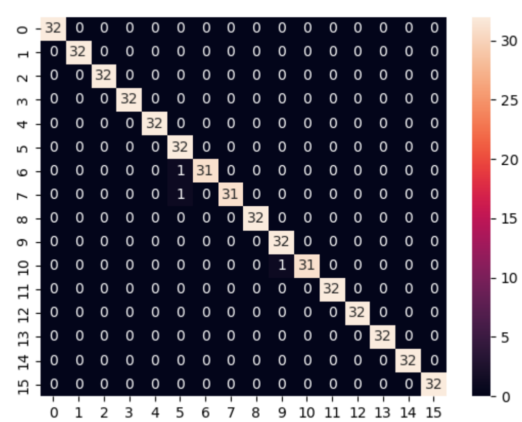


Fig. 4. Confusion Matrix with RMSprop.

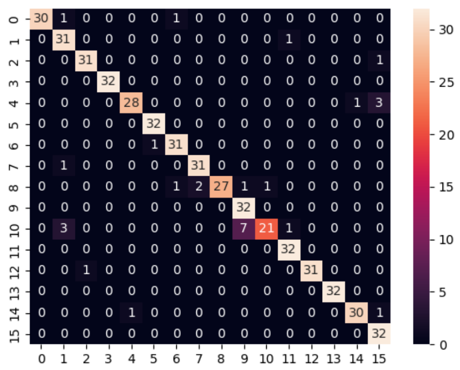


Fig. 5. Confusion Matrix with SDG.

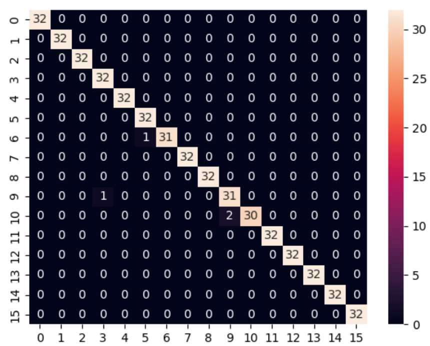


Fig. 6. Confusion Matrix with ADAM.

TABLE II. TRAINING AND TESTING ACCURACIES USING THE THREE OPTIMIZERS

Guninthalu	Training Accuracy			Testing Accuracy		
	Adam Optimizer	**RMSprop Optimizer**	**SGD Optimizer**	**Adam Optimizer**	**RMSprop Optimizer**	**SGD Optimizer**
	94.5	96	93.8	84.63	89	83.21
	94.55	96.02	94.1	84.82	89.05	83.36
	94.6	96.1	94.21	84.9	90.02	83.39
	94.2	96.06	82.21	84.7	89.5	83.29
	92.6	94.29	81.64	83.5	88.4	82.06
	92.42	94.2	80.64	82.4	87.21	80.62
	95	96.5	93.25	85	91.2	83.12
	93.3	93.4	91.6	83.22	87.4	81.62
	95.2	97	94.32	85.3	90.2	83.46
	93.3	94.9	91.52	83.22	87.4	81.44
	95.2	96.2	93.89	85.2	89.64	83.62
	95.72	96.26	94.12	85.63	90.05	83.78
	94.41	95.62	92.98	84.32	88.64	82.36
	94.64	95.71	93.02	84.41	88.73	82.54
	94.2	94.9	92.91	84.21	88.32	82.25
	94.3	95.2	93.61	84.33	88.42	82.35
	95.62	96.22	94.01	85.62	89.21	83.14
	95.63	96.34	94.31	85.72	89.32	83.28
	95.68	96.41	94.65	85.91	89.64	83.36
	96.22	96.76	94.91	86.45	89.84	84.31
	96.34	96.82	93.8	86.54	90.01	83.21

TABLE III. PRECISION USING THE THREE OPTIMIZERS

Guninthalu	Precision		
	Adam Optimizer	**RMSprop Optimizer**	**SGD Optimizer**
	0.835	0.895	0.821
	0.861	0.902	0.846
	0.87	0.912	0.851
	0.84	0.906	0.83
	0.828	0.892	0.819
	0.824	0.872	0.808
	0.841	0.91	0.824
	0.83	0.875	0.811
	0.862	0.882	0.842
	8.83	0.875	0.85
	0.842	0.912	0.83
	0.876	0.923	0.852
	0.824	0.884	0.806
	0.856	0.891	0.841
	0.826	0.872	0.802
	0.831	0.884	0.811
	0.846	0.916	0.831
	0.859	0.922	0.842
	0.869	0.934	0.851
	0.876	0.954	0.859
	0.885	0.961	0.821

TABLE IV. RECALL USING THE THREE OPTIMIZERS

Guninthalu	Recall		
	Adam Optimizer	RMSprop Optimizer	SGD Optimizer
	0.812	0.89	0.795
	0.836	0.893	0.821
	0.842	0.897	0.837
	0.821	0.894	0.811
	0.819	0.881	0.809
	0.804	0.864	0.782
	0.826	0.916	0.812
	0.806	0.864	0.783
	0.829	0.871	0.816
	0.806	0.864	0.772
	0.824	0.912	0.804
	0.841	0.925	0.815
	0.808	0.884	0.794
	0.824	0.89	0.813
	0.824	0.884	0.818
	0.831	0.889	0.824
	0.836	0.91	0.827
	0.846	0.921	0.831
	0.854	0.933	0.842
	0.862	0.941	0.851
	0.868	0.948	0.795

TABLE V. LOSS USING THE THREE OPTIMIZERS

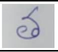
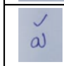
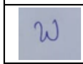
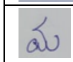
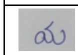
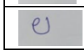
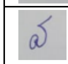
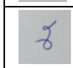
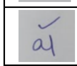
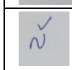
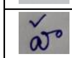
Guninthalu	Loss		
	Adam Optimizer	RMSprop Optimizer	SGD Optimizer
	0.63	0.39	0.68
	0.68	0.41	0.71
	0.71	0.5	0.74
	0.64	0.44	0.67
	0.6	0.4	0.62
	0.54	38	0.56
	0.7	0.47	0.72
	0.59	0.39	0.61
	0.68	0.46	0.71
	0.59	0.39	0.62
	0.65	0.41	0.68
	0.71	0.45	0.74
	0.59	0.37	0.61
	0.61	0.42	0.64
	0.58	0.36	0.6
	0.6	0.38	0.62
	0.65	0.41	0.67
	0.71	0.42	0.74
	0.75	0.44	0.77
	0.78	0.47	0.81
	0.79	0.51	0.68

TABLE VI. TOTAL PRECISION, RECALL, ACCURACY AND F1 SCORE WITH ADAM, RMSPROP AND SDG OPTIMIZERS

Optimizer	Total Precision	Total Recall	Accuracy	F1 Score
ADAM	0.9924	0.9921	99.21	0.9923
RMSprop	0.9944	0.9941	99.41	0.9942
SGD	0.948	0.943	94.33	0.945

TABLE VII. DETAILS OF VARIOUS GUNINTHALU OF PROPOSED DATABASE

Hand Written Characters	Train-Test Split	No of Train Samples	No of Test Samples
	80-20	220416	55104
	80-20	220416	55104
	80-20	220416	55104
	80-20	220416	55104
	80-20	220416	55104
	80-20	220416	55104
	80-20	220416	55104

	80-20	220416	55104
	80-20	220416	55104
	80-20	220416	55104
	80-20	220416	55104
	80-20	220416	55104
	80-20	220416	55104
	80-20	220416	55104
	80-20	220416	55104
	80-20	220416	55104
	80-20	220416	55104
	80-20	220416	55104
	80-20	220416	55104
	80-20	220416	55104
	80-20	220416	55104

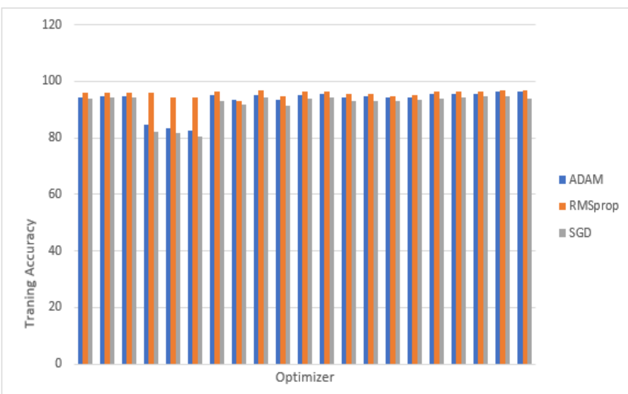


Fig. 7. Training Accuracy with ADAM, RMSprop & SGD Optimizer on 21 Guninthalu.

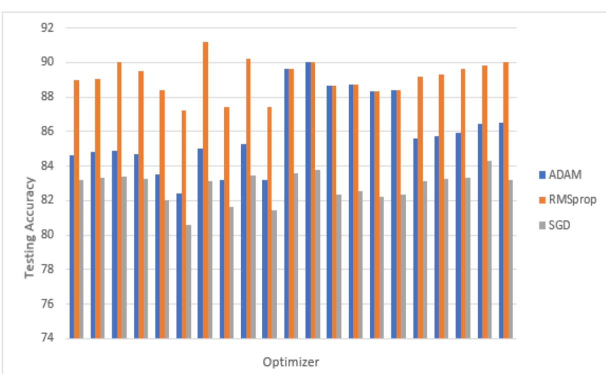


Fig. 8. Testing Accuracy with ADAM, RMSprop & SGD Optimizer on 21 Guninthalu.

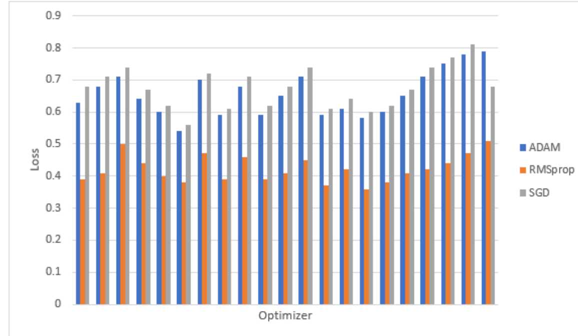


Fig. 9. Loss with ADAM, RMSprop & SGD Optimizer on 21 Guninthalu.

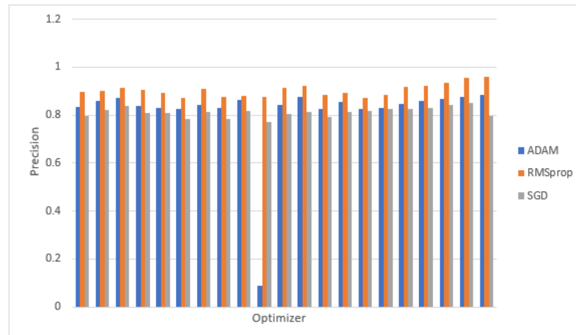


Fig. 10. Precision with ADAM, RMSprop & SGD Optimizer on 21 Guninthalu.

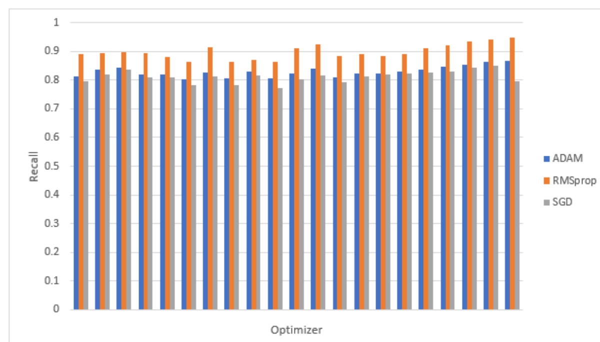


Fig. 11. Recall with ADAM, RMSprop & SGD Optimizer on 21 Guninthalu.

The training, testing accuracy, loss, precision, and recall with Adam optimizer, RMSprop optimizer, and SDG optimizer are shown in Fig. 7 to 11. The results show that the RMSprop optimizer has a higher testing accuracy than the ADAM and SGD optimizers. The mean values acquired with all the guninthalu are used to determine the overall training, testing accuracy, precision, and recall. Tables III to VI shows that when compared to ADAM and SGD optimizers, the RMSprop optimizer performs well on all guninthalu.

There are 820 different handwritten characters for each gunintham used to train the model. Each of the 21 guninthalu contains 16 characters, for a total of 21X16X820 = 275520 handwritten characters, which are used to train the model as shown in Table VII. After training, the model is checked using a 20% data mean and 55104 characters. Accuracy, Precision, Loss, and Recall are used to evaluate the model after it has been evaluated on 55104 data points.

B. Augmentation

Data augmentation is a method of artificially creating fresh training data from existing data. This is performed by transforming examples from the training data into fresh and unique training examples using domain-specific techniques.

C. Dropout

Dropout is a technique for preventing overfitting in a model. Dropout operates by setting the outgoing edges of hidden units at random. As the augmentation factor is increased, the validation accuracy improves. The test accuracy increased to 90.12 with ADAM optimizer, 94.26 with RMSprop, and 86.72 with SGD optimizer after data augmentation as shown in Fig. 12 to 14.

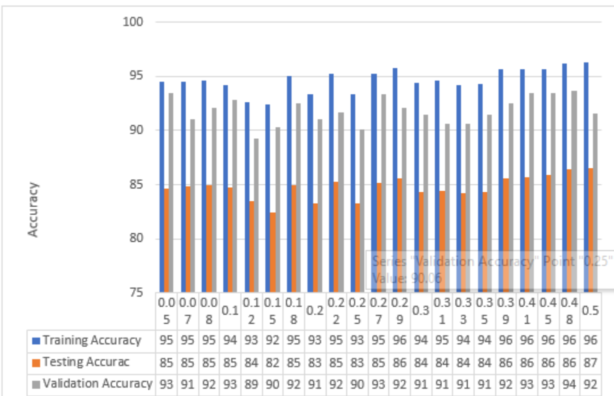


Fig. 12. Dropout in Layers with ADAM Optimizer.

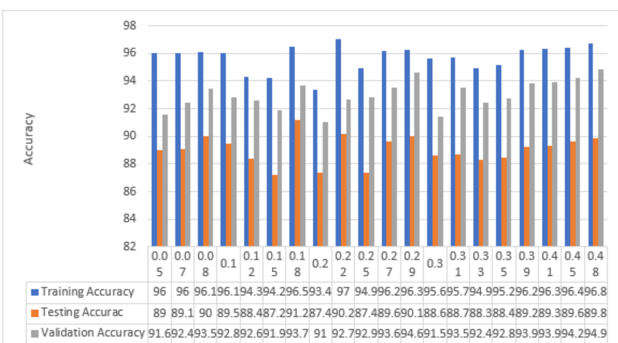


Fig. 13. Dropout in Layers with RMSprop Optimizer.

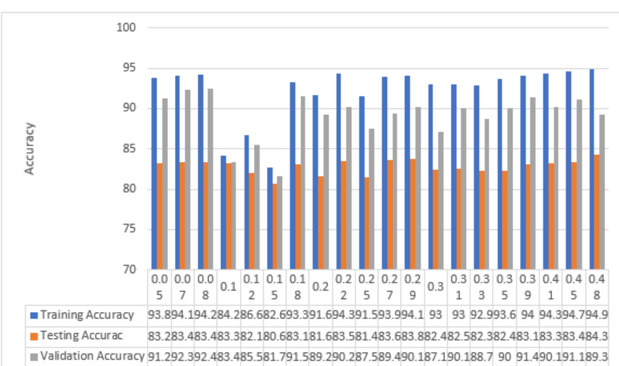


Fig. 14. Dropout in Layers with SGD Optimizer.

VI. Conclusion

With CNN, our goal is to improve the quality of handwritten Telugu gunintham identification. The contributions of Adam, RMSprop, and SGD were critical in improving the model's accuracy and performance. The dropout and activation functions ReLu and Sigmoid are used to prevent overfitting. In CNN models, RMSprop optimizer outperformed Adam and SGD in terms of accuracy. This model might be adjusted to enhance the identification of handwritten Telugu characters. In the future, we want to continue experimenting with new approaches to improve the identification of Telugu handwritten guninthalu.

References

[1] Kou, F., Du, J., He, Y., & Ye, L. "Social Network Search Based on Semantic Analysis and Learning." CAAI Transactions on Intelligence Technology. 1 (2016) 293-302.

[2] Srinivas, S., Sarvadevabhatla, R. K., Mopuri, K. R., Prabhu, N., Kruthiventi, S. S., &Babu, R. V. (2016) "A taxonomy of deep convolutional neural nets for computer vision." https://doi.org/10.3389/frobt.2015.00036.

[3] Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., &Torralba, A. (2014) "Object detectors emerge in deep scene cnns." https://arxiv.org/abs/1412.6856.

[4] Wang, Y., & Wu, Y. "Scene Classification with Deep Convolutional Neural Networks." (2014).

[5] Artha Andriyanto, Antoni Wibowo NS Norhaslinda Zainal Abidin "Sectoral Stock Prediction Using Convolutional Neural Networks with Candlestick Patterns as input Images" International Journal of Emerging Trends in Engineering Research 8(2020) 2249-2252.

[6] P.V. Ramana Murthy, Ch. G.V.N. Prasad "Recognition of Online Handwritten Telugu Letters For Different Domains And Organizations" Journal of Critical Reviews 6 (2019.) 33-40.

[7] P. Sujatha, D. Lalitha Bhaskari "Telugu and Hindi Script Recognition using Deep learning Techniques" International Journal of Innovative Technology and Exploring Engineering (IJITEE) 8 (2019) 1758-1764.

[8] Buddaraju Revathi, G.Naveen Kishore, V Dheera "A Survey On OCR For Telugu Language" International Journal Of Scientific & Technology Research 8 (2019) 559-562.

[9] B. Hari Kumar, P. Chitra "Survey Paper Of Script Identification Of Telugu Language Using OCR" International Journal of Electronics and Communication Engineering (IJECE) 8 (2019) 15-20.

[10] Chandra Prakash Konkimalla et al., Optical Character Recognition (OCR) for Telugu: Database, Algorithm and Application arXiv:1711.07245 (2018).

[11] Chirag I Patel, Ripal Patel, Palak Patel "Handwritten Character Recognition using Neural Network" International Journal of Scientific & Engineering Research 2 (2011) 1-6.

[12] A Ram Bharadwaj, A. Venugopal, Ch. Surya Kiran, M. V. Nageswara Rao "Telugu text extraction and recognition using convolutional and recurrent neural networks" International Journal of Engineering and Advanced Technology (IJEAT) 8 (2019) 1449-1451.

[13] N. Shaffi and F. Hajamohideen, "uTHCD: A New Benchmarking for Tamil Handwritten OCR," in IEEE Access, vol. 9, pp. 101469-101493, 2021, doi: 10.1109/ACCESS.2021.3096823.

[14] Richard HR Hahnloser, Rahul Sarpeshkar, Misha A Mahowald, Rodney J Douglas,and H Sebastian Seung "Digital selection and analogue amplification coexistin a cortex-inspired silicon circuit" Nature 405 (2000), 947.

[15] Diederik P. Kingma, Jimmy Ba, "Adam: A Method for Stochastic Optimization", arXiv:1412.6980.

[16] Praveen Kumar Kollu, R. Satya Prasad "Intrusion Detection System Using Recurrent Neural Networks and Attention Mechanism", International Journal of Emerging Trends in Engineering Research 7(2019) 178-182.

Information Security Enhancement by Increasing Randomness of Stream Ciphers in GSM

Ram Prakash Prajapat[1]
SDE, BSNL
Jodhpur, Rajasthan
India

Dr. Rajesh Bhadada[2]
Professor and Head
MBM Engg. College
Jodhpur, Rajasthan, India

Arjun Choudhary[3]
Centre for Cyber Security
Sardar Patel University of Police
Jodhpur, India

***Abstract*—Information security is a crucial issue and needs to be addressed efficiently. Encryption of the original information is used to ensure privacy during exchange of information. In GSM (Global System for Mobile) standard, once the voice traffic initiates after signaling, encryption comes into the picture to ensure privacy during the call, after authentication. Here in this process, the plaintext is encrypted in to cipher-text using stream ciphers. For stronger security, strong ciphers with strong randomness are required. Linear Feedback Shift Registers (LFSRs) based A5 algorithm family is used for encryption in GSM. There are many shortcomings of this cipher and with these, privacy can't be assured. Some ways are proposed in this paper to ensure better security by enhancing the randomness of the generated bit stream being used for encryption. These are incorporation of user's current location, reuse of already generated 32 bit SRES during authentication process and conversion of linear FSRs into nonlinear FSRs. Statistical Test Suite NIST is used to test the various properties of random bit stream and an attempt has been made to achieve better randomness, hence more security.**

***Keywords*—Security; encryption; A5/1 stream cipher; randomness; NIST test suite**

I. Introduction

Although with the advancement of technology, many vulnerabilities of security threats of GSM have been addressed in EDGE, 3G (HPA/HSDPA) & 4G (LTE) but this is still relevant as large number of people use GSM specially in rural areas. In addition to this, stream ciphers are used in many other wireless applications like modems / routers, smart appliances & security devices. Few algorithms like A3, A8 & A5 are used in GSM for authentication & encryption process over A_{bis} air interface between user mobile (MS) and base station (BS). The details of inputs, outputs & use of these algorithms are described in Table I. Here the SRES is Signed Response of 32 bits, K_c is a 64 bits Cipher Key and RAND is a 128 bits random number. Fig. 1 shows that the combination of K_c, RAND and SRES is called "Triplet".

A5 is mainly responsible for encryption as shown in Fig. 2. Here by using Cipher Key K_c of 64 bits along with TDMA Frame number F_n, a 228 bits pseudo random number PRAN is generated which is XORed with 228 bits plain-text in bit-by-bit manner to get cipher-text. This cipher form of information after encryption is finally transmitted over the air interface between the user mobile station & base station.

Two stage security, i.e. "Authentication" & "Encryption" is implemented in GSM. Initially, the access of the network resources is granted to any new or existing subscriber on its request after authentication process on every location update.

During this "Authentication Process", the core network challenges MS and in response to this, MS sends SRES. This is matched with the SRES available with itself and grants the access on matching only as described in Fig. 3. After getting the access of the network, the encryption process takes place to ensure the privacy during the call. Here in this process, the plaintext is encrypted in to cipher-text using stream ciphers of A5/1 algorithm. In the same way, decryption occurs at the other end to reconstruct the original information.

As the transmission of information is bursty in nature in GSM, the 114 bit frame sequence in downlink (BS to MS) & the same way a 114 bit frame sequence in uplink (MS to BS) is transmitted every 4.6 milliseconds. K_c is produced and mixed with a publically known TDMA frame number Fn for each frame for every new voice call.

The configuration of A5 was never shared by the ETSI [1]. Although, it got reverse engineered & became available to all. Many crypto attacks occurred by cryptanalysts to reveal the internal structure of this algorithm and hackers & intruders managed to decrypt the transmitted information by users. Golic [2, 3] presented the functional design in the year 1994 and later Marc Briceno [4] revealed the entire design by reverse engineering in the year 1999 [15,16,17,19,20].

TABLE I. Algorithms used for Information Security in GSM

Algorithm	Inputs	Output	Purpose
A3	K_i (128 Bits)	SRES (32 Bits)	Authentication Process
	RAND (128 Bits)		
A8	K_i (128 Bits)	K_c (64 Bits)	Cipher Key Generation (K_c)
	RAND (128 Bits)		
A5	K_c (64 Bits)	Cipher Text (228 Bits)	Encryption Process (Voice & Data)
	TDMA Frame No. (22 Bits)		
	Plain Text (228 Bits)		

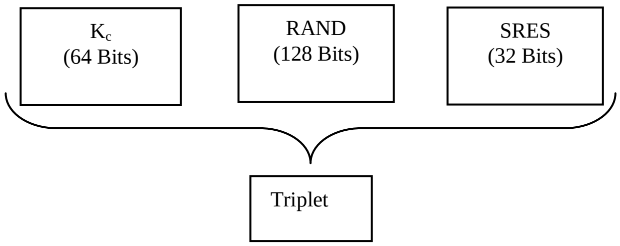


Fig. 1. Triplet Details.

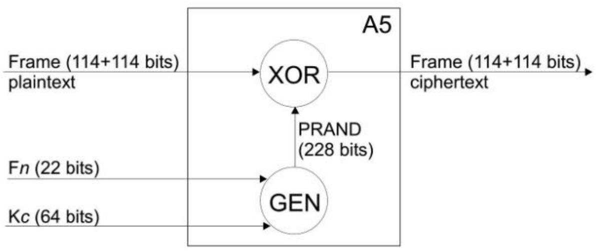


Fig. 2. Encryption by A5/1 in GSM.

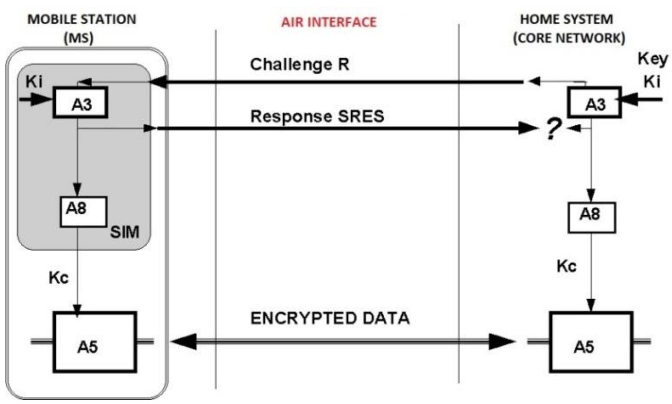


Fig. 3. Authentication and Encryption Process.

Biryukov, Wagner and Shamir [6], Ekdahl and Johansson [10], Dunkelman and Biham [5], Maximov & team [11] and Barkan [14] contributed lot. Contribution of Andrew Rukhin et al [8], N. Komninos [9], Basar Kasim [12, 13] is also remarkable. Very recently, Darshana Upadhyay, Priyanka Sharma & Sharada Valiveti [18], Sattar B. Sadkhan [21,24,26] & Nagendar Yerukala [27] made some modifications by changing of feedback taps in A5/1 stream cipher to improve the randomization property to make it robust to attacks [22,23].

Many cryptanalysts proved that due to weaknesses of this stream cipher, information security can be compromised in GSM [25, 28, 29]. These weaknesses are:

1) Weak Linear Complexity (LC).
2) Poor clocking system (Majority Function Rule).
3) Clocking period is too short.
4) Poor clocking taps selection.
5) Collision issue.

Using an improved clocking system with a combinational function of high correlation immunity and high algebraic degree, the security can be increased [6, 11].

Logisim simulator (primarily developed by Dr. Carl Burch) is used to realize the structure of the proposed A5/1 algorithm and its randomness parameters are analyzed by NIST Suite [7]. Brief description about the internal structure of A5/1 is given in Section II, modifications are proposed in Section III, observations and randomness analysis is given in Section IV and Section V concludes the results.

II. Internal Structure of A5/1

It has three Linear Feedback Shift Registers (LFSRs) of different bit lengths to generate a pseudo random binary stream. The total bit length of this cipher is 64 bits in which LFSR-1 (R1) has 19 bits, LFSR-2 (R2) has 22 bits, and LFSR-3 (R3) has 23 bits as depicted in Fig. 4.

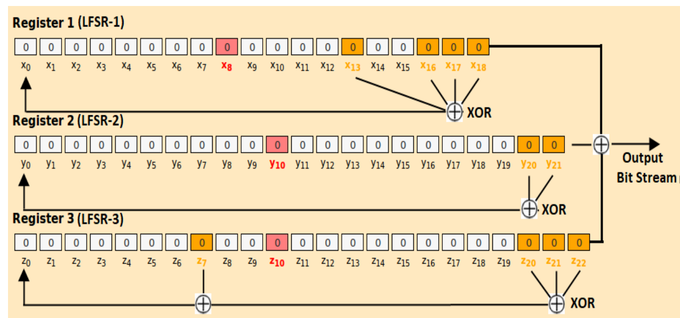


Fig. 4. Internal Design of A5/1.

Following its own feedback polynomial, periodic bit sequence is generated by these registers. The feedback bit positions are predefined at 13, 16, 17, 18 and 20, 21 and 7, 20, 21, 22 for R1, R2 and R3, respectively.

Similarly, the single clocking taps of these registers are also predefined at tap 8, 10, and 10 for R1, R2 and R3, respectively. Bit length, clocking bit, tap bits and primitive polynomial are shown in Table II.

TABLE II. Information Table for LFSRs

LFSR	Length (in bits)	Clocking Bit	Tap Bits	Primitive Feedback Polynomials
R1	19	8	13, 16, 17, 18	$x^{19}+x^{18}+x^{17}+x^{14}+1$
R2	22	10	20, 21	$x^{22}+x^{21}+1$
R3	23	10	7, 20, 21, 22	$x^{23}+x^{22}+x^{21}+x^{8}+1$

TABLE III. Majority Rule Truth Table

Clocking Bit			Majority Function	Clocked LFSR		
R1	R2	R3		R1	R2	R3
0	0	0	0	Yes	Yes	Yes
0	0	1	0	Yes	Yes	No
0	1	0	0	Yes	No	Yes
0	1	1	1	No	Yes	Yes
1	0	0	0	No	Yes	Yes
1	0	1	1	Yes	No	Yes
1	1	0	1	Yes	Yes	No
1	1	1	1	Yes	Yes	Yes

The clocking mechanism of each register is decided by Majority Rule as shown in the truth table Table III below. For each cycle, only those registers will be clocked and updated whose clocking bit values have majority. The majority value m is decided by m = *maj* (C1,C2,C3), here C1, C2 and C3 are the clocking bits of all three registers.

In simple words, "at least two out of three" is the majority rule i.e. the majority among these bits. If two or more clocking bits are 1, the majority value m will be 1, and similarly if two or more are 0, the majority value m will be 0.

Thus, in this mechanism, two or more, whose clocking bit is the equal to m, will be clocked at each clock cycle. Every register has the clocking probability of 0.75 and non-clocking probability of 0.25. The majority rule function can be realized using logic gates as shown below in Fig. 5 (Logisim simulator).

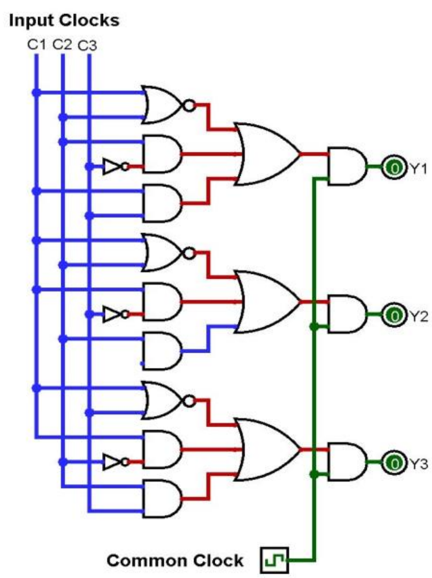


Fig. 5. Realization Majority Function in Logisim.

All three registers are reset by setting a zero value. In next 86 cycles, K_c and Fn are loaded bit by bit with regular clocking. The output ignored during initial stage of the first 100 clock cycles and during this period the irregular clocking continues for all three LFSRs as per the majority rule. Now the required random bit stream of 228 bits is obtained for 228 clock cycles [28]. The same steps are repeated for the next frame.

III. Proposed Cipher

To overcome some of the problems mentioned above in previous section, the following schemes / modifications are proposed to increase the randomness:

1) MOD-I: Here in this scheme as shown in Fig. 6, the nonlinearity is introduced in the feedback path of the shift registers by adding universal gates. Thus, the LFSRs have been converted into NLFSRs. By this, the randomness of the generated bit stream will be improved.

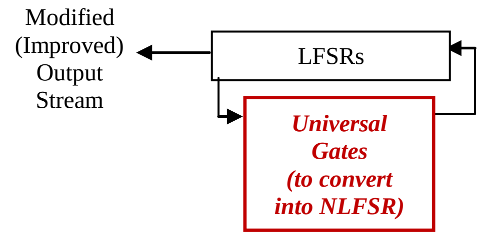


Fig. 6. Proposed MOD-I Scheme.

Polynomial equations of this MOD-I scheme will have the impact of NAND and NOR logic gates in it.

2) MOD-II: Here in this scheme, the 32 bit SRES is reused in the feedback path of the shift registers, which is already generated during the authentication process by A3 algorithm, as shown in Fig. 7. The SRES is XOR'ed in the feedback unit of LFSR through a NAND gate on bit by bit basis. This scheme reuses the output of another algorithm, hence increases the randomness of the cipher key.

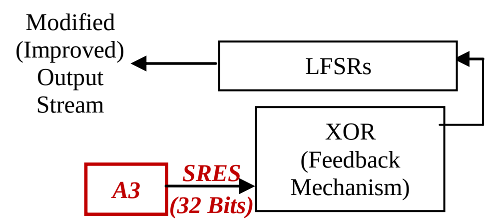


Fig. 7. Proposed MOD-II Scheme.

3) MOD-III: The user location is mixed with the bit stream generated using XOR (only last 32 bits). This works as a key feature as generally users have different locations and intruders cannot crack it easily. This is very important proposal, because the location of each individual user is not known to intruders and many times dynamic in nature. The CGI changes with the movement of the user and this makes the bit stream more complex. The idea is shown in Fig. 8.

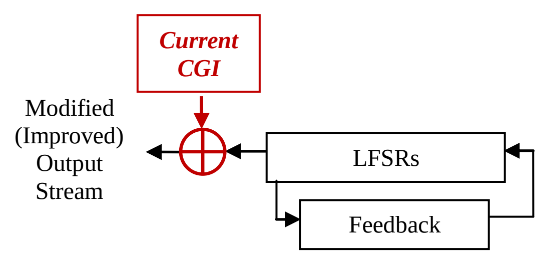


Fig. 8. Proposed MOD-III Scheme.

4) MOD-C: In this, all the above three modifications are combined to get the simultaneous impact of all above modifications.

The idea of combining all three modifications is shown in Fig. 9.

The proposed A5/1 cipher is realized and simulated in Logisim as shown in Fig. 10.

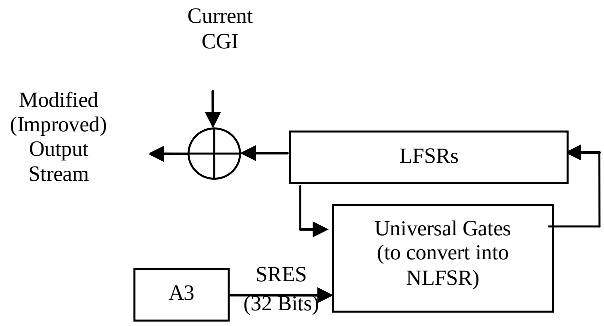


Fig. 9. Proposed Scheme of MOD-C.

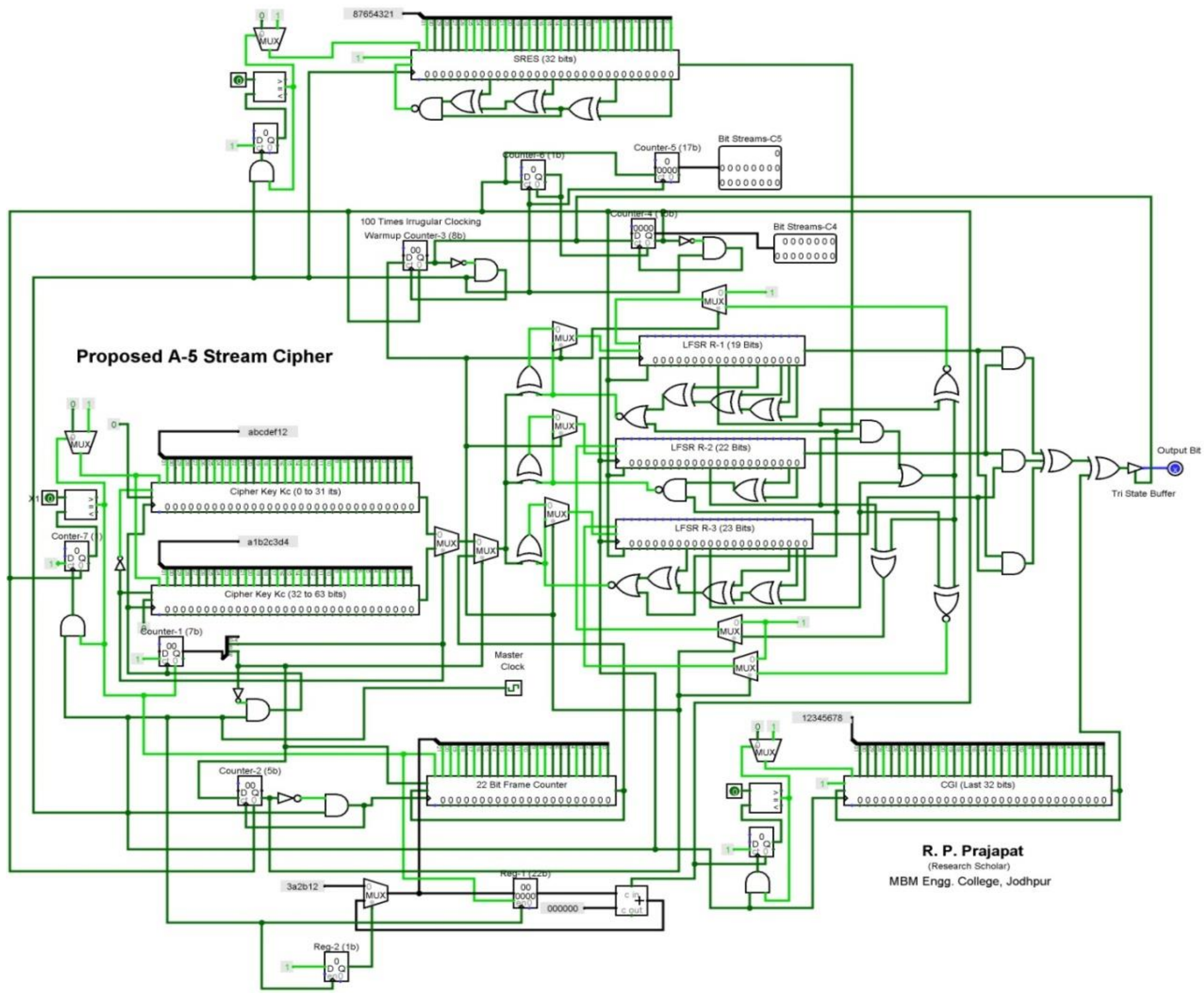


Fig. 10. Proposed Scheme of MOD-C in Logisim.

IV. Observations and Randomness Analysis

In a broader sense, the randomness of a data set is the lack of predictability, i.e. more uncertainty or more entropy. If a data set has more randomness means it has more encryption capability, hence more security. Such tests are carried out to check recognizable or repetitive patterns in any data set under test. Randomness is related to the theory of information entropy, probability, and chance. Entropy is a measuring tool for randomness.

NIST Suite is a statistical test suite based on Linux operating system and is used for statistical parameters testing of the output bit stream of both the actual and proposed scheme of A5/1 cipher. This test suite is also used in various cryptographic applications [7]. Performance comparison has also been done between these two with respect to various parameters. Following are the main tests which were carried out for the randomness test:

1) The frequency test (monobit & within a block): It tests the balance of 0's and 1's in the bit stream. The equilibrium between 0's and 1's should be maintained for a perfectly random data set. Therefore, the probability of availability of 0's and 1's should be close to 0.5 [7].

2) The cumulative sums test (cusums): It tests the randomness of a sequence of 0's and 1's which are called "random walks" or "partial sequences". It tells that the sum of the partial sequences is too large or too small [7].

3) The runs test: It analyzes the occurrence of similar patterns that are separated by different patterns [7].

4) The DFT (spectral) test: It finds out the patterns which are periodic in nature in a random bit sequence. The repetitive or periodic patterns close to each other are detected in this test [7].

5) The serial test: It detects the pair or patterns like 00, 01, 10, 11, 100 & 101 etc and checks the balances with its complimentary pair/pattern [7].

6) The linear complexity test: It is directly related to the bit length of the LFSRs used to generate the random bit stream [7].

The P-value parameter is defined for all these tests in NIST suite. It shows the probability of a bit stream of being random in nature. For an ideal random bit set, this P-value is 1 and if it is 0, then the bit stream is completely nonrandom. Thus, a higher value or close to 1 is desirable.

Different sizes of data (up to 10,000 blocks of 114 bit i.e. 10,000 x 114 =10, 00,000 consecutive bits) are used during the statistical tests both for the actual and proposed cipher [28]. The observations of different tests conducted upon the generated bit stream are as follows:

As the LFSRs of the actual cipher have been converted into NLFSRs in MOD-I scheme, we see a slight increase in the P-values of various tests. The 32 bit SRES is reused in feedback path of the shift registers in MOD-II and again slight increment in the P-values of various tests. As the last 32 bits of current location (CGI) of the user incorporated in output bit stream the slight increment in the P-values of various tests can be observed I MOD-III. Because all proposed modifications are implemented here simultaneously in MOD-C, a huge increase in the P-values of various tests can be observed.

The details of observations of different sizes of the data set are also provided in Table V for this MOD-C scheme.

TABLE IV. OBSERVATIONS OF THE PROPOSED SCHEMES

Test Parameter	**Actual A5/1** ***(P-Value)***	**Proposed MOD-I Scheme** ***(P-Value)***	**Proposed MOD-II Scheme** ***(P-Value)***	**Proposed MOD-III Scheme** ***(P-Value)***	**Proposed MOD-C Scheme** ***(P-Value)***
Frequency	0.46	0.55	0.58	0.57	0.77
Block Frequency	0.82	0.87	0.87	0.86	0.92
Cumulative Sum	0.44	0.61	0.62	0.64	0.80
Runs	0.90	0.91	0.92	0.91	0.94
Spectral DFT	0.95	0.96	0.95	0.96	0.98
Serial	0.96	0.97	0.96	0.97	0.98
Linear Complexity	0.79	0.82	0.81	0.82	0.88

TABLE V. STATISTICAL TEST RESULTS BY NIST TEST SUITE (A : ACTUAL A5/1 & P : PROPOSED A5/1)

MOD-C Scheme															
Bit Stream	Bits	**Frequency**		**Block Frequency**		**Cumulative Sum**		**Runs**		**DFT**		**Serial**		**Linear Complexity**	
		A	**P**	**A**	**P**	**A**	**P**	**A**	**P**	**A**	**P**	**A**	**P**	**A**	**P**
100	114	0.43	0.76	0.80	0.89	0.42	0.77	0.83	0.89	0.92	0.96	0.91	0.95	0.73	0.84
500	114	0.47	0.79	0.79	0.86	0.41	0.76	0.82	0.88	0.93	0.96	0.92	0.96	0.74	0.85
1000	114	0.49	0.81	0.82	0.91	0.43	0.78	0.83	0.89	0.92	0.97	0.95	0.98	0.75	0.86
5000	114	0.48	0.80	0.80	0.90	0.43	0.79	0.85	0.91	0.93	0.98	0.93	0.96	0.76	0.87
8000	114	0.43	0.78	0.81	0.90	0.41	0.77	0.89	0.93	0.94	0.98	0.96	0.98	0.76	0.88
10000	114	0.46	0.77	0.82	0.92	0.44	0.80	0.90	0.94	0.95	0.98	0.96	0.98	0.79	0.88
Average		0.46	0.79	0.81	0.90	0.42	0.78	0.85	0.91	0.93	0.97	0.94	0.97	0.76	0.86

V. Conclusion

The improvements in P-values of various tests of all four modifications (MOD-I, MOD-II, MOD-III & MOD-C) have been described in different tables of previous section. Based on these test results, it is stated that there is slight improvement in the randomness of generated bit stream of all three MODs but when all MODs are combined simultaneously in MOD-C a huge improvement can be observed in the P-values of Cumulative Sum test, Frequency test and most importantly in Linear Complexity test. There is a good balance between 0's and 1's (results of Frequency test), better random walks (results of Cumulative test), less occurrence of similar patterns and that too are well separated by different patterns (results of Run tests), low periodic patterns (results of DFT test) and enhanced entropy (results of Linear Complexity test) in random bit stream generated by proposed cipher.

An effort is made in this paper to improve the randomness by making three modifications MOD-I, II, & III and then comparing the NIST test results. These modifications are the incorporation of nonlinearity I feedback path of LFSRs, reusing the SRES of A3, and inclusion of current CGI of the user, respectively to improve the entropy. After that, all three modifications are implemented simultaneously in a combined manner to achieve better results. These are the major improvements and contributions in the proposed cipher scheme.

The weakness issues mentioned in early part of this paper are addressed significantly by these proposed schemes. Simulation & testing results confirmed it.

The test results of MOD-C scheme for different data sizes as shown in Table V and depicted in graphical form in Fig. 11 above. The P-values of all the tests of proposed A5/1 have been increased by a significant value in comparison with the original A5/1 cipher. For better randomness, higher P-values are desired and better randomness means stronger encryption and enhanced security. Therefore, as per the observations and test results, it is concluded that the proposed scheme of cipher is having better security against the cryptographic attacks with respect to the actual A5/1 cipher due to increased randomness (at the cost of slight increment in hardware).

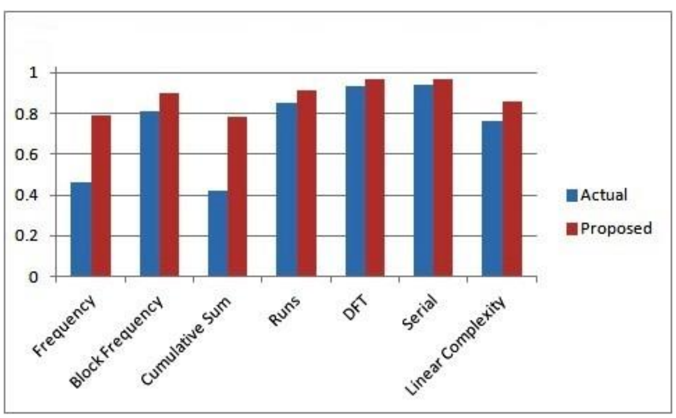


Fig. 11. Comparative Analysis of Test results of MOD-C Scheme.

References

[1] GSM Recommendations 02.09, Security Aspects by European Telecommunications Standards Institute (ETSI), June 1993.

[2] Jovan Dj. Golić, "On the security of shift register based key stream generators," Fast Software Encryption-Cambridge' 93, Lecture Notes in Computer Science, vol. 809, R. J. Anderson ed., Springer-Verlag, pp. 90–100, 1994.

[3] Jovan Dj. Golić, "Cryptanalysis of alleged A5 stream cipher," Advances in Cryptology, proceedings of EUROCRYPT'97, LNCS, vol. 1233, pp.239–255, Springer-Verlag, 1997.

[4] M. Briceno, I. Goldberg, D. Wagner, A pedagogical implementation of the GSM A5/1 and A5/2 voice privacy encryption algorithms, (1999), http://cryptome.org/gsm-a512.htm, originally on www.scard.org.

[5] E. Biham, and O. Dunkelman, Cryptanalysis of the A5/1 GSM stream cipher, Progress in Cryptology, proceedings of INDOCRYPT'00, LNCS, pp. 43–51, Springer-Verlag, 2000.

[6] A. Biryukov, A. Shamir, and D. Wagner, Real time cryptanalysis of A5/1 on a PC, Advances in Cryptology, proceedings of Fast Software Encryption'00, LNCS, pp.1–18, Springer-Verlag, 2001.

[7] C. Burch, "Logisim : A Graphical System for Logic Circuit Design and Simulation," J. Educational Resources in Comput., vol. 2(1), pp. 5-16, 2002.

[8] Andrew Rukhin et al, NIST, A Statistical Test Suit for random and pseudorandom number generators for cryptographic applications. NIST Special Publication 800-22, with revisions dated May 15, 2001.

[9] N. Komninos, B. Honary & M. Darnel1, "Security enhancements for A5/1 without loosing hardware efficiency in future", 3G Mobile Communication Technologies Conference, 2002, Page No. 324-328.

[10] P. Ekdahl, and T. Johansson, Another attack on A5/1, IEEE Transactions on Information Theory, vol. 49, pp. 284-289, 2003.

[11] A. Maximov, T. Johansson, and S. Babbage, An improved correlation attack on A5/1, proceedings of SAC 2004, LNCS, vol.3357, pp.1–18, Springer-Verlag, 2005.

[12] Basar Kasim & Levent Ertaul, "GSM Security", International Conference on Wireless Networks (ICWN), 2005, pp. 555-561.

[13] I. Erguler and E. Anarim, A modified stream generator for the GSM encryption algorithms A5/1 and A5/2, 13th European Signal Processing Conference (EUSIPCO'05), September, 2005.

[14] E. Barkan, and E. Biham, Conditional estimators: an effective attack on A5/1, proceedings of SAC 2005, LNCS, vol. 3897, pp. 1-19, Springer-Verlag, 2006.

[15] Elad Barkan, Eli Biham & Nathan Keller, "Instant Ciphertext-Only Cryptanalysis of GSM Encrypted Communication", Technion - Computer Science Department - Technical Report CS-2006-07 - 2006.

[16] Sukalyan Goswami, Subarna Laha, Satarupa Chakraborty & Ankana Dhar, "Enhancement of GSM Security Using Elliptic Curve Cryptography Algorithm" 3rd Int. Conference on Intelligent Systems Modelling and Simulation, IEEE, 2012 Page No. : 639-644.

[17] Mahdi Daghmechi Firoozjaei & Javad Vahidi, "Implementing Geo-encryption in GSM Cellular Network", IEEE, 2012, Page No. 299-302.

[18] Prof. Darshana Upadhyay et al, Randomness analysis of A5/1 Stream Cipher for secure mobile communication, IJCSC Vol. 5, pp. 95-100, Sept. 2014.

[19] Pankaj et el., "Design of Enhanced Pseudo-Random Sequence Generator usable in GSM Communication", IEEE WiSPNET 2016 conference.

[20] Nibras Hadi Jawad, "Simulation and Developed A5/3", International Conference on Current Research in Computer Science and Information Technology (ICCIT), Slemani - Iraq, 2017.

[21] Sattar B. Sadkhan et al., "A DNA-Sticker Algorithm for Cryptanalysis LFSRs and NLFSRs Based Stream Cipher" International Conference on Advanced Science and Engineering (ICOASE), Iraq, 2018.

[22] Martin Jurecek, Jirí Bucek, and Róbert Lórencz, "Side-Channel Attack on the A5/1 Stream Cipher" 2019 22nd IEEE Euromicro Conference on Digital System Design (DSD), 2019.

[23] Sattar B. Sadkhan et al., " Proposed enhancement of A5/1 stream cipher" 2nd International Conference on Engineering Technology and their Applications 2019-IICET2019-IRAQ, 2019.

[24] Sattar B. Sadkhan & Zainab Hamza, Proposed Enhancement of A5/1 stream cipher, 2019 2nd International Conference on Engineering Technology and its Applications (IICETA).

[25] Martin Jurecek, Jirí Bucek and Róbert Lórencz, Side-Channel Attack on the A5/1 Stream Cipher, 2019 22nd Euromicro Conference on Digital System Design (DSD), 2019.

[26] Sattar B. Sadkhan, "A proposed Development of Clock Control Stream Cipher based on Suitable Attack" 2020 1st. Information Technology To Enhance e-learning and Other Application (IT-ELA, 2020.

[27] Nagendar Yerukala, V Kamakshi Prasad, and Allam Apparao, "Performance and Statistical Analysis of Stream Ciphers in GSM Communications" Journal of Communications Software and Systems, Vol. 16, No. 1, March 2020.

[28] Ram Prakash Prajapat, Rajesh Bhadada and Giriraj Sharma, Security Enhancement of A5/1 Stream Cipher in GSM Communication & its Randomness Analysis" 2021 IEEE 6th International Forum on Research and Technology for Society and Industry (RTSI), 6-9 Sept. 2021.

[29] Yi Qian, Feng Ye and Hsiao-Hwa Chen, Cryptographic Techniques, Security in Wireless Communication Networks (SWCN), 2022.

Plant Disease Detection using AI based VGG-16 Model

Anwar Abdullah Alatawi[1], Shahd Maadi Alomani[2], Najd Ibrahim Alhawiti[3], Muhammad Ayaz[4]
Industrial Innovation and Robotic Center (IIRC), University of Tabuk, Tabuk 47731, Saudi Arabia[1, 2 3]
Faculty of Computers and Information Technology, University of Tabuk, Tabuk, Saudi Arabia[1, 2, 3]
Sensor Networks and Cellular Systems (SNCS), Research Center, University of Tabuk, Tabuk, 47431, Saudi Arabia[4]

***Abstract*—Agriculture and modern farming is one of the fields where IoT and automation can have a great impact. Maintaining healthy plants and monitoring their environment in order to identify or detect diseases is essential in order to maintain a maximum crop yield. The implementation of current high rocketing technologies including artificial intelligence (AI), machine learning, and deep learning has proved to be extremely important in modern agriculture as a method of advanced image analysis domain. Artificial intelligence adds time efficiency and the possibility of identifying plant diseases, in addition to monitoring and controlling the environmental conditions in farms. Several studies showed that machine learning and deep learning technologies can detect plant diseases upon analyzing plant leaves with great accuracy and sensitivity. In this study, considering the worth of machine learning for disease detection, we present a convolutional neural network VGG-16 model to detect plant diseases, to allow farmers to make timely actions with respect to treatment without further delay. To carry this out, 19 different classes of plants diseases were chosen, where 15,915 plant leaf images (both diseased and healthy leaves) were acquired from the Plant Village dataset for training and testing. Based on the experimental results, the proposed model is able to achieve an accuracy of about 95.2% with the testing loss being only 0.4418. The proposed model provides a clear direction toward a deep learning-based plant disease detection to apply on a large scale in future.**

Keywords*—*Machine learning; VGG-16; disease detection; convolutional networks; Plant Village; modern farming

I. INTRODUCTION

Agriculture has always been a basic human need ever since humans' existence as plants were a primary source of food. Even nowadays, agriculture is still considered an essential food resource and is the center of several aspects in humans' lives [1]. As a matter of fact, agriculture serves as the pillar of economy in many countries regardless of their developmental stages. The various domains that show the importance of agriculture include the fact that agriculture is a main source of livelihood where approximately 70% of the population depends on plants and their cultivation for livelihood. This great percentage reflects on agriculture being the most important resource that can actually stand a chance in the face of the rapidly increasing population [2].

One of the most critical challenges that face agriculture and affects it trade is plant diseases and how to timely detect them and deal with them to improve the health of crops. By definition, plant disease in a type of natural problems that occur in plants affecting their overall growth and might lead to plant death in extreme cases. Plant diseases can occur throughout the different stages of plant development including seed development, seedling, and seedling growth [3]. When diseased, plants go through different mechanical, morphological, and biochemical changes [4, 5]. Truthfully, there are two main types of plant stress classified as biotic stress represented by living creatures that interact with plants in a way that negatively affects their growth [6] such as bacteria, viruses, or fungi [7], or abiotic stress represented by the collection of non-living factors or the environmental factors. Fig. 1 illustrates the collection of factors that contribute to plant diseases.

Typically, the commonly used approach for farmers, scientists, and even breeders, to detect and identify plant disease was the manual inspection of plants. Of course, this process requires expertise and knowledge for the proper detection. With time, manual inspection became tiresome and time consuming and not as quite efficient especially when large amounts of plants needed to be inspected. Another factor that proves the inefficiency of manual inspection is the similar conditions that might be caused by different pathogens that might look alike in their effect on the plant [6].

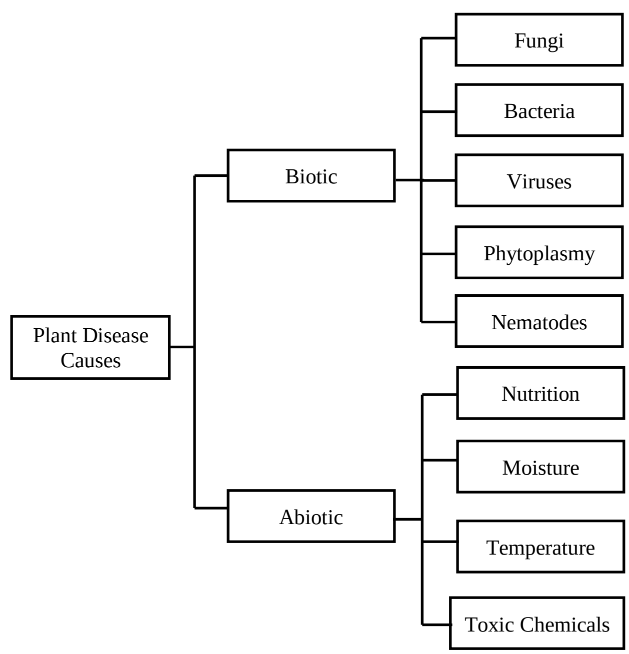


Fig. 1. Plant Disease Causes Detailed as Biotic and Abiotic Factors.

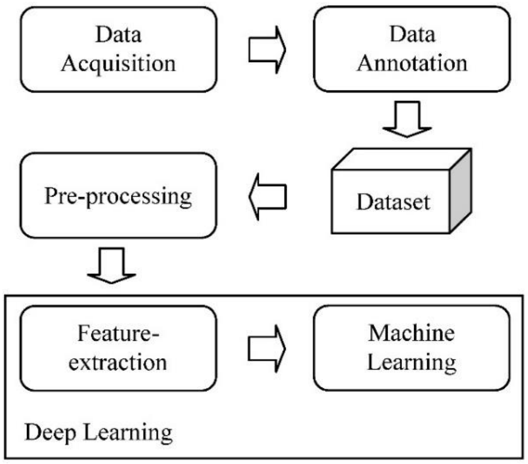


Fig. 2. Steps in the Implementation of Machine Learning Models in Plant Disease Detection.

For this reason, humans needed a better suited technique that can deliver effective plant detection results in less time.

Technology and the rise of both deep and machine learning came as a benefit in various fields, most importantly medicine. Nonetheless, advanced technological approaches can also be implemented for the purpose of detecting diseases in plants. Thus, machine learning and deep learning approaches can be considered non-destructive disease detection methods since they are based on image-processing techniques, as opposed to destructive serology or molecular methods [8]. However, for these techniques to work, the disease must have already caused a visible change in the plant, particularly in the leaf area or stem [9].

Artificial intelligence, computer vision and machine learning utilizations can greatly enhance the process of plant disease detection, and is already applied in multiple research papers [10]. Such technologies are capable of not only detecting the presence of a disease, but it is also possible to determine its severity, and to classify exactly which kind of disease is present in a given plant sample [11].

Based on their depth, the plant disease detection methods can be divided into shallow architectures and deep architectures. Basic machine learning methods like Random Forest (RF), Support Vector Machine (SVM), Naïve Bayes (NB), and K-Nearest Neighbor (KNN) rely on specific design intended for features such that good features and patterns must be recognized. These specific features include hue saturation value (HSV), Histogram of Oriented gradient (HOG), linear binary pattern (LBP), and red-green-blue RGB color features [12]. In machine learning, according to the complexity of the classifier, the more data is required for its training in order to achieve satisfactory results. The Fig. 2 illustrates the general flow of the implementation of machine learning techniques in detecting plant diseases. Initially, the data is acquired and labeled according to disease classes or healthy class. A specific dataset is then created for the model, where the input images can be pre-processed before feature extraction can take place. Machine learning algorithms are capable of recognizing the changes in features upon comparison, and thus determining the output as diseased or healthy [13].

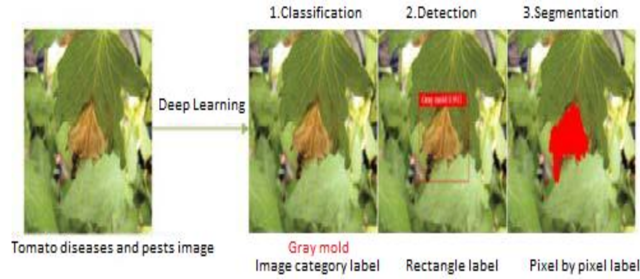


Fig. 3. Plant Disease Detection Phases through Deep Learning.

On the other hand, deep architectures like CNN (Convolutional Neural Networks) have also been heavily used in studies that are concerned with plant disease detection. These deep architectures differ from the shallow ones by not requiring hand-designed features since deep learning algorithms are able to learn the features themselves. Thus, deep learning approaches undergo three basic stages in detecting plant diseases (see Fig. 3): classification, detection, and segmentation [14].

After SVM machine learning approach was the most commonly used one for so long, approximately after the year 2015 CNN replaced SVM as the most popular ML technique for detection of diseases. CNN is considered state-of-the-art model that has been used in plant disease detection nowadays, especially since this task requires dealing with image data applications [13].

CNN can execute tasks such as classification of images, segmentation, object detection, and recognition. In their structure, CNNs are made up of artificial neural networks where tens and even hundreds of layers are used [15, 16]. CNNs is made up of an input layer, several convolutional layers, along with pooling layers in between them, and finally full connection layer in addition to activation function layers, and output layer. There exist several forms of CNN architectures like VGG-16, Inception-V3, ResNet50, and AlexNet. However, CNN architectures need large data numbers which is often considered as a challenge [17-19].

Since agriculture is essential there's a need to provide methods that enhance the agricultural methods in terms of planting, monitoring crop environment, detecting plant disease, and even harvesting. These important details led to significant research to be conducted and several papers to be published with the purpose of providing solutions to these agricultural challenges. This study proposes a model based on CNN, namely VGG-16 architecture in order to detect and classify a total of 19 plant conditions (several crop types and diseases) with the best accuracy possible.

Our contributions in this study can be summed up as follows:

1) Updating a large dataset based on Plant Village. The dataset comprises 15 thousand images of plant leaves which are captured on the field, which means that they are photographed within their surroundings, and thus it is efficient in terms of not needing to isolate the plant for disease detection.

2) Implementing the proposed VGG-16 model which is an effective convolutional neural network architecture, and it achieves a great accuracy. Our proposed model is capable of scanning through thousands of leave images in order to identify if a plant has a certain type of disease based on its leaf image. The proposed model achieves a great accuracy of detection among 19 different disease classes in a short period of time, and it doesn't require a long time in training either.

The arrangement of the current paper is as follows: section two is a description of some of the published similar studies about ML and DL in plant disease detection. Section three describes the proposed methodology including the dataset and the proposed model. Results are provided and compared theoretically with some of existing techniques in section four, while section five concludes the article by sharing the future research intentions.

II. Related Work

A large number of studies have been performed and published about the detection of diseases in plants, especially since disease detection and classification can help in treating the diseased plant. The more technology emerges, the more it will be implemented in the agricultural field, one of the reasons being the continuous enhancement of already existing machine learning and deep learning methods that identify and classify plant diseases. This section describes five studies conducted in this regard.

The study by Ashwin et al. [20] discussed the development of a machine learning approach for the detection of diseases in Soybean plant in Mazandaran province in Iran. The study aimed to illustrate the importance of incorporating physiological features as well as the morphological features in realizing the diseases presence in a plant. The dataset used by the authors contains two randomly chosen subsets one for the healthy soybean plants (negative) and one for the diseased soybean plants (positive), which makes up to 2,500 images total. These images were collected from 10 different regions in the Mazandaran province. Several features were selected based on one-way ANOVA F1-score which include stem length, root length, number of pods per plant, empty pods per plants and seed oil content. Six different machine learning techniques with ten-fold cross validation and internal cross validation, which are the binary Logistic Regression, Multilayer Perceptron with Adam optimization, Random Forest, Gradient Tree Boosting (GBT), and Support Vector Machine. These ML techniques are evaluated based on confusion matrix, precision, F1-score, AUC, etc. Out of the six techniques, GBT achieved the best results of 96.79% accuracy, 96.68% F1-score, and 97% sensitivity and specificity depending on 21 features. The same technique achieved similar results (96.13%) depending on 12 morphological features.

Xian et al. [21] relied on a supervised machine learning termed Extreme Learning Machine "ELM" to determine the presence or absence of tomato diseases. The developed system was based on analyzing tomato leaves; the whole dataset was acquired from Kaggle, specifically the Plant Village dataset. The dataset comprises 10 classes, which makes a total of 1,000 images. The model is a feed forward neural network made up of several hidden nodes with weights that connect and the complete neural network has no iterations. Image pre-processing is done on the images in the dataset including image resizing, segmentation, color space conversion, and HSV. A scatter plot of HSV is done on the resized image of tomato late blight, where a mask can be performed on the segmented image. 7,000 images of tomato leaves were selected for training (70%) whereas 3,000 were selected for testing (30%). The features HSV histogram, Haralick textures, and Colour Moments were selected and the labels were transformed into numerical values via the One Hot Encoding method. In the ELM model, using 1024 neurons and Sigmoid activation function yielded the best disease detection accuracy of 84.94%. When compared to Decision Tree model, ELM achieves better overall accuracies, 77.8% vs. 84.94%, respectively, in all of the 10 classes. On the other hand, SVM is slightly better in terms of accuracy (91.43%) when compared with this ELM model.

In their paper discussing plant disease detection, Bedi et al. [22] proposed a detection model based on convolutional auto encoder "CAE" and CNN as two parts of a hybrid system. The system was implemented on leaf images of peach plants specifically, but it can also be used to spot diseases in other plants. The dataset consists of two different classes healthy peach leaves (2160) and diseased leaves with bacterial spots (2297) which sums up 4457 images in total. 70% of the images in the dataset were used for training the system whereas the remaining 30% were kept for testing the detection performance. After gathering the data into the dataset, the usual pre-processing procedures are performed on the images. The next step is the training of the CAE model and getting compressed domain representations of leaf images from which, a CNN model will be trained afterwards. Upon testing, new leaf images will be used as input and their respective compressed representations will be compared for the CNN to determine if the leaf is healthy or not based on whether a defect is found. The CAE network was made up of 14 layers while the hybrid system was made up of 17 total layers through concatenating layers of the CAE with the CNN layers. Adam optimizer and binary cross entropy were used on the proposed system, where parameters such as NRMSE loss and accuracy measures were used for evaluation. As a result, the testing NRMSE loss was 0.0607, the testing accuracy reached 98.38%, and the F1-score was 98.36%.

Jeyalakshmi et al. [23] developed an approach based on machine learning for classification of diseases of potato and grape crops depending on their leaf images. The dataset was acquired from Pant Village dataset, consisting of 1000 healthy leaf images, as well as 2000 images of diseased potato leaves, and 3270 of diseased grape leaves. In the potato subset, there exist three different classes, while there are four classes in the grape leaves subset. From each RGB leaf images, the background was removed through using "enhanced GrabCut algorithm". Several features were selected from the images such as the intensities of red, green and blue, weighted mean and weighted standard deviation (belonging to histogram features), and texture features including contrast, entropy, etc. After that, the features were normalized and fed into several classifiers, namely KNN, Naïve Bayes, and SVM. It was observed that the number of features (ranging between 7 and

13) affects the accuracy of detection such that the higher the number of features, the better the accuracy. For instance, the highest accuracies were achieved by KNN at 94% considering 13 features, and 96.8% by SVM similarly in potato crops. On the other hand, in grape crops, SVM was superior achieving 96.02% accuracy with respect to 13 features.

Lamba et al. [24] published a paper discussing the implementation of several machine learning algorithms as well as deep learning techniques to properly detect diseases in crops. The images comprising the dataset are from Kaggle and rice dataset. In fact, four different datasets are created which are: the rice dataset containing 120 images, the pepper dataset containing 1,994 images, potato dataset containing 2,152 images, and tomato dataset containing 16,072 images. As usual, the images are pre-processed accordingly, and then auto-color correlogram filter is applied as an additional pre-processing step. After that, classification took place through machine learning techniques, or through deep learning methods with various activation functions to assess the overall performances. With respect to machine learning, fifteen different algorithms were tested including bayes net, random forest, and iterative optimized classifier to name a few. On the other hand, in the deep learning method, different activation functions were used including SoftMax, Softsign, ReLu, etc. The evaluation of performance was done relying on confusion matrix, specificity, F1-score, and accuracy. As a matter of fact, each of the models (ML and DL) was implemented on each one of the four datasets. In Pepper dataset, random forest and DL-Softsign achieved the best results as ML and DL techniques respectively. In rice dataset, random forest and DL-Softmax achieved the best results. SVM and DL-Softsign were responsible for the better results in the tomato dataset, and SVM and DL-Hard-tanh outperformed the rest in the potato dataset. In conclusion, from the large set of experiments, the authors concluded that in binary classification, DL with Softsign is the best, whereas Softmax is better suited for multi-class classifications. Table I compares between theses similar systems.

From the analysis of these previous studies, a group of limitations can be seen, as most of the studies focus on detecting the diseases in one type of crops. On the other hand, logically, most farms are interested in growing more than one crop, especially if the farmer is using these crops for the local market, whereas in some cases a field might be planted with one specific type of crop which is grown for a huge demand and can be used for exportation and national trade. However, the aim is to create a model that can be used by the various farmers, whether they have a small farm growing multiple crops, or they had large farms with more than one type of crop. For this reason, we chose to take into consideration a group of plants to be analyzed and to train our model to detect their diseases respectively. Furthermore, the second aim is to create a reliable model which achieves very accurate results with minimal loss and minimal false detection (both false positive and false negative), for this reason the majority of the studies focus on CNN models for the disease detection class. Similarly, our chosen CNN architecture in VGG-16 still stands as one of the best computer vision models. VGG provides better results while having smaller convolutional layers since it doesn’t rely on a large number of hyper-parameters, yet it has several 3x3 filters with stride 1, and max layers and padding with stride 2 that are distributed throughout the architecture. In addition, the VGG16 model not only perform image recognition and detection, but it is also capable of localization, meaning that in our case and objectives, the model can be improved to detect the exact location of the classified disease on the plant leaf.

TABLE I. COMPARISON BETWEEN SIMILAR SYSTEMS

Study	Year of Publication	Crop Type	Dataset	Technique	Performance Results
A Machine Learning Approach to Prediction of Soybean Disease” [20]	2021	Soybean	Real samples from Mazandaran province	LR-L1, LR-L2, MLP, RF, GBT and SVM	Accuracy: 96.7% F1-score: 96.5%
Plant Diseases Classification using Machine Learning [21]	2021	Tomato	Plant Village	Extreme Learning Machine (ELM)	Accuracy: 84.94% F1-score: -
Plant diseases detection using hybrid model based on convolutional autoencoder and convolutional neural network [22]	2021	Peach	Plant Village	CAE and CNN	Accuracy: 98.38% F1-score: 98.36%
An effective approach to feature extraction for classification of plant diseases using machine learning [23]	2020	Potato and grape	Plant Village	Naïve bayes, K nearest neighbor and support vector machine classifier	Accuracy: 96% F1-score: 95%
“Classification of plant diseases using machine and deep learning [24]	2021	Rice, pepper, potato and tomato	Plant Village	Auto-color correlogram and deep learning	Accuracy: 99.4% F1-score: 95%

III. Implementation of the Model

In order to maintain constant care of plants in any farm, especially a vertical farm, you need to constantly keep a close eye on the crops and their leaves as well as their stems and all the surrounding conditions. Of course, detecting plant diseases from looking into plant leaves manually takes a lot of time and effort from the farmer, which is why automation could be a game changer in the field of plant disease detection. Computational techniques can be applied such that the images of the plants are being taken and analyzed or screened for diseases in a time efficient manner, without facing human effort or human error challenges.

In this project, we implement several deep learning technologies in order to classify the diseases in crops. Among machine learning techniques, deep learning is an interesting area since it can easily recognize patterns and perform complicated procedures, which makes it suitable for the disease classification task. In deep learning, deep neural networks are created, among which is the CNN, short for Convolutional Neural Networks, that can easily analyze images upon training. In our project, VGG-16 was selected to be the CNN classifier of different plant diseases.

A. Dataset

The chosen dataset for our study is the Plant Village dataset which contains numerous images of plant diseases totaling for 15,915 images. From these images, the dataset is divided into nineteen different classes, each class resembles a disease, and these classes do not overlap, meaning that a single image only belongs to one class and cannot simultaneously belong to two or more classes.

The nineteen classes include tomato, grape, apple, tomato, and corn diseases as well as healthy classes. The Fig. 4 and Fig. 6 demonstrate in detail the different classes and to which crop they belong.

The number of images per class varies between the classes, as shown in the Fig. 5. After the data had been selected, they are collected in a csv file, where each image is labeled with a path that indicates the health conditions of the crop within the image.

Fig. 4. Crop Disease Classes in Plant Village Dataset-1.

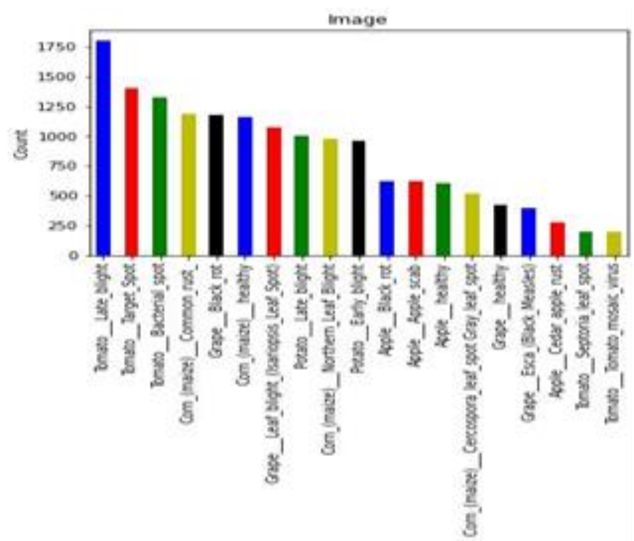


Fig. 5. Distribution of Images in Plant Village Dataset into Nineteen different Classes.

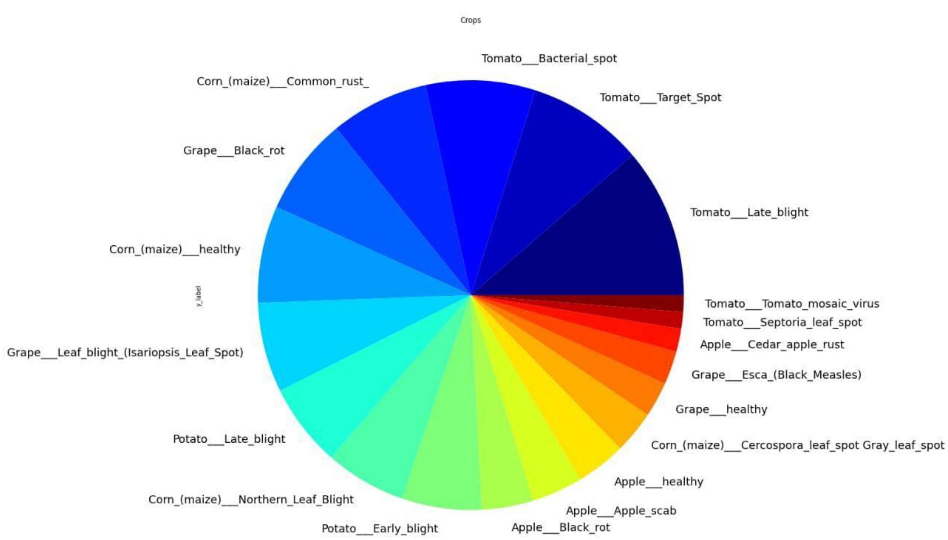


Fig. 6. Crop Disease Classes in Plant Village Dataset-2.

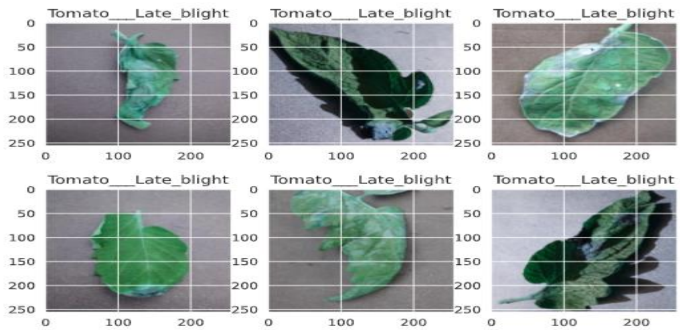


Fig. 7. Leaves with Diseases.

The Fig. 7 illustrates some images of the tomato late blight class which refers to a disease in the tomato crop.

In order for the data to be convenient for our chosen model, all the text value had to be converted into numeric values that are suitable for the model. Moreover, the majority of dataset's images (80%) were utilized in the training procedure, and the rest 20% were kept for testing.

B. Proposed Model

VGG-16 is one of the most commonly used CNN architecture, especially since it works well with the ImageNet, which is large project utilized for visual object recognition procedures and it is considered one of the best models to be proposed so far due to its extreme usefulness in the image classification's field in the deep learning domain. Initially, this model was created by Karen Simonyan and Andrew Zisserman in 2014, where they developed in during their work in Oxford University titled "Very Deep Convolutional Networks for Large-Scale Image Recognition". In fact, "V" means Visual," G" Geometry while "G" stands for research group who contributed in the development of this Convolutional Neural Network model, whereas the number 16 refers to the neural network layer's number. ImageNet is so large that it contains more than fourteen million images distributed over thousand classes. This architecture is one of the top 5 models in terms of performance achievement in the ImageNet dataset, where its accuracy reached 92.7%. As an approach for the AlexNet enhancement, this architecture was submitted to ImageNet.

Large Scale Visual Recognition Challenge (ILSVRC), where this model has replaced the large kernel-sized filters of numbers 11 and 5 in both first and second convolutional layer, respectively by a multiple three × three kernel-sized filters consecutively. Moreover, the training of this model was trained by utilizing NVIDIA Titan Black GPUs for several weeks.

C. VGG-16 Architecture and Training Procedure

The training Procedure is made up of three consecutive steps as shown in Fig. 8:

- Preparing the images.
- Classifying the photos.
- Printing the decision.

Image Processing: The input of the convents is 224 × 224 RGB image with a fixed size where the value of each pixel is subtracted from the RGB mean value of the training image.

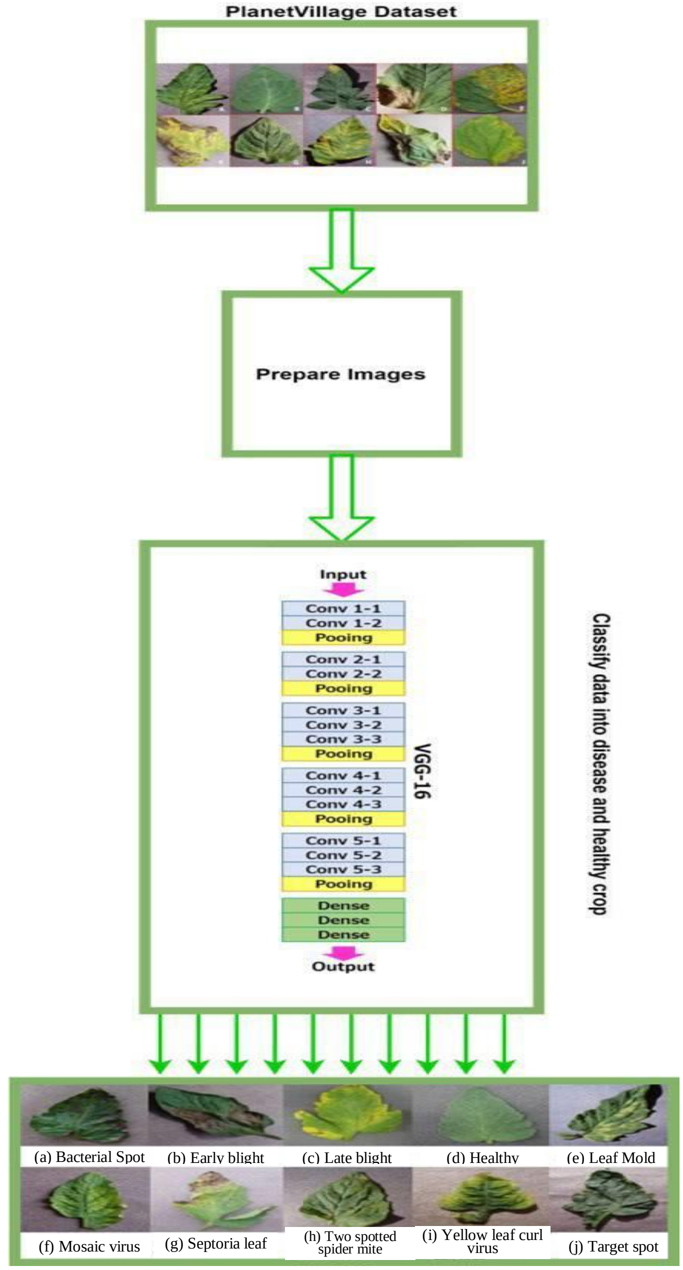


Fig. 8. Training Steps.

Classifying the data: The proposed model is made up of thirteen convolution layers, two batch normalization layers, along with five max-pooling layers and three full connection layers.

The processed image passes through several convolutional layers that contains filters that are characterized by a receptive field of size 3 × 3 for capturing the notions of left and right, up and down along with the center. Despite its small size of the mentioned filter, this filter is accompanied by the same efficiency as that of a receptive field of size 7 × 7 due to its deep characteristics such as including more nonlinearities and lesser parameters. In addition to that, a 1x1 convolution filter was used as an input channel's linear transformation in a certain configuration. On the other hand, both spatial padding and the convolution stride are fixed to 1 pixel for 3 × 3 convolutional layers, in which the spatial resolution's preservation becomes easy to occur. Also, spatial pooling is easier in case of a five max-pooling layers' addition after some of the convolutional layers and the Max-pooling layer takes place over a 2×2-pixel window, with stride 2.

In addition to that, a total of three varying FC (Fully Connected) layers in depths are fixed behind a group of convolutional layers, where the first two FC layers is made up of 4096 channels per FC layer, and the third performs 1000-way ILSVRC classification and is made up of 1000 channels for each class. Finally, the final layer is the soft-max layer, it's important to say the Fully Connected Layer's configuration does not vary among different networks.

The architecture of the VGG-16 model is portrayed in details in the Fig. 9.

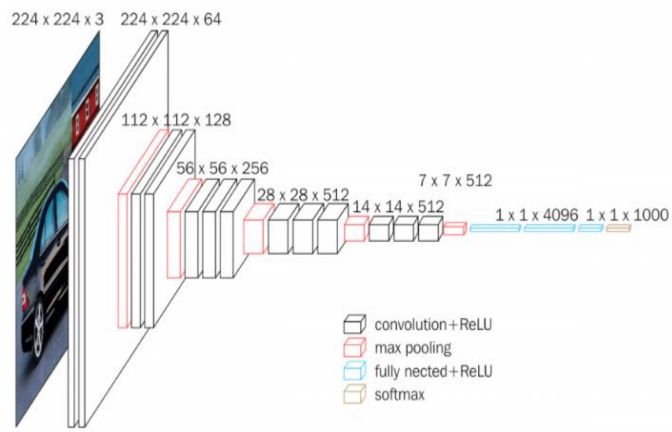


Fig. 9. VGG-16 Detailed Architecture showing the Various Layer and their Positioning.

D. Activation Function Used

Two activation Functions were used for our model training where the Softmax activation and the ReLU function.

The ReLU function was used at the fully connected layers, where the ReLU or "Rectified Linear Unit" is one of the popular activation functions used in Neural Networks and specifically in Convolutional Neural Networks and is defined as in (1):

$$y = max\ (0, x) \tag{1}$$

Moreover, the Softmax activation function is used for the output layers and this activation function is a type of logistic regression that is able of normalizing the inputted vector to a new vector where its probability distribution is equal to 1 and it is defined as in (2):

$$\sigma(\vec{a})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{2}$$

E. Loss Functions used

In the machine learning domain, the cost functions tend to optimize the model in the training procedure and the aim of the training procedure is to minimize the loss function and the model obtained is better as much as we tend to minimize this loss function. Therefore, one of the most important loss functions is the Cross Entropy Loss Function where it is used for Classification model's optimization and the complete understanding of this loss function depends on the Softmax activation function understanding. Moreover, in our project, the Sparse Categorical Cross Entropy is used for training our model where it has the same loss function as that of the cross entropy as in (3):

$$Loss = -\sum_{i=1}^{output\ size} y_i log\hat{y_i} \tag{3}$$

However, the truth labelling procedure is what differs between the two loss functions, where in the case of a one hot encoded true labels ([1,0,0], [0,1,0] and [0,0,1] in 3 classification problem) the categorical cross entropy is used, while the cross entropy is used in the case of an integer truth labels coding ([1],[2],[3]).

F. Accuracy Calculation

The accuracy of the model is calculated as in (4):

$$Accuracy = \frac{True\ Positive}{True\ Positive + True\ Negative} \times 100 \tag{4}$$

IV. Results

The performance of our proposed model was assessed in terms of Loss Function and Accuracy, which was measured during the training/validation step as well as during the testing step. The Fig. 10 shows the performance during the training phase.

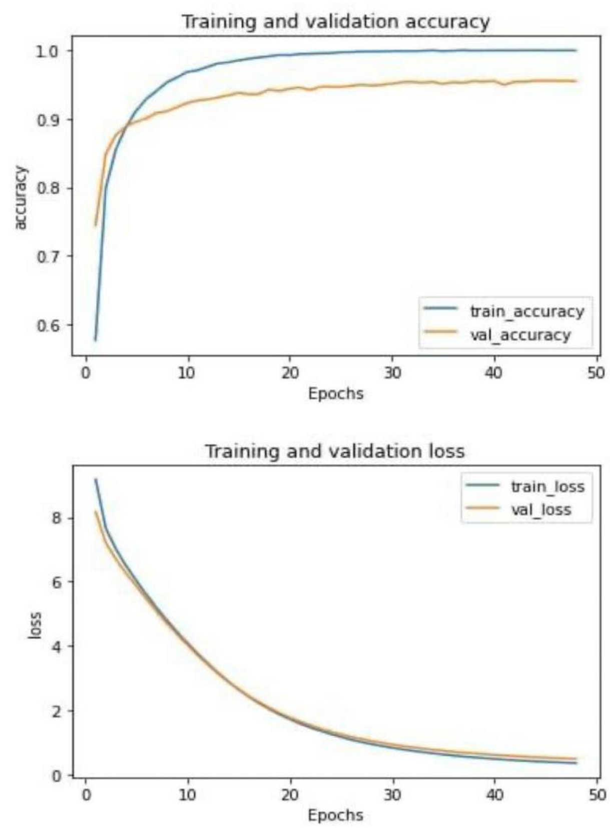


Fig. 10. The Performance of Proposed Model in Terms of Training and Validation Loss and Accuracy.

In the testing phase, the proposed VGG-16 model achieves the 93.5% accuracy, taking into consideration that the model performs classification on 19 different classes. The performance of our model is excellent for detecting diseases automatically from plant images. On the test set, which is excellent accuracy for classifying 19 items of plant images where the Test loss has reached 0.44 while the Test accuracy has reached 0.95. Moreover, the test accuracy was also computed for each of the 19 cases as well as the estimated distribution for the prediction and probability is shown in the Fig. 11 and Fig. 12.

A. *Discussion*

In the light of these results, and taking into consideration all of the parameters such as the big dataset and the variety of used crops, we can compare our results with those of similar studies that also used VGG networks for plant disease detection. For instance, the study by Iftikhar Ahmad implemented the VGG-16 model among others for the detection of diseases that infect tomato leaves, where the VGG16 network achieved a 76.2% accuracy by using feature extraction on augmented field dataset of 15 thousand images, and 84% accuracy for VGG16 when using parameter tuning [25]. Another study also used VGG and achieved an accuracy of 95.32% when using the Plant Village dataset of 4062 grape leaf images [26]. Narayani Patil and his colleagues also used VGG-16 networks for detection of plant diseases, and it achieved 97.53 % accuracy [27]. Finally, a pretrained VGG-16 model achieved an accuracy of 94.3% compared to 96.5% for the fine-tuned VGG-16 while using the Plant Village dataset but only for tomato leaves [28]. Taking into consideration that in our model we used a large dataset and multiple crops, it can be concluded that our VGG-16 model performed better in terms of accuracy than these studies.

Increasing agricultural movement is becoming more and more important by the day due to the rapid increase in population which increases the food demand, and ultimately means the need for more crops. Maintaining crop health is a very important field, since growing the plants in their ideal conditions leads to better yield which comes with greater benefits. New technologies are invested in the agricultural domain to take care of the wellbeing of plants.

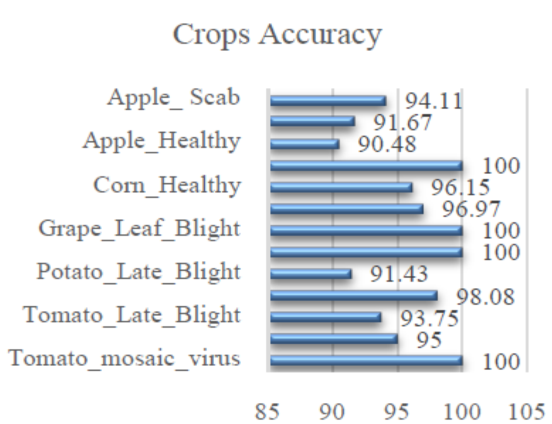


Fig. 11. Crops Accuracy.

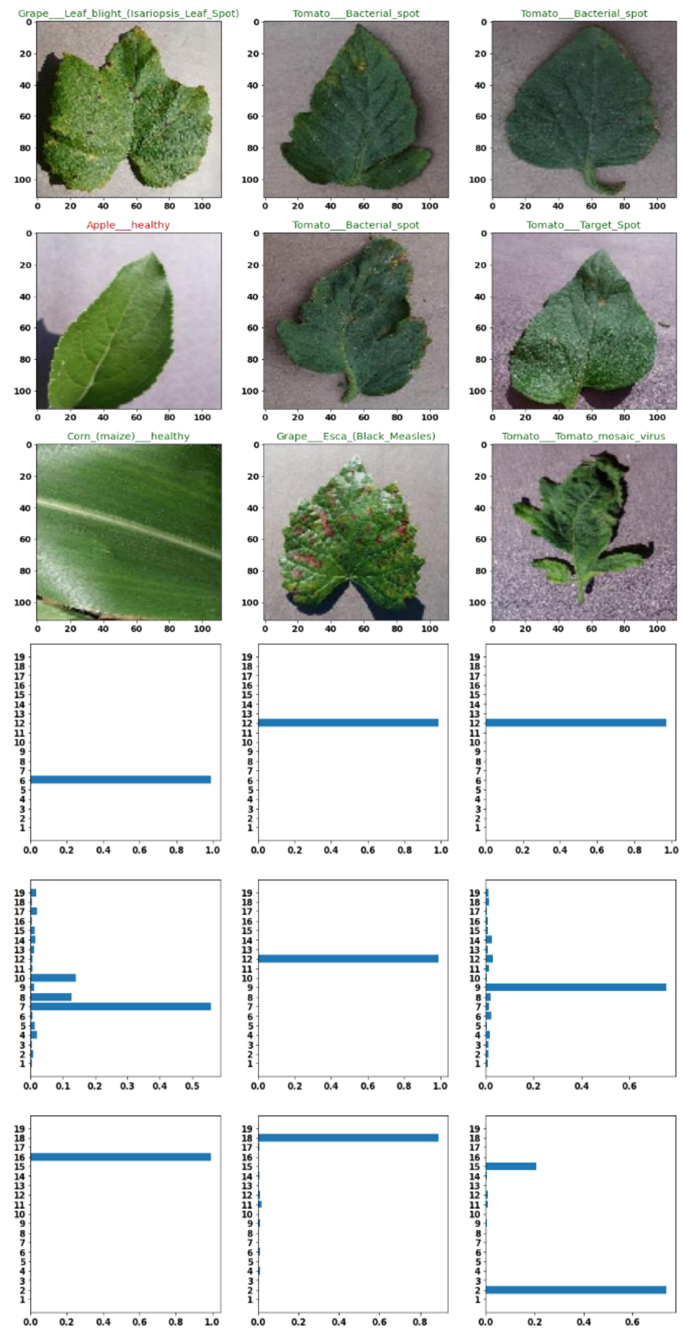


Fig. 12. Probability Estimated Distribution.

One of the domains where technology can have a large impact is the detection of plant diseases. Machine learning and deep learning are often implemented in systems for this specific purpose especially since ML techniques have an important image processing capability.

V. Conclusion

In this study, we proposed a system for the detection of plant diseases through analyzing leaf images of plants to determine not only if they are healthy or diseased, but rather to classify which kind of disease exists in each crop type. Our model is based on a VGG-16 architecture that classifies 19 classes of plant diseases, according to the data acquired from the Plant Village dataset. The model was able to achieve a 95.2% accuracy with a loss of 0.4418.

Despite achieving a high accuracy with a low loss, our model faces some limitations since the input images must have certain illumination conditions and a complex background behind them due to the fact that they are collected from actual leaves from planted plants. These conditions pose as a challenge for any model used for plant disease detection and they can be considered as areas for improvement when designing or trying to enhance the existing model. Furthermore, in the future studies, our efforts will be focused on achieving more precise disease detection, particularly through training our machine learning model to identify the exact location of the disease on each leaf, especially if more than one disease is detected in one plant leaf. In addition to that, the plant disease dataset can be further increased to take into consideration even more plant diseases and to incorporate additional crop types. Moreover, we can consider some advanced methods to increase the accuracy of processing of leaf images by applying technologies like Faster region-based convolutional neural network (Faster R-CNN), which is a unified network designed for object recognition. Faster R-CNN creates a network that proposes a region for the detection which is then fed to the developed model for training, and after that according to the features, the optimal detection region is selected for classification purposes. Another method that can be implemented is the You Only Look Once YOLO technology which presents a very fast detection in real time with approximately 45 frames for second. Another technique similar to YOLO is the Single Shot Detector (SSD) which provides a fast detection of objects from a single frame. SSD achieves its high accuracy by producing detection at different scales and separates between the predictions by aspect ratio.

REFERENCES

[1] Ramsanthosh, A. S. Laxmi, S. A. Chepuri, V. Santosh, V. Kothareddy, and S. Chopra, "Review and further prospects of plant disease detection using machine learning," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 7, pp. 105-115, 2021.

[2] B. V. Gokulnath and G. Usha Devi, "A survey on plant disease prediction using machine learning and deep learning techniques," Inteligencia Artificial, vol. 23, pp. 136-154, 2020.

[3] J. Liu and X. Wang, "Plant diseases and pests detection based on deep learning: A review," Plant Methods, vol. 17, p. 22, 2021.

[4] T. Su, W. Li, P. Wang, and C. Ma, "Dynamics of peroxisome homeostasis and its role in stress response and signaling in plants," Frontiers in Plant Science, vol. 10, p. 705, 2019.

[5] Ayaz, Muhammad, et al. "Internet-of-Things (IoT)-based smart agriculture: Toward making the fields talk." IEEE access 7 (2019): 129551-129583.

[6] S. Roy, R. Ray, S. R. Dash, and M. K. Giri, "Plant disease detection using machine learning tools with an overview on dimensionality reduction," in Data Analytics in Bioinformatics, R. Satpathy, T. Choudhury, S. Satpathy, S. N. Mohanty, and X. Zhang, Eds. Beverly, MA: Scrivener Publishing LLC, 2021, pp. 109-144.

[7] D. B. Collinge, H. J. L. Jørgensen, M. A. C. Latz, A. Manzotti, F. Ntana, E. C. Rojas, et al., "Searching for novel fungal biological control agents for plant disease control among endophytes," in Endophytes for a Growing World, B. R. Murphy, F. M. Doohan, M. J. Saunders, and T. R. Hodkinson, Eds. Cambridge: Cambridge University Press, 2019, pp. 25-51.

[8] A. M. Abdu, M. M. Mokji, and U. U. Sheikh, "Machine learning for plant disease detection: an investigative comparison between support vector machine and deep learning," IAES International Journal of Artificial Intelligence, vol. 9, pp. 670-683, 2020.

[9] L. Mohamed, E. Ahmed, and A. Mohamed, "Deep learning in plant diseases detection for agricultural crops: A survey," International Journal of Service Science, Management, Engineering, and Technology, vol. 11, pp. 41-58, 2020.

[10] R. D. S. Chowdary, V. Yogeshwari, V. Varsha, M. R. B. Priya, and S. Chinnappan, "Plant disease detection using image processing and machine learning algorithm," Journal of Xidian University, vol. 14, pp. 79-91, 2020.

[11] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," Computers and Electronics in Agriculture, vol. 147, pp. 70-90, 2018.

[12] B. S. Kusumo, A. Heryana, O. Mahendra, and H. F. Pardede, "Machine learning-based for automatic detection of corn-plant diseases using image processing," in 2018 International Conference on Computer, Control, Informatics and its Applications (IC3INA), Tangerang, Indonesia, 2018, pp. 93-97.

[13] H. F. Pardede, E. Suryawati, D. Krisnandi, R. S. Yuwana, and V. Zilvan, "Machine learning based plant diseases detection: A review," in 2020 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET), Tangerang, Indonesia, 2020, pp. 212-217.

[14] M. Turkoglu, D. Hanbay, and A. Sengur, "Multi-model LSTM-based convolutional neural networks for detection of apple diseases and pests," Journal of Ambient Intelligence and Humanized Computing, 2019.

[15] M. Syarief and W. Setiawan, "Convolutional neural network for maize leaf disease image classification," Telkomnika (Telecommunication, Computing, Electronics and Control), vol. 18, pp. 1376-1381, 2020.

[16] Sharif, Z., Jung, L. T., Razzak, I., & Alazab, M. (2021). Adaptive and Priority-based Resource Allocation for Efficient Resources Utilization in Mobile Edge Computing. IEEE Internet of Things Journal.

[17] A. Abdalla, H. Cen, L. Wan, R. Rashid, H. Weng, W. Zhou, et al., "Fine-tuning convolutional neural network with transfer learning for semantic segmentation of ground-level oilseed rape images in a field with high weed pressure," Computers and Electronics in Agriculture, vol. 167, p. 105091, 2019.

[18] Uddin, M., et al. "Cloud-connected flying edge computing for smart agriculture." Peer-to-Peer Networking and Applications 14.6 (2021): 3405-3415.

[19] Sharif, Z., Jung, L. T., & Ayaz, M. (2022, January). Priority-based Resource Allocation Scheme for Mobile Edge Computing. In 2022 2nd International Conference on Computing and Information Technology (ICCIT) (pp. 138-143). IEEE.

[20] N. Ashwin, U. K. Adusumilli, N. Kemparaju, and L. Kurra, "A machine learning approach to prediction of soybean disease," International Journal of Scientific Research in Science, Engineering and Technology, vol. 9, pp. 78-88, 2021.

[21] T. S. Xian and R. Ngadiran, "Plant diseases classification using machine learning," Journal of Physics: Conference Series, vol. 1962, p. 012024, 2021.

[22] P. Bedi and P. Gole, "Plant disease detection using hybrid model based on convolutional autoencoder and convolutional neural network," Artificial Intelligence in Agriculture, vol. 5, pp. 90-101, 2021.

[23] S. Jeyalakshmi and R. Radha, "An effective approach to feature extraction for classification of plant diseases using machine learning," Indian Journal of Science and Technology, vol. 13, pp. 3295-3314, 2020.

[24] M. Lamba, Y. Gigras, and A. Dhull, "Classification of plant diseases using machine and deep learning," Open Computer Science, vol. 11, pp. 491-508, 2021.

[25] I. Ahmad, M. Hamid and S. Yousaf, "Optimizing Pretrained Convolutional Neural Networks for Tomato Leaf Disease Detection," Complexity, 2020.

[26] K. Akshai and J. Anitha, "Plant disease classification using deep learning," in International Conference on Signal Processing and Communication (ICPSC), 2021.

[27] N. Patil, S. Kelkar and M. Ranawat, "Krushi Sahyog: Plant disease identification andCrop recommendation using Artificial Intelligence," in 2nd International Conference for Emerging Technology (INCET), 2021.

[28] H. Agh Atabay, "Deep residual learning for tomato plant leaf disease identification," Journal of Theoretical and Applied Information Technology , 2017.

Analysis of Factors Influencing the COVID-19 Mortality Rate in Indonesia using Zero Inflated Negative Binomial Model

Maria Susan Anggreainy[1], Abdullah M. Illyasu[2], Hanif Musyaffa[3], Florence Helena Kansil[4]
Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480[1, 3, 4]
Tokyo Institute of Technology, Yokohama 226-8502, Japan[2]

***Abstract*—This research aims to create a model, analyze the factors that influence the COVID-19 mortality rate in Indonesia. There are five independent variables and one dependent variable used in the research. The independent variables used are the percentage of poor people, the percentage of households using shared toilet facilities, the percentage of households using wood as the main fuel for cooking, the percentage of the population whose drinking water source comes from pumped water and the percentage of population who have health insurance from private insurance. While the dependent variable used is the Annual Parasite incidence COVID-19. The results obtained are as follows. First, a Zero-Inflated Negative Binomial regression model was obtained for the case of COVID-19 morbidity where this model could overcome overdispersion and excess zero values in observations. Second, there are 4 independent variables that have a significant effect on the count model and there is no independent variable that has a significant effect on the Zero inflation model. Third, a web application is produced that can display the Zero-Inflated Negative Binomial regression model (ZINB).**

Keywords*—Mortality rate; overdispersion; zero-inflated negative binomial; Poisson regression; correlation

I. Introduction

Health is very important for every human being. With a healthy body, the soul will be good, and the mind will be in balance. Having a healthy body and soul can support human activities without any obstacles. Steps that need to be taken to maintain health include exercising, getting appropriate nutrition coverage and guarding yourself from habits that damage the body. Public awareness is needed to carry out a healthy lifestyle to avoid disease. COVID-19 has become a health emergency in the past one month since the first case in Indonesia was reported in Depok in March 2020. The problem that Indonesia is currently facing is that the government is still not effective in conducting testing and COVID-19 cases and there is no enforcement of rules regarding social distancing and mobility [1].

Research on the analysis of factors that influence COVID-19 in Indonesia is carried out using several different methods. The Zero-Inflated method used by the researcher in this research is the Zero-Inflated Negative Binomial method. This research was conducted to model the COVID-19 mortality rate using the web-based Zero-Inflated Negative Binomial method. The Zero-Inflated Negative Binomial method is used because there is overdispersion in the Poisson regression model and the data used has an excessive value of 0 [2]. Overdispersion is a condition where the value of the variance is greater than the average [3]. One of the causes of overdispersion is the number of observations that are zero in the dependent variable or variable Y. Zero-Inflated Negative Binomial Regression (ZINB) is a model formed from a mixed distribution of Poisson gamma [3]. If y_i is a discrete independent random variable with i= 1, 2, n the null value of the observation is assumed to appear in two appropriate ways for separate states. The first state called zero state occurs with probability pi and produces only zero observations, while the second state is called Negative Binomial state occurs with probability (1 - pi) and has a Negative Binomial distribution with mean μ, with $0 \leq p$i ≤ 1 [4]. Estimated ZINB regression parameters using the Maximum Likelihood Estimation (MLE) method with the EM (Expectation Maximization) and Newton Raphson Algorithm procedures [5]. This method is usually used to estimate the parameters of a model whose density function is known.

The use of the Zero-Inflated Negative Binomial (ZINB) method has been carried out by several researchers [6][7][8]. In research conducted concluded that the ZINB regression model is more appropriate to be used to model data on the number of maternal deaths in Bali Province which contains many zero values and experiences overdispersion [9]. In another research conducted and concluded that the Poisson regression model does not meet the overdispersion hypothesis, so another model is used, the proposed model is the Zero-Inflated Poisson (ZIP) model, but there is still overdispersion on the ZIP models [10].

To overcome this problem, the ZINB model and the hurdle negative binomial (HNB) model are used [11]. The Akaike Information Criterion (AIC) value of the ZINB model is lower than the value of the HNB model [12]. This shows that the ZINB model is best used in data on the incidence of diptheria in Indonesia [13]. The HNB model can control zero values and overdispersion, just like the ZINB model [14]. However, in the data on the incidence of diphtheria in Indonesia, the ZINB model is more suitable to control the value and over-dispersion of the data concludes that the appropriate model for the frequency of traveling during the last six months in South Tapanuli regency, North Sumatra, for the March 2016 period is the Zero-Inflated Negative Binomial model [15].

The independent variables used in this research were the percentage of poor population, the percentage of households using shared toilet facilities, the percentage of the population whose drinking water source came from pump water and the percentage of the population who had health insurance from private insurance. The dependent variable used in this research is the COVID-19 mortality rate. Based on the data obtained, there are areas where the COVID-19 mortality rate is zero. This causes the data used to be overdispersion or the average value and variance are different. Therefore, the formulation of the problem in this research is:

1) Create a Zero-Inflated Negative Binomial model for the COVID-19 mortality rate in Indonesia.

2) Analyzing the Factors Affecting the COVID-19 mortality rate in Indonesia.

II. MATERIALS AND METHODS

A. Research Variable

The type of data used in this research is secondary data. The variables used in this research consisted of the dependent variable and the independent variable (can be seen in Table I). Measurement scale of the variables is ratio.

TABLE I. EXPLANATION OF THE VARIABLES USED

Variable Name	Variable type.	Definition
COVID-19 mortality rate	Dependent	API is the morbidity rate per 1000 population at risk in one year.
Percentage of poor people	Independent	Population whose average monthly per capita expenditure is below the poverty line. The unit used is Percent (%).
Percentage of households that use shared toilet facilities	Independent	MCK stands for bathing, washing, and toileting is one of the public facilities that are shared by several families for the purpose of bathing, washing, and defecating in certain residential locations which are considered to have a fairly dense population and low level of economic capacity. The unit used is Percent (%).
Percentage of households that use wood as the main fuel for cooking	Independent	Until now, household energy needs in rural areas are still supported by firewood and agricultural waste. In rural areas, especially in remote areas, people still use more than 60% of their energy needs from firewood or biomass. The unit used is Percent (%).
Percentage of population whose drinking water source comes from pumped water	Independent	Pump water is water that comes from pumping from a water source in the ground, then distributed into existing water pipes in the house or in the water tank. The unit used is Percent (%).
Percentage of population who have health insurance from private insurance.	Independent	Health Insurance is an insurance that provides insurance to the insured to replace any medical expenses which include hospital treatment costs, surgery costs and drug costs. The unit used is Percent (%).

B. Research Steps

The data analysis technique used in this research is descriptive analysis, Poisson regression and Zero-Inflated Negative Binomial. The following is a more detailed explanation of the steps taken by researchers in completing this research.

1) Perform secondary data collection.

2) Perform data processing so that the data used.

3) Determine the type of variable from each variable used.

Conducting descriptive analysis by calculating the Mean, Median and standard deviation values. In addition, it determines the maximum and minimum values of each variable.

The following is an explanation of the steps taken by researchers in making this Zero-Inflated Negative Binomial model (can be seen in Fig. 1):

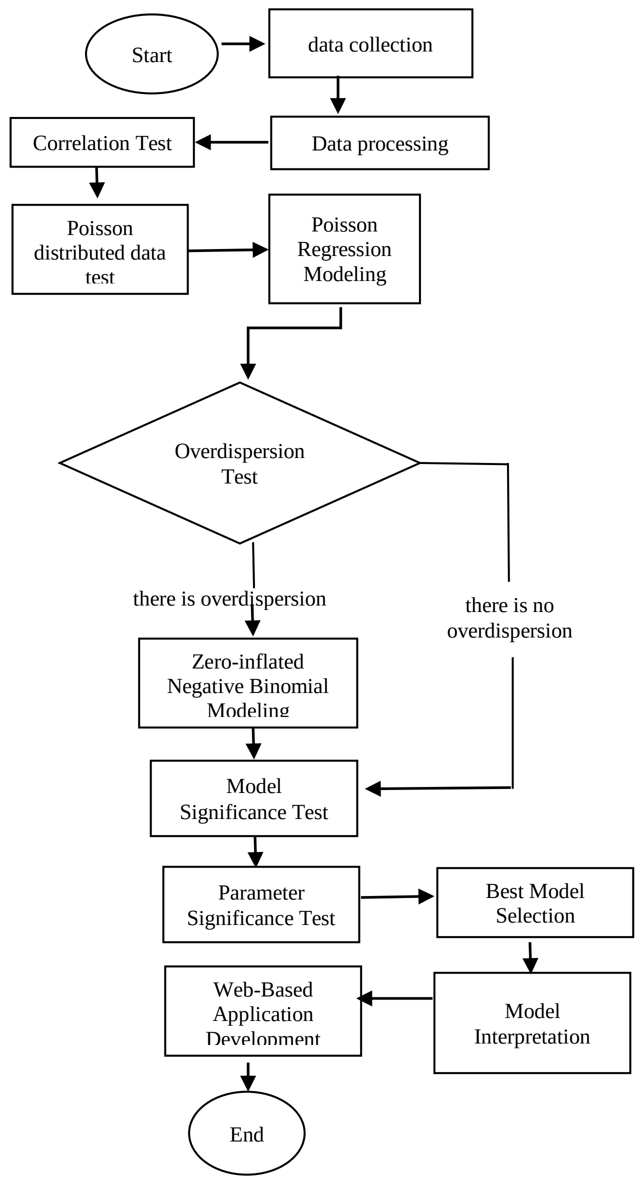


Fig. 1. Flowchart of Research Steps.

1) Perform the Spearman rank correlation test to find out whether there is a relationship between the dependent variable and the independent variable with the following steps:

- Specifies H_0 and H_1.
- Calculate ρ_{count} according to equation.

Equation (1) if there is no the same rank.

$$\rho_{calculate} = 1 - \frac{6\sum d^2}{n(n^2-1)} \quad (1)$$

Equation (2) If there is the same rank.

$$\rho_{calculate} = \frac{\frac{\sum(R_x-\overline{R_x})(R_y-\overline{R_y})}{n}}{\sqrt{\frac{\sum(R_x-\overline{R_x})^2}{n}}\sqrt{\frac{\sum(R_y-\overline{R_y})^2}{n}}} \quad (2)$$

- ρ_{count} =Spearman Correlation Coefficient.
- d = Rank difference for each data.
- n = Number of Samples or data.
- R_x= Rating for each data on Variable X.
- $\overline{R}_x$ = Average Rating for each data on Variable X.
- R_y= Rank for each data on Variable Y.
- $\overline{R}_y$ =Average Rating for each data on variable Y.
  - H_0: $\rho 0$ which means there is no correlation betweenvariables X and Y.
  - H_1 : $\rho \neq 0$ which means there is a correlation betweenthe variables X and Y.
- Determine the $\rho_{(table)}$.
- Comparing the value of ρ_{count} with the value of $\rho_{(table)}$.
- Draw conclusions based on the results of the comparison.

2) Determine the estimated regression coefficient by using the Maximum Likelihood Estimation method and Newton Raphson iteration [16].

3) Create a Poisson Regression model using Equation (3).

$$\ln \hat{\mu}_i = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j x_{ij} \text{ , } i = 1, 2, \ldots\ldots, n \text{ and } j = 1, 2, \ldots\ldots, \quad (3)$$

4) where the coefficients used are the coefficients obtained is step 3.

5) Create a Poisson Regression model using Equation (3).

$$\ln \hat{\mu}_i = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j x_{ij} \text{ , } i = 1, 2, \ldots\ldots, n \text{ and } j = 1, 2, \ldots\ldots, p \quad (3)$$

where the coefficients used are the coefficients obtained in step 3.

6) Perform the Overdispersion test on Poisson regression with the following steps [17]:

- Specifies H_0 and H_1.
- Determine the value of Deviance (θ_1) and Pearson Chi-Squared (θ_2).
- Compares the values of θ_1 and $\theta_2$with 1.
- Draw conclusions based on the results of the comparison.

7) If Overdispersion is found, then the model made is Zero Inflated Negative Binomial.

C. Analyzing

After the previous steps have been taken, the steps taken to analyze the factors that influence the COVID-19 mortality rate are as follows:

1) Determine the estimated regression coefficient by using the Maximum Likelihood Estimation method and Newton Raphson iteration.

2) Create a Zero-Inflated Negative Binomial model using Equations (4).

$$logit\ \hat{\pi}_i = \hat{\gamma}_0 + \sum_{j=1}^{p} \hat{\gamma}_j x_{ij} \text{ , } i = 1,2,\ldots.,n \text{ and } j = 1,2,\ldots.,p \quad (4)$$

p : number of independent variables.

n : number of observations.

$\hat{\beta}$: estimated ZINB regression model parameters.

$\hat{\gamma}$: estimated ZINB regression model parameters.

3) Conducting the test simultaneously with the following steps:

a) Specifies H_0 and H_1.

b) Calculating the value of the G test statistic according to Equation (5).

$$W = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}\right)^2 \quad (5)$$

The test criteria is reject H_0 if W > $t(\frac{\alpha}{2}; n-1)$. Denial H_0 at the level of significance α means that a certain j-th independent variable has a significant contribution to the dependent variable Y.

c) Determining the value $X^2_{(\alpha,n-k-1)}$.

d) Comparing the value of the G test statistic with the value of $X^2_{(\alpha,n-k-1)}$.

e) Drawing conclusions based on comparison results.

4) Conduct a parameter significance test for each of the independent variables in the former model to find out which variables influence the COVID-19 mortality rate. The test steps are as follows:

a) Specifies H_0 and H_1.

b) Calculating the value of the Wald test statistic according to Equation (5).

c) Looking for value $t(\frac{\alpha}{2};n-1)$.

d) Comparing the statistical value of the Wald test of each independent variable with the value of $t(\frac{\alpha}{2};n-1)$.

e) Determining the independent variables that affect the dependent variable based on the comparison results.

5) Choose the best model based on Akaike Information Criterion (AIC) based on Equation (6).

$$AIC = 2k - 2\ln(\hat{L}) \tag{6}$$

k = Number of parameter estimates in the model $\hat{L}$= The maximum value of the model Likelihood function

6) Make an interpretation of the model that has been made.

III. Result and Discussion

In this research, a web-based application was made. Users must register their personal data first as a complete account. After the account has been successfully created, the user will be directed to the Login page to input the registered Username and Password. If the Username and Password match, the User can access the features contained in the application. After successful login, the user will be directed to the home menu. In the home menu there is a Navigation bar that contains the features found in the application. Users can see the results of predictions of mortality rates that have been done before or Users can make new interpretations of COVID-19 mortality rates based on existing variables. After the user has finished using the application, the user can log out of the account by pressing the logout button. Use cases in this research can be seen in Fig. 2.

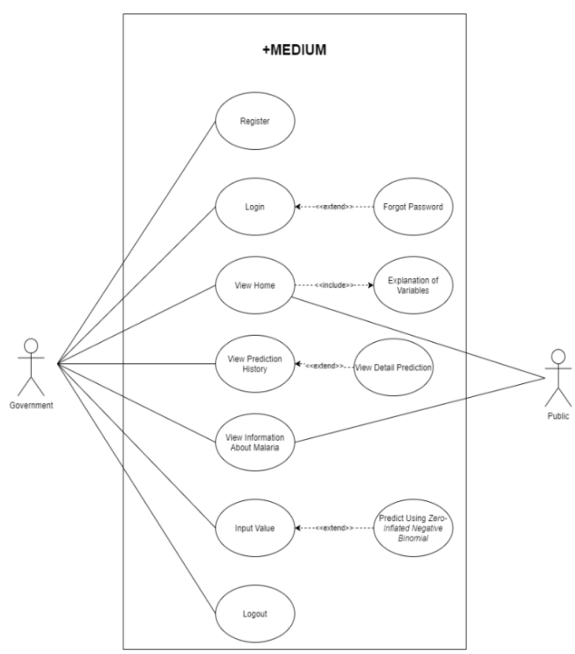


Fig. 2. Use Case.

A. Correlation Analysis

The independent variables tested were the percentage of poor people (X_1), the percentage of households using shared toilet facilities (X_2), the percentage of households using wood as the main fuel for cooking (X_3), the percentage of the population whose drinking water source came from pumps (X_4) , the percentage of the population who had health insurance from Private Insurance (X_5) , Area Height (X_6) , Number of Hospitals (X_7), Number of Doctors (X_8) and Percentage of Households with Ground Floor (X_9) . For example, the correlation value of X1 and Y is -0.4570287, then the interpretation of the correlation value is that the relationship between the percentage of the poor and the COVID-19 mortality rate is in the moderate category and contradicts each other where if the percentage of the poor increases, then the mortality rate will decrease (can be seen in Table II).

Based on Table I, the correlation value of several independent variables is greater than $(0.05,29)$ which means that several independent variables have a significant effect on the dependent variable. The influential variable is Percentage of poor people (X_1), the percentage of households using shared toilet facilities (X_2), the percentage of households using wood as the main fuel for cooking (X_3), the percentage of the population whose drinking water source came from pumps (X_4), and the percentage of the population who had health insurance from Private Insurance (X_5). Therefore, the variables Area Height (X_6), Number of Hospitals (X_7), Number of Doctors (X_8) and Percentage of Households with Ground Floor (X_9) are excluded from the modeling to be carried out.

B. Poisson Distributed Data Test

From the test, the results of the D_n test statistic value of 0.4286. These results were compared with the statistical value of the Kolmogorov-Smirnov test (0.05;7) which was worth 0.48343. Because the statistical value of the D_n test is smaller than the statistical value of the Kolmogorov-Smirnov test, it can be concluded that H_0 a failed to reject or the data came from a population that followed the Poisson distribution. Based on the conclusion, the regression model that will be made is Poisson regression modeling.

TABLE II. Spearman Correlation Test

Variabl e	$\boldsymbol{\rho}$	Direction	$\boldsymbol{\rho}(\mathbf{0.05,29})$	Conclusio n
$X1$	0.4570287	Negative	0.368	There is a relationshi p
$X2$	0.435146	Positive	0.368	There is a relationshi p
$X3$	0.3951663	Negative	0.368	There is a relationship
$X4$	0.4761548	Positive	0.368	There is a relationshi p
$X5$	0.4286245	Positive	0.368	There is a relationshi p
$X6$	0.290077569	Negative	0.368	No connection
$X7$	0.292051614	Positive	0.368	No connection
$X8$	0.146538363	Positive	0.368	No connection
$X9$	0.196191728	Negative	0.368	No connection

TABLE III. POISSON REGRESSION MODELING RESULTS

Variable	Coefficient	Standard Error	Test StatisticValue	Conclusion
Intercept	3.772434	0.318566	11.842	Significant
X_1	-0.004478	0.013024	- 0.344	Not Significant
X_2	0.078039	0.021383	3.650	Significant
X_3	-0.034350	0.003827	- 8.975	Significant
X_4	0.085689	0.011711	7.317	Significant
X_5	-1.014189	0.201304	- 5.038	Significant

Table III shows the coefficient values, standard error, test statistics and conclusions from each Poisson regression variable. The influential variable is Percentage of poor people (X_1), the percentage of households using shared toilet facilities (X_2), the percentage of households using wood as the main fuel for cooking (X_3), the percentage of the population whose drinking water source came from pumps (X_4), and the percentage of the population who had health insurance from Private Insurance (X_5). The model formed based on the values shown in Table III is as follows:

$$\ln \hat{\mu_i} = 3.772434 - 0.004478 * X_1 + 0.078039 * X_2 - 0.034350 * X_3 + 0.085689 * X_4 - 1.014189 * X_5.$$

C. Zero-Inflated Negative Binomial Modeling

The Zero-Inflated Negative Binomial (ZINB) regression model is a regression model that can be used to model data with the dependent variable having a Poisson distribution, many observations that are zero in the dependent variable and overdispersion occurs.

Table IV displays the coefficient values, Standard error, test statistics, and conclusions from each of the Zero-Inflated Negative Binomial regression variables. Based on the table, it is known that the percentage of the population who has health insurance from private insurance does not have a significant effect on both models. So that the Zero-Inflated Negative Binomial modeling was carried out without using the variable Percentage of the population who had health insurance from private insurance. The results of modeling can be seen in Table V.

TABLE IV. RESULTS OF THE ZERO-INFLATED NEGATIVE BINOMIAL REGRESSION MODEL

Variable	Coefficient	StandardError	Wald TestStatistic Value	Conclusion
β_0	2.26342	0.33687	6.719	Significant
β_1	0.48196	0.06754	7.136	Significant Influenceon the Dependent Variable for the Count Model.
β_2	0.12367	0.05760	2.147	Significant Influenceon the Dependent Variable for the Count Model.
β_3	-0.24037	0.04078	-5.894	Significant Influenceon the Dependent Variable for the Count Model.
β_4	0.13436	0.02281	5.891	Significant Influenceon the Dependent Variable for theCount Model.
β_5	-0.49147	0.48896	-1.005	No Significant Effecton the Dependent Variable for theCount Model.
γ_0	-15.2554	5294.8945	-0.003	Not Significant
γ_1	2.0205	709.9964	0.003	No Significant Effecton the Dependent Variable for the Zero-InflationModel.
γ_2	-13.0143	3093.7702	-0.004	No Significant Effecton the Dependent Variable for the Zero-InflationModel.
γ_3	-0.2961	287.7473	-0.001	No Significant Effecton the Dependent Variable for the Zero-InflationModel.
γ_4	-1.2309	2178.5769	-0.001	No Significant Effecton the Dependent Variable for the Zero-InflationModel.
γ_5	34.4325	8887.1670	0.004	No Significant Effecton the Dependent Variable for the Zero-InflationModel.

TABLE V. ZERO-INFLATED NEGATIVE BINOMIAL MODELING RESULTS WITHOUTVARIABLE PERCENTAGE OF POPULATION WHO HAVE HEALTH INSURANCE FROM PRIVATE INSURANCE

Variable	Coefficient	StandardError	Wald Test Statistic Value	Conclusion
β_0	2.13703	0.32859	6.504	Significant
β_1	0.52405	0.06850	7.650	Significant Influenceon the Dependent Variable for theCount Model.
β_2	0.14371	0.05981	2.403	Significant Influenceon the Dependent Variable for theCount Model.
β_3	-0.26860	0.04007	-6.704	Significant Influenceon the Dependent Variable for the Count Model.
β_4	0.15242	0.01937	7.870	Significant Influenceon the Dependent Variable for the Count Model.
γ_0	-27.395	24062.983	-0.001	Not Significant
γ_1	7.081	2213.102	0.003	No Significant Effecton the Dependent Variable for the Zero-InflationModel.
γ_2	-17.144	697.278	-0.025	No Significant Effecton the Dependent Variable for the Zero-InflationModel.
γ_3	-2.119	617.880	-0.003	No Significant Effect on the Dependent Variable for theZero-Inflation Model.
γ_4	-9.382	2751.073	-0.003	No Significant Effect on the Dependent Variable for the Zero-InflationModel.

The model formed can be as follows:

a) Count Model Coefficient

$$\hat{\mu} = exp(2.13703 + 0.52405 * X_1 + 0.14371 * X_2 - 0.26860 * X_3 + 0.15242 * X_4)$$

b) Zero-Inflation Model Coefficient

$$\hat{\pi} = \frac{exp(-27.395 + 7.081 * X1 - 17.144 * X2 - 2.119 * X3 - 9.382 * X4)}{1 + exp(-27.395 + 7.081 * X1 - 17.144 * X2 - 2.119 * X3 - 9.382 * X4)}$$

D. Best Model Selection

The best model selection is done by looking at the AIC (Akaike Information Criterion) value. The selection of the best model is done by comparing the 2 models that have been formed, namely the Poisson Regression model and the Zero-Inflated Negative Binomial Model. The AIC values of the two models can be seen in Table VI.

TABLE VI. COMPARISON OF AIC VALUES

Regression Model	**AIC Values**
Regresi Poisson	817.4561
Zero-Inflated Negative Binomial (Full Model)	65.1253
Zero-Inflated Negative Binomial (Without X_5)	62.1329

The AIC value in Table VI shows that the lowest AIC value is the Zero-Inflated Negative Binomial model without a Variable Percentage of the population who has health insurance from private insurance. These variables were excluded because they had no significant effect on the Count Model and Zero-Inflation Model. Therefore, the Zero-Inflated Negative Binomial modeling was carried out without using the Variable Percentage of the population who had health insurance from private insurance.

E. Model Interpretation

The ZINB model is used to deal with overdispersion in the Poisson Regression model. The ZINB model is divided into two components, namely the count model for and the zero inflation model for. The interpretation of the model formed from ZINB is based on the odd ratio value as seen from the exp (β) value.

1) The interpretation of the count model coefficient is as follows :

a) The constant is 2.13703, meaning that if the variables are Percentage of poor people X_1 , Percentage of households using shared MCK facilities (X_2), Percentage of households using wood as the main fuel for cooking (X_3), and Percentage of population whose drinking water source comes from Pump Water $((X_4)$, is zero, then the COVID-19 mortality rate is worth exp(2.13703) = 8.474232.

b) The coefficient of $X_2$is 0.52405, meaning that every 1 percent increase in the percentage of poor people (X_2), will increase the COVID-19 mortality rate by exp (0.52405) = 1.688854 times the original COVID-19 mortality rate, if other variables are constant.

c) The coefficient of X_3 is -0.26860, meaning that every 1 percent increase in the percentage of households that use wood as the main fuel for cooking (X_3), it will reduce the COVID-19 mortality rate by exp (-0.26860) = 0.764449 times the original COVID-19 mortality rate, if the variable is else constant value.

d) The X_4 coefficient is 0.15242, meaning that every 1 percent increase in the Percentage of the Population whose drinking water source comes from Pump Water (X_4) will increase the COVID-19 mortality rate by exp (0.15242) = 1.164649 times the original COVID-19 mortality rate, if other variables are constant.

2) The interpretation of the zero inflation model coefficient is as follows:

a) The constant is -27.395 , meaning that if the variables are Percentage of poor people (X_1), Percentage of households using shared MCK facilities (X_2), Percentage of households using wood as the main fuel for cooking (X_3), and Percentage of population whose drinking water source comes from from the Air Pump (X_4), is zero, then the value of the COVID-19 mortality rate is exp(- 27.395) = 1.266E-12.

b) The coefficient of X_1 is 7.081, meaning that every 1 percent increase in the percentage of poor people (X_1), will increase the chance of the COVID-19 mortality rate to zero by exp (7.081) = 1189,157 times, if other variables are constant.

c) The coefficient of X_2 is -17,144, meaning that every 1 percent increase in the percentage of households using shared MCK facilities (X_2), will reduce the chance of the COVID-19 mortality rate to zero by exp (-17,144) = 3.58E-08 times, if other variables are constant.

d) The coefficient of X_3 is -2.119, meaning that every 1 percent increase in the percentage of households that use wood as the main fuel for cooking (X_3), it will reduce the chance of the COVID-19 mortality rate to zero by exp (-2.119) = 0.1201 times, if other variables are worth constant.

e) The coefficient of X_4, is -9.382, meaning that every 1 percent increase in the Percentage of the Population whose drinking water source comes from Pump Water (X_4) will reduce the chance of the COVID-19 mortality rate to zero by exp (-9.382) = 8.42E-05 times, if other variables are worth constant.

F. Parameter Significance Test Results

Based on Table VII, it can be concluded that in the count model there are 4 variables that have a significant effect on the COVID-19 mortality rate. In the zero-inflation model, there are no independent variables that affect the COVID-19 mortality rate. Based on the two models, it can be concluded that the variables used are not appropriate for the zero-inflation model.

TABLE VII. PARAMETER SIGNIFICANCE TEST RESULTS

Variable Name	**Count Model Coefficient**	***Zero-Inflation Coefficient***
Percentage of poor people (X_1)	Significant	Not significant
Percentage of households using shared toilet facilities (X_2)	Significant	Not significant
Percentage of households using wood as the main fuel for cooking (X_3)	Significant	Not significant
Percentage of Population whose drinking water source comes from Pump Water (X_4)	Significant	Not significant

IV. CONCLUSION

The conclusions obtained from this research are as follows the factors that influence the COVID-19 mortality rate in the count model are percentage of poor people ($X1$), the percentage of households using shared toilet facilities ($X2$), the percentage of households using wood as the main fuel for cooking ($X3$), and the percentage of the population whose drinking water source came from pumps (X_4). In the Zero-Inflation model, there are no factors that affect the COVID-19 mortality rate. so that the ZINB regression model used is the count model. Third, based on the evaluation of user satisfaction, the designed application has been able to help predict COVID-19 mortality and assist in providing information and insight to the public about COVID-19.

ACKNOWLEDGMENT

This work is supported by Research and Technology Transfer Office, Bina Nusantara University as a part of Bina Nusantara University's International Research Grant (PIB 2021) with contract number: 017/VR.RTT/III/2021.

REFERENCES

[1] Anggreainy, M.S., Musyaffa, H. Illyasu, A.M, "COVID-19 patient mortality risk classification using Linear Regression and Exponential Smooting Methods". 2021 International Seminar on Machine Learning, Optimization, and Data Science, ISMODE 2021, 2022, pp. 173-177.

[2] Mekuanint Simeneh Workie., Abebaw Gedef Azene. "Bayesian Zero-Inflated Regression Model with application to under five child mortality". Journal of Big Data volume 8, Article number : 4 (2021).

[3] Muggy David Cristian Ginzel, Asep Saefuddin, Erfiani, Determination of Extracurricular Activities for High School Students Based on Multiple Intelligence , Xplore: Journal of Statistics: Vol. 9 No. 1 (2020).

[4] Moh. Irfan Agus Saputro., M.Farhan Qudratullah. "Estimation of zero-inflated negative binomial regression parameters using the maximum likelihood method (case study : factors affectiong infant mortality in Wonogiri in 2015)". Proceeding International Conference on Science and Engineering (ICSE).2021-02-28.

[5] Shakhawat Hossain., Hatem A. Howlader. "Estimation Technique for Regression Model with Zero-inflated Poisson Data". International Journal of Statistics and Probability; Vol.4, No. 4; 2015.

[6] Preisser JS, Das K, Long DL, Stamm JW. A marginalized zero-inflated negative binomial regression model with overall exposure effects. The University of North Carolina at Chapel Hill Department of Biostatistics Technical Report Series. 2014;43.

[7] Pornpop Saengthong, Ampai Thongteeraparp, Winai Bodhisuwan."The zero inflated negative binomial – Crack distribution: Some properties and parameter estimation".Journal of Science and Technology · November 2015.

[8] Long DL, Preisser JS, Herring AH, Golin CE. A Marginalized Zero-Inflated Poisson Regression Model with Random Effects. *Journal of the Royal Statistical Society, Series C (Applied Statistics)* (published online April 30, 2015). DOI: 10.1111/rssc.12104.

[9] Made Susilawati, I Gusti Ayu Made Srinadi. "Perbandingan Regresi Zero Inflated Poisson (Zip) Dan Regresi Zero Inflated Negative Binomial (Zinb) Pada Data Overdispersion (Studi Kasus: Angka Kematian Ibu di Provinsi Bali). E-Jurnal Matematika 5(4):133. November 2016. DOI:10.24843/MTK.2016.v05.i04.p132.

[10] Puspita Rahayu, Kusman Sadik, Indahwati."Overdispersion study of poisson and zero-inflated poisson regression for some characteristics of the data on lamda, n, p"/ International Journal of Advances in Intelligent Informatics 2(3):140. November 2016. DOI:10.26555/ijain.v2i3.73.

[11] Drajat Indra Purnama. "Comparison of Zero Inflated Poisson (ZIP) Regression, Zero Inflated Negative Binomial Regression (ZINB) and Binomial Negative Hurdle Regression (HNB) to Model Daily Cigarette Consumption Data for Adult Population in Indonesia". Jurnal Matematika Statistika dan Komputasi 17(3):357-369. May 2021. DOI:10.20956/j.v17i3.12278.

[12] Osuji G. A., Okoro C. N., Obubu M., Obiora-Ilouno H. O."Effect of Akaike Information Criterion on Model Selection in Analyzing Auto-crash Variables".Effect of Akaike Information Criterion on Model Selection in Analyzing Auto-crash Variables. January 2016.

[13] Mulya Rahma Karyanti, Erni Juwita Nelwan, Iqbal Zein Assyidiqie, Hindra Irawan Satari, Sri Rezeki Hadinegoro. "Diphtheria Epidemiology in Indonesia during 2010-2017".Acta medica Indonesiana. 1 Jul 2019.

[14] Cindy Xin Feng. "A comparison of zero-inflated and hurdle models for modeling zero-inflated count data". Journal of Statistical Distributions and Applications volume 8, Article number: 8 (2021).

[15] Naufal Rifki Fauzan;Daniel Alexander;Muhammad Siraz Hafizh;Maria Susan Anggreainy."Covid-19 Vaccine Tweets - Sentiment Analysis". 2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI). DOI: 10.1109/ICCSAI53272.2021.9609712.

[16] Raden Aurelius Andhika Viadinugroho. "Calculate Maximum Likelihood Estimator with Newton-Raphson Method using R". Towards Data Science. 1 Maret 2021.

[17] Dewi Retno Sari Saputro, Ade Susanti, and Nafisa Berliana Indah Pratiwi."The handling of overdispersion on Poisson regression model with the generalized Poisson regression model". AIP Conference Proceedings Volume 2326, Issue 1, https://doi.org/10.1063/5.0040330.

Automatic Healthy Sperm Head Detection using Deep Learning

Ahmad Abdelaziz Mashaal[1], Mohamed A. A. Eldosoky[2], Lamia Nabil Mahdy[3], Kadry Ali Ezzat[4]
Biomedical Engineering Department, Faculty of Engineering, Helwan University, Cairo, Egypt[1, 2]
Biomedical Engineering Department, Higher Technological Institute, 10th of Ramadan City, Egypt[3, 4]

***Abstract*—Infertility is one of the diseases in which researchers are interested. Infertility disease is a global health concern, and andrologists are constantly looking for more advanced solutions for this disease. The intracytoplasmic sperm injection (ICSI) process is considered as one of the most common procedures for achieving fertilization. Sperm selection is performed using visual assessment which is dependent upon the skills of the laboratory technicians and as such prone to human errors. Therefore, an automatic detection system is needed for quick and more accurate results. This study utilizes a deep learning technique for the classification of heads of human sperms which indicate the healthy human sperms. The Convolutional Neural Network (CNN) model of visual Geometry Group of 16 layers (VGG16) was used for classification, and it is one of the best architectures used for image classification. The dataset consists of 1200 images of human sperm heads divided into healthy and unhealthy. Here, the VGG16 model is fine-tuned and achieved an accuracy of 97.92% and a sensitivity of 98.82%. Moreover, it achieved an F1 score of 98.53%. The model is an effective and real-time system for detecting healthy sperms that can be injected into eggs for achieving successful fertilization. This model quickly recognizes healthy sperms and makes the sperm selection process more accurate and easier for the andrologists.**

Keywords*—Infertility; sperm morphology; deep learning; human sperm head; healthy sperms

I. INTRODUCTION

The inability of couples to achieve pregnancy after 1 year of regular coition is known as infertility [1]. Infertility is the problem of most couples around the world, about 30-40% of them related with male factor abnormalities [2]. The most important role in the intracytoplasmic sperm injection method is the selection of the best sperms that can be injected into the oocytes. Embryologists select the best sperm depending on the morphology of sperms by visual assessment. This process is time-consuming and so tiring compared to machine learning approaches [3]. Machine learning depends on manual extraction of features such as head area, length, and width of sperms [4]–[6]. The human sperms are composed of head, midpiece and tail. The head of human sperm contains the deoxyribonucleic acid which carries the genetic instructions needed for reproduction. The midpiece part has the mitochondria to supply the tail with energy required for movement. The tail part provides the sperm with motility required for its movement to oocytes for fertilization [34]. The tip of head contains acrosome which secretes enzymes that are useful for penetration and make the penetration process easier for human sperms.

Intracytoplasmic sperm injection (ICSI) process is considered as an optimum choice with male infertility diseases. ICSI process is done when the sperm count is low, and it is also preferable when the sperms have low motility according to the semen analysis. In Vitro Fertilization (IVF) process is different from ICSI since the fertilization is performed by the sperms themselves in the test tube without injection with any needles after keeping the human sperms with oocytes inside one tube [35]. Intrauterine Insemination (IUI) process is less costly in which the semen is injected directly into the uterine cavity and this method increases the success rate of pregnancy with some cases [34].

The rest sections of this paper is organized as follows: Section 2 discusses the literature review. Section 3 describes the methodology, dataset, and explanation of preprocessing stages in detail. Section 4 explains the proposed model for healthy sperm detection. Section 5 shows and discuss the results of this proposed model. Section 6 concludes the paper and mentions the scope of future work.

II. LITERATURE REVIEW

There are several researches about automatic detection of the best sperms used for the intracytoplasmic sperm injection process using machine learning and deep learning. In one of these studies, J Riordon et al. improved a model for classification of sperms using VGG16 using versatile dataset with accuracy reached 94.1% [9]. Javadi et al. proposed a method for sperm evaluation with 3 labels which are acrosome, head and vacuole [10]. Mirroshandel et al. improved automatic detection algorithm for human sperm images with accuracy reached 93.2% with tail and neck [11]. Revollo, Natalia V., et al. improved an automatic system for human sperm head detection using machine learning techniques in which the sperms were labeled manually achieving an accuracy of 92% [36]. Bijar et al. proposed a model for segmentation of different parts of sperms with high accuracy [12]. Erfan Miahi et al. developed a neural architecture search based on genetic algorithm for detection of different parts of human sperm with an accuracy of 91.66% and 77.33% in the vacuole and head abnormality detection, respectively [37]. Prabaharan, L., and A. Raghunathan. proposed a convolutional neural network for abnormal sperm detection based on morphology achieving an accuracy of 98.99% [38].

There are some researchers who used K-nearest neighbor and principal component for sperm analysis [7]. Several researches have proven the relationship between ICSI success and the morphology of the sperm head [8]. In some studies, stained sperms were used, and they are not useful for the intracytoplasmic sperm injection (ICSI) process, but this method has the advantage of detecting healthy sperms easily and it is more suitable for low quality and noisy images [14].

There are also many researchers used versatile types of deep learning architectures; liu et al. used AlexNet architecture for classification of human sperms [13]. McCallum et al. improved a VGG16 model for detecting the best sperms of highest integrity of DNA [14]. Ghasemian et al. proposed a model for recognition of sperm abnormality with low computation time [15]. Some studies used segmentation procedures, shaker et al. proposed a method for detection and segmentation of human sperms achieving accuracy of 92% for sperm head segmentation [16]. Urbano et al. presented a model for tracking of human sperms [17]. Mirsky et al. developed a system for healthy and unhealthy sperm detection based on morphology using machine learning with precision of 90% [18]. Thirumalaraju et al. presented a deep learning model with accuracy of 89% in recognition of good sperms [19]. These algorithms could automate the process of sperm recognition with its types [14]. For computer vision tasks, there are many algorithms for transfer learning performed by researchers to find out the best processes for the image classifiers [22], [23]. The design of the neural network affects the accuracy of the classification, depending also on the type of dataset. Baker et al. used reinforcement learning in designing neural network architectures [24], zoph et al. also updated a neural network architecture using reinforcement learning [25]. Zoph et al. used versatile methods for producing the best architectures for classification of images and proposed learning architectures for image recognition [21].

Mathematical morphology is so important for sperm recognition and detection of healthy and unhealthy sperms. Rodríguez et al. used Lambertian model and sperm cell segmentation based on mathematical morphology [28]. There are several research studies that are interested in counting the sperms using versatile methods. One of these works is the method of Susrama, I. G., K. E. Purnama, and M. H. Purnomo, which proposed a model for human sperm concentration and number using Otsu's threshold obtaining an accuracy of 91% [29]. There are some models for processing microscopic videos for detecting the motility of sperms [30]. It is helpful for tracking of the spermatozoa through the videos and this analysis is needed for motility assessment. Boumaza et al. improved an automatic system for concentration assessment based on decision trees machine learning [31]. Ilhan et al. presented an approach for analyzing spermiogram tests with high correlation with visual assessment results and proposed an algorithm for motile sperm detection [32]. Lv, Qixian, et al. proposed an algorithm for automatic segmentation of sperm head based on the U-Net network [39]. Chandra, Satish, et al. utilized different deep learning models, and both ResNet50 and VGG19 models achieved an accuracy of 71%, 87.33%, 73% for acrosome, vacuole and head label, respectively [40].

The intracytoplasmic morphologically selected sperm injection (IMSI) usually is done using magnification of 10,000 and it has the property of motile spermatozoa selection properly [26]. Several researches demonstrated the effectiveness of IMSI over ICSI process due to the power of magnification of IMSI process and many factors that differ from patient to another. So, IMSI process in some cases showed clinically significant difference [26]. There are many factors that influence the success of intracytoplasmic sperm injection process, the World Health Organization (WHO) defined the criteria of reference values for the characteristics of semen. The reference volume of semen is 1.5 mL, reference sperm count is 15 million per mL and minimum total motility of 40% [27], [34]. The abnormalities of sperm heads can cause failure of fertilization, and then failure of ICSI process [8].

III. Material and Methods

A. Sperm Head Dataset

The Dataset used is freely available for public, collected from Hannam Fertility Centre in Canada and utilized by McCallum et al [14]. Some processing work was done on dataset before using, and then the dataset has been classified into healthy and unhealthy human sperm heads with the help of andrologists as shown some of dataset in Fig. 2. Google Colab in python is mainly used for implementing this paper with the aid of MATLAB program. Dataset is partitioned into training set, test set and validation set. The dataset consists of 1200 images of size 150x150 pixels and partitioned as 720 images for training, 240 images for test purpose and 240 images for validation purpose, as represented in Fig. 1, and Table I.

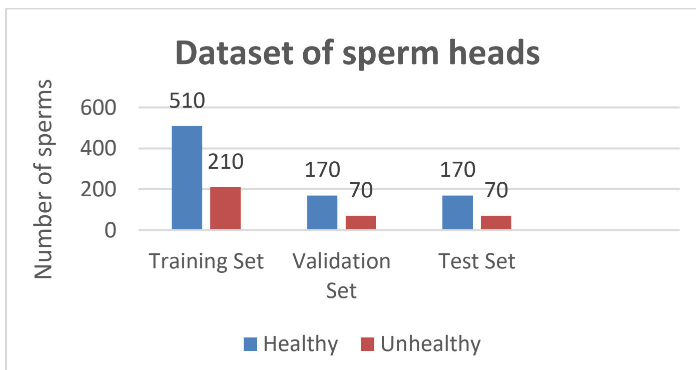


Fig. 1. Overview of Sperm Head Dataset with Healthy and Unhealthy Classes.

TABLE I. Partitioning of Dataset

Dataset Partitions	Number of Images	Percentage
Training Dataset	720	60%
Validation Dataset	240	20%
Test Dataset	240	20%
Total Dataset	1200	100%

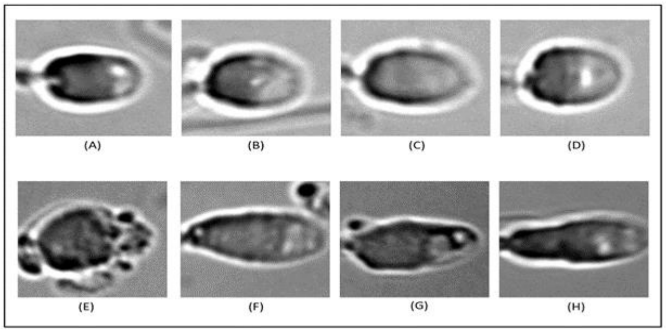


Fig. 2. A Selection of Sperm Head Dataset. (A), (B), (C), (D) Healthy Sperm Head Images. (E), (F), (G), (H) Unhealthy Sperm Head Images.

B. Image Preprocessing

Preprocessing of images is an important step for image enhancement (e.g., noise removal), the median filter is used for this purpose [33] as represented in Fig. 3. Image resizing also is necessary in the preprocessing steps. Dataset images were scaled up to 224x224 pixels according to VGG16 requirements. Image resizing has an important role in image processing technique to enlarge or reduce the dataset. Image interpolation can be divided into two different ways, both image down sampling and image up sampling are necessary when resizing the dataset for matching the requirements of the deep learning model and differ according to each model and the field of study.

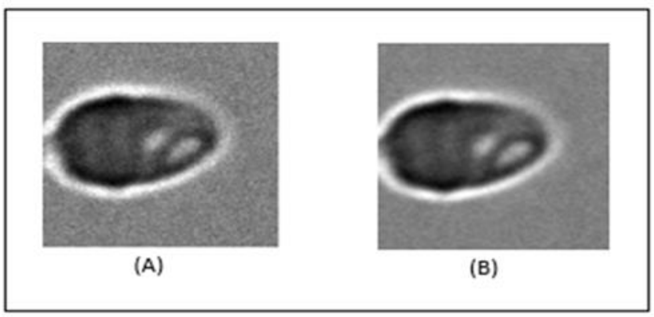


Fig. 3. Preprocessing of Input Images. (A) Noisy Image before Preprocessing. (B) Image after Applying Median Filter for Noise Removal.

The main preprocessing methods used after denoising process are explained as the following.

1) Otsu's thresholding process: In this technique, the thresholding on grayscale image is processed. Otsu's method is used to perform automatic image thresholding by obtaining the image histogram and computing the threshold value (t) then the image pixels is replaced into white or black depending on the value of the threshold (t) [29]. The image pixels whose saturation is greater the threshold is replaced into white and image pixels whose saturation is lower than the threshold is replaced into black color. It searches for the threshold that reduces the intra-class variance as shown.

$$\sigma_\omega^2(t) = \omega_0(t)\sigma_0^2(t) + \omega_1(t)\sigma_1^2(t) \quad (1)$$

where σ_ω^2 is the intra-class variance, ω_0 and ω_1 are the probabilities of the two classes separated by a threshold (t) and σ_0^2 and σ_1^2 are the variances of these two classes.

This method is considered as one of segmentation techniques in image processing and this algorithm can be described as the following.

Otsu's thresholding algorithm:

a) Given grayscale image.

b) Calculate the histogram of this image.

c) Computing the probabilities of each intensity level

d) Set up initial ω_i and μ_i at t = 0

e) Calculate ω_i and μ_i at all possible thresholds. (t = 1, t = 2, … till the maximum intensity)

f) Calculate $\sigma_b^2(t)$ as the following equation:

$$\sigma_b^2(t) = \sigma^2 - \sigma_\omega^2(t) = \omega_0(t)\omega_1(t)[\mu_0(t) - \mu_1(t)]^2 \quad (2)$$

where σ_b^2 is the inter-class variance, ω is the class probabilities and μ is the class means.

g) Choose the best threshold value for optimum results corresponding to the maximum $\sigma_b^2(t)$ and apply this threshold to image [29].

2) Area opening method: Area opening technique is used for removal of objects with area smaller than a specified parameter from the foreground of binary images [33]. Mathematically, opening a set A by a structuring element B can be given by the following equation.

$$A \circ B = (A \ominus B) \oplus B \quad (3)$$

where the symbol $\ominus$ denotes to erosion and $\oplus$ denotes to dilation.

Area opening technique is utilized for removing vacuoles that may appear in the head of sperms in order not to affect the accuracy of the classifier of the proposed model as shown in Fig. 4, and this method contributes to get high accuracy.

3) Image complement process: The complement of a black and white image is converting the zeros to ones and ones to zeros [33]. Image complement is performed for getting white sperm head and black background as shown in Fig. 5. The image complement is used as a final step before training in which data augmentation is done through.

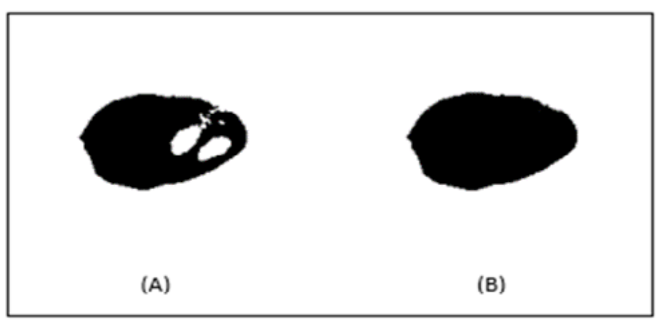


Fig. 4. Area Opening of Input Images. (A) Image after Applying Otsu's Thresholding Segmentation Method and before Applying Area Opening. (B) Image after Applying Area Opening for Removing Vacuoles in Sperm Head.

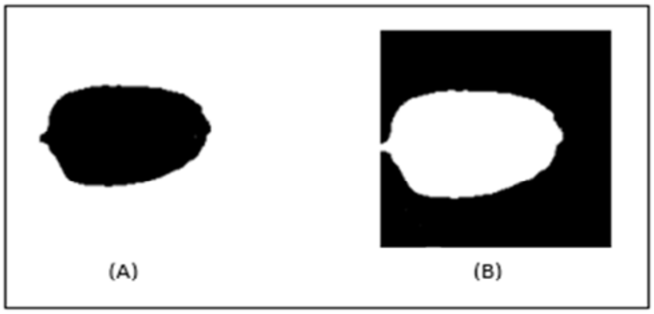


Fig. 5. Image Complement of Input Images. (A) Image after Applying Area Opening Process. (B) Image after the Complement Process.

4) Data augmentation: Image data augmentation is a technique used to increase the dataset by modifying images. Image data augmentation is performed only on training dataset. This is done by using ImageDataGenerator that is provided by Keras deep learning library. ImageDataGenerator class is used for augmentation of images with some parameters of image augmentation (e.g., shift, zoom and rotation) as represented in Table II. Some of these modifications for each parameter and its effect on the original images are shown in Fig. 6.

TABLE II. PARAMETERS OF DATA AUGMENTATION

Parameters	**Values**
Rescale	1./255
Width shift range	0.2
Height shift range	0.2
Horizontal flip	True
Zoom	0.25
Rotation	-30°
Image fill	"Nearest"

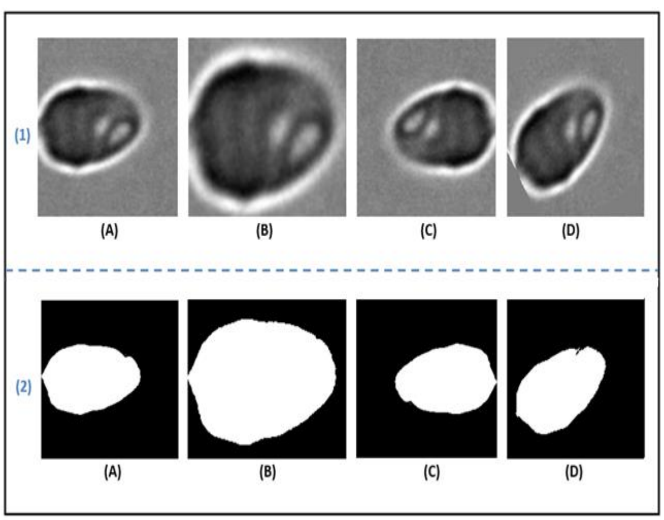


Fig. 6. Some Types of Image Augmentation Applied to Dataset. The First Row in Group (1) Including the Raw Dataset before Segmentation and the Second Row in Group (2) Including the Images after Segmentation (A) Original Image without Augmentation. (B) Image with Zoom Augmentation. (C) Image with Horizontal Flip. (D) Image with -30° Rotation.

IV. PROPOSED HEALTHY SPERM HEAD DETECTION MODEL

The deep learning architecture VGG16 is a convolutional neural network model proposed by Simonyan et al. [20] and pre-trained on ImageNet. The size of the input image of this architecture is fixed 224x224x3 then the image is passed through a stack of convolutional layers and the filter used of size 3x3 [9]. At first it has two convolution layers with pooling layer then another two convolution layers with pooling followed by three convolution layers with pooling repeated three times and finally three layers of fully connected (FC) attached with the output layer (e.g., sigmoid function) as shown in Fig. 7. In this proposed model, new layers are considered. Flatten layer, dense, dropout, dense and dropout layer are added respectively, and then dense layer with sigmoid activation function is used as represented in Fig. 8.

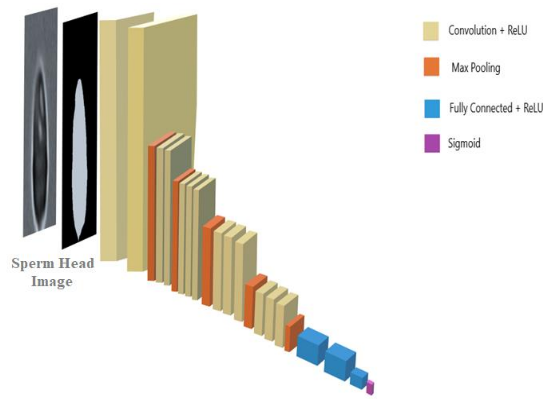


Fig. 7. General Representation of VGG16 Architecture.

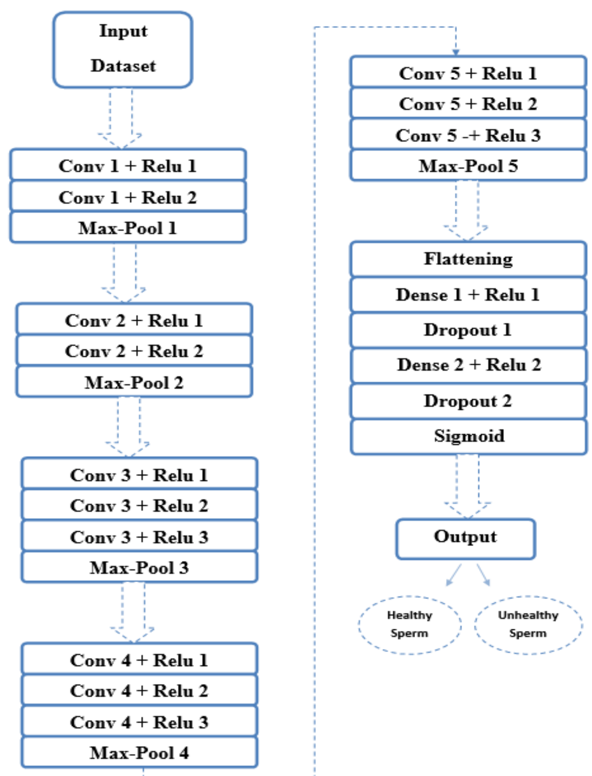


Fig. 8. Fine-Tuned Proposed Model for Healthy Sperm Head Detection.

The features of sperm head were extracted by convolutional layers, each convolutional layer has the same size of kernels [10]. Given an image H of size (u, v), the convolution can be estimated as follows,

$$G(u,v) = (H * Q)(u,v) = \sum_x \sum_z M(u - X, v - Z) Q(x,z) \quad (4)$$

where Q is the convolution kernel of size x, G is the feature map of the output, H is the feature map of the input and (u, v) is the size of image.

In this proposed model, binary cross entropy loss function is performed since we have only 2 output categories, binary cross entropy loss function is used during adjusting the weights of the model, and then the loss can be minimized [10], this function is effective for optimizing the model and it is represented as the following formula.

$$E = -\frac{1}{n} \sum_{i=1}^{n} [\, y_i \cdot log(\hat{y}_i) + (1 - y_i) \cdot log(1 - \hat{y}_i) \,] \quad (5)$$

where n is the number of scalar values in the model output, y_i is the true label, $\hat{y}_i$ is the predicted label.

The proposed algorithm demonstrated its effectiveness in classification of healthy and unhealthy sperm head then detecting the best healthy sperm head which indicates the best healthy sperm that can be used for successful fertilization as proved in andrology [34]. The main steps of the convolution and the VGG16 algorithm can be summarized as the following.

1) **Steps of Convolution algorithm are:**

a)	The first step is to multiply the filter with the region of the input image of the same size using the convolution formula.
b)	Each element is multiplied with an element in the corresponding location then the results will be summed and give one output value.
c)	Repeating these steps by sliding the filter across the image until getting the final output.

2) **Main steps of VGG16 algorithm are:**

a)	2D Convolution as previously described.
b)	Max Pooling. This operation is used for down sampling feature maps by calculating the maximum value in each patch of a feature map.
c)	Flattening, converting pooled feature map to a one-dimensional array (vector).
d)	Dense layer. It's a layer that consists of neurons that detect input from neurons that's in the previous layers.
e)	Dropout Layer. Used to prevent a model from overfitting.
f)	Activation function that finally used for defining the output [10]. In the last layer, the sigmoid function is used for the classification of sperms healthy or unhealthy. The sigmoid function is expressed as the following, $$S(x) = \frac{1}{1+e^{-x}} \quad (6)$$ where $S(x)$ is the sigmoid function of a variable x

In this paper, the best sperms could be recognized using this proposed model. At the first, the dataset was uploaded. The preprocessing steps were done to be handled by VGG16 and they were effective in this process. The noises in the image were removed using median filter, the sperm heads were segmented using Otsu's thresholding process, area opening was implemented for removing objects in the heads, the images were complemented, and then the model training and evaluation were done. Consequently, any image can be inserted after preprocessing for detection as shown in the proposed model flowchart in Fig. 9. There are also preprocessing steps should be done to the input images.

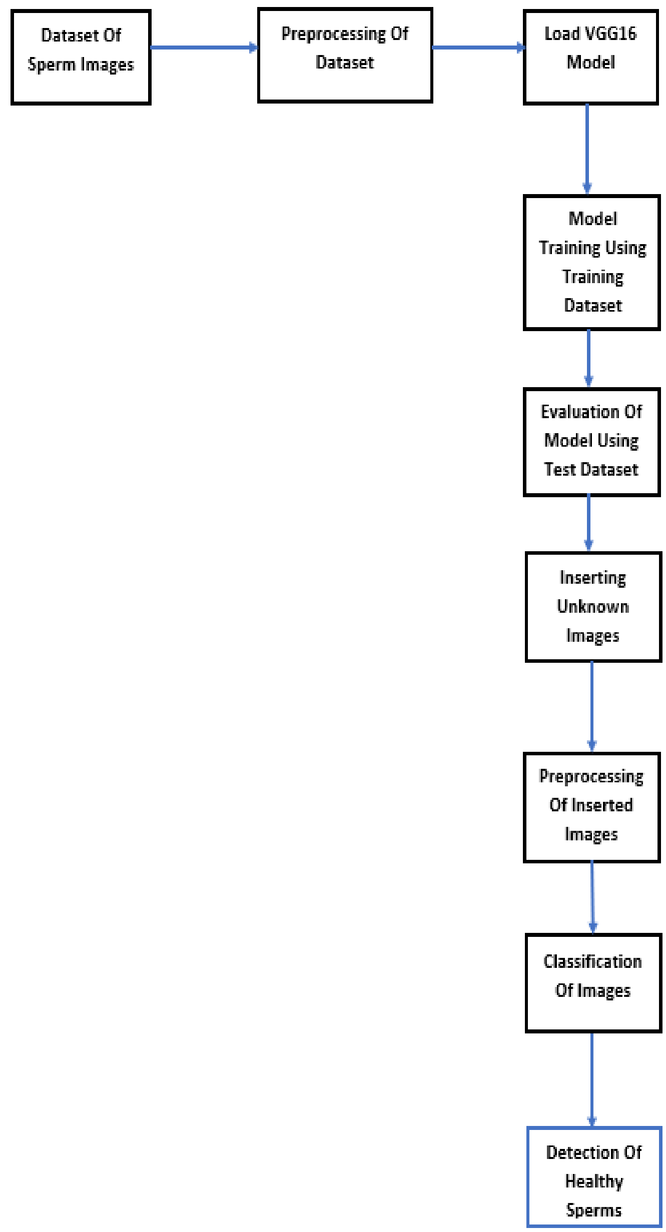


Fig. 9. Flowchart of the Proposed Model for Healthy Sperm Detection.

V. Results and Discussion

The proposed model for classification of sperm heads is effective after using 1200 images of healthy and unhealthy sperm heads resulting in high accuracy. This proposed model is trained on a large dataset, and it is a real time system for andrologists for classification of healthy and unhealthy sperms. In Fig. 10, the results of the proposed model are shown for 4 unknown images of healthy and unhealthy sperm heads which are detected accurately. In this study, the median filter demonstrated the effectiveness of image denoising. Otsu's thresholding process is done for segmentation then area opening technique is done and both have proved their success with the dataset. The accuracy of model reaches 97.92%, sensitivity equals 98.82%, precision equals 98.25%, specificity equals 95.71% and F1 Score equals 98.53% as shown in Table IV. The following formulas [7]–[11] respectively represents Accuracy, Sensitivity, Precision, Specificity and F1 Score. The following confusion matrix shows the values of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) [11]. The confusion matrix of our study is shown in Table III with 168 true positive and 67 true negative. The performance of our model is shown in Fig. 11 with high training and validation accuracy after training with 200 epochs respectively reached 99.51% and 98.48%. The training and validation loss respectively reached 0.9817 and 1.4251.

TABLE III. The Confusion Matrix on the Test Set

		Predicted	
		Healthy	***Unhealthy***
Actual	***Healthy***	TP = 168	FN = 2
	Unhealthy	FP = 3	TN = 67

$$Accuracy = \frac{TN+TP}{TN+TP+FP+FN} \tag{7}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{8}$$

$$Precision = \frac{TP}{TP+FP} \tag{9}$$

$$Specificity = \frac{TN}{TN+FP} \tag{10}$$

$$F1\ Score = \frac{2*\ Precision*\ Sensitivity}{(Precision+Sensitivity)} \tag{11}$$

TABLE IV. Evaluation Parameters for the Proposed Model

Classifier	Accuracy	Sensitivity	Precision	Specificity	F1 Score
Proposed Method	97.92%	98.82%	98.25%	95.71%	98.53%

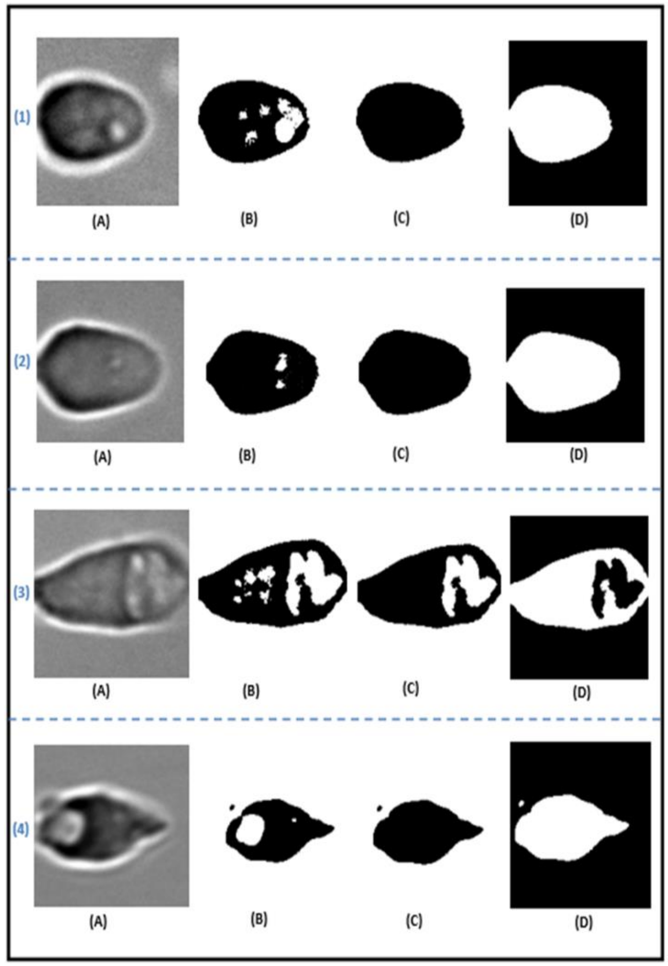


Fig. 10. Results of the Proposed Model with Preprocessing Steps of Four Images in which the Images in Groups (1) and (2) are Healthy Sperm Heads and the Images in Groups (3) and (4) are Unhealthy Sperm Heads. (A) Original Images before Segmentation. (B) Images after Applying Otsu's Thresholding Segmentation. (C) Images after Applying Area Opening. (D) Images after the Complement Method.

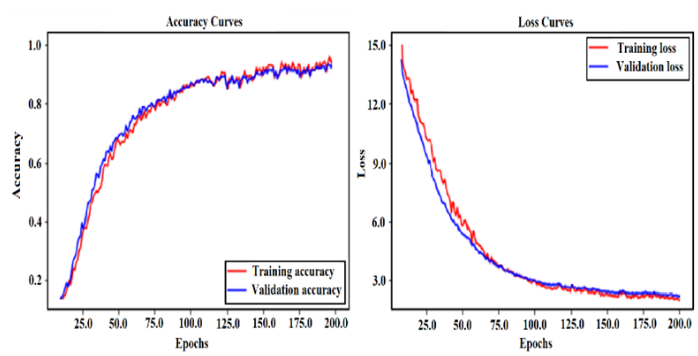


Fig. 11. Performance of the Proposed Model.

VI. Conclusion and Future Work

In conclusion, the best sperms required for the ICSI process can be selected using the proposed deep learning model. Laboratory specialists always look for simple and precise methods for healthy sperm recognition for injection into oocytes process. This paper proposed a deep learning method based on VGG16 and its results have been approved by the andrologists. In this article, deep learning approach demonstrated the effectiveness of healthy sperm detection rather than machine learning. The fine-tuned VGG16 model resulting in high accuracy of 97.92% and this is one of the main advantages of this proposed model compared to the results of other models in literature review section. The VGG16 transfer learning model has proven its computationally efficiency for sperm head recognition. Assisted reproductive technology (ART) is necessary for finding more tools for the detection of the best sperms for increasing fertilization, pregnancy, and live birth rates. This proposed approach demonstrated its advantages of getting rapid results with high accuracy for detecting the healthy sperm heads which indicate to the healthy sperms that can be used in the intracytoplasmic sperm injection process.

For future work, different deep learning architectures can be used with a comparative study among them with using versatile dataset for better and more accurate evaluation. Taking another part of sperm into consideration with different preprocessing methods can be useful.

References

[1] Isidori, Aldo, Maurizio Latini, and Francesco Romanelli. “Treatment of male infertility.” Contraception 72.4 (2005): 314-318.

[2] Stouffs, Katrien, et al. “Is there a role for the nuclear export factor 2 gene in male infertility?.” Fertility and sterility 90.5 (2008): 1787-1791.

[3] Freund, M. “Standards for the rating of human sperm morphology. A cooperative study.” International journal of fertility 11.1 (1966): 97-180.

[4] Chang, Violeta, et al. “Gold-standard for computer-assisted morphological sperm analysis.” Computers in biology and medicine 83 (2017): 143-150.

[5] Yi, W. J., K. S. Park, and J. S. Paick. “Parameterized characterization of elliptic sperm heads using Fourier representation and wavelet transform.” Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol. 20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No. 98CH36286). Vol. 2. IEEE, 1998.

[6] Chang, Violeta, et al. “Automatic classification of human sperm head morphology.” Computers in biology and medicine 84 (2017): 205-216.

[7] Li, Jiaqian, et al. “Human sperm health diagnosis with principal component analysis and K-nearest neighbor algorithm.” 2014 International Conference on Medical Biometrics. IEEE, 2014.

[8] Monte, Giuseppe Lo, et al. “Focus on intracytoplasmic morphologically selected sperm injection (IMSI): a mini-review.” Asian journal of andrology 15.5 (2013): 608.

[9] Riordon, Jason, Christopher McCallum, and David Sinton. “Deep learning for the classification of human sperm.” Computers in biology and medicine 111 (2019): 103342.

[10] Javadi, Soroush, and Seyed Abolghasem Mirroshandel. “A novel deep learning method for automatic assessment of human sperm images.” Computers in biology and medicine 109 (2019): 182-194.

[11] Mirroshandel, Seyed Abolghasem, and Fatemeh Ghasemian. “Automated morphology detection from human sperm images.” Intracytoplasmic sperm injection. Springer, Cham, 2018. 99-122.

[12] Bijar, Ahmad, et al. “Segmentation of sperm's Acrosome, nucleus and mid-piece in microscopic images of stained human semen smear.” 2012 8th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP). IEEE, 2012.

[13] Liu, Rui, et al. “Automatic Microscopy Analysis with Transfer Learning for Classification of Human Sperm.” Applied Sciences 11.12 (2021): 5369.

[14] McCallum, Christopher, et al. “Deep learning-based selection of human sperm with high DNA integrity.” Communications biology 2.1 (2019): 1-10.

[15] Ghasemian, Fatemeh, et al. “An efficient method for automatic morphological abnormality detection from human sperm images.” Computer methods and programs in biomedicine 122.3 (2015): 409-420.

[16] Shaker, Fariba, S. Amirhassan Monadjemi, and Ahmad Reza Naghsh-Nilchi. “Automatic detection and segmentation of sperm head, acrosome and nucleus in microscopic images of human semen smears.” Computer methods and programs in biomedicine 132 (2016): 11-20.

[17] Urbano, Leonardo F., et al. “Automatic tracking and motility analysis of human sperm in time-lapse images.” IEEE transactions on medical imaging 36.3 (2016): 792-801.

[18] Mirsky, Simcha K., et al. “Automated analysis of individual sperm cells using stain-free interferometric phase microscopy and machine learning.” Cytometry Part A 91.9 (2017): 893-900.

[19] Thirumalaraju, Prudhvi, et al. “Automated sperm morpshology testing using artificial intelligence.” Fertility and sterility 110.4 (2018): e432.

[20] Simonyan, Karen, and Andrew Zisserman, A. “Very deep convolutional networks for largescale image recognition.” arXiv 1–14, https://doi.org/10.1016/j.infsof.2008.09.005 (2014).

[21] Zoph, Barret, et al. “Learning transferable architectures for scalable image recognition.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[22] Cai, Han, et al. “Efficient architecture search by network transformation.” Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 32. No. 1. 2018.

[23] Real, Esteban, et al. “Large-scale evolution of image classifiers.” International Conference on Machine Learning. PMLR, 2017.

[24] Baker, Bowen, et al. “Designing neural network architectures using reinforcement learning.” arXiv preprint arXiv:1611.02167 (2016).

[25] Zoph, B., Le, Q.V. “Neural architecture search with reinforcement learning.” arXiv preprint arXiv:1611.01578 (2016).

[26] Babayev, Samir N., Chan Woo Park, and Orhan Bukulmez. “Intracytoplasmic sperm injection indications: how rigorous?.” Seminars in reproductive medicine. Vol. 32. No. 04. Thieme Medical Publishers, 2014.

[27] Cooper, Trevor G., et al. “World Health Organization reference values for human semen characteristics.” Human reproduction update 16.3 (2010): 231-245.

[28] Medina-Rodríguez, Rosario, et al. “Sperm cells segmentation in micrographic images through lambertian reflectance model.” International Conference on Computer Analysis of Images and Patterns. Springer, Cham, 2015.

[29] Susrama, I. G., K. E. Purnama, and M. H. Purnomo. “Automated analysis of human sperm number and concentration (oligospermia) using otsu threshold method and labelling.” IOP Conference Series: Materials Science and Engineering. Vol. 105. No. 1. IOP Publishing, 2016.

[30] Haugen, Trine B., et al. “Visem: A multimodal video dataset of human spermatozoa.” Proceedings of the 10th ACM Multimedia Systems Conference. 2019.

[31] Boumaza, Karima, Abdelhamid Loukil, and Kaouthar Aarizou. “Automatic human sperm concentrartion in microscopic videos.” Medical Technologies Journal 2.4 (2018): 301-307.

[32] Ilhan, Hamza Osman, and Nizamettin Aydin. “A novel data acquisition and analyzing approach to spermiogram tests.” Biomedical Signal Processing and Control 41 (2018): 129-139.

[33] Zhou, Huiyu, Jiahua Wu, and Jianguo Zhang. Digital Image Processing: Part II. Bookboon, 2010.

[34] World Health Organization. “WHO laboratory manual for the examination and processing of human semen.” (2010).

[35] Rodriguez-Wallberg, Kenny A., Hanna P. Nilsson, and Tekla Lind. “Live birth and pregnancy rates after in vitro fertilization/intracytoplasmic sperm injection in women with previous unilateral oophorectomy: a systematic review and meta analysis.” Fertility and Sterility (2022).

[36] Revollo, Natalia V., et al. “Supervised Machine Learning Classification of Human Sperm Head Based on Morphological Features.” Trends and Advancements of Image Processing and Its Applications. Springer, Cham, 2022. 177-191.

[37] Miahi, Erfan, Seyed Abolghasem Mirroshandel, and Alexis Nasr. “Genetic Neural Architecture Search for automatic assessment of human sperm images.” Expert Systems with Applications 188 (2022): 115937.

[38] Prabaharan, L., and A. Raghunathan. “An improved convolutional neural network for abnormality detection and segmentation from human sperm images.” Journal of Ambient Intelligence and Humanized Computing 12.3 (2021): 3341-3352.

[39] Lv, Qixian, et al. “An Improved U-Net for Human Sperm Head Segmentation.” Neural Processing Letters 54.1 (2022): 537-557.

[40] Chandra, Satish, et al. “Prolificacy Assessment of Spermatozoan via state-of-the-art Deep Learning Frameworks.” IEEE Access 10 (2022): 13715-13727.

RTL Design and Testing Methodology for UHF RFID Passive Tag Baseband-Processor

Enabling Internet-of-Things (IoT) Technology

Syifaul Fuada[1]
Program Studi Sistem Telekomunikasi
Universitas Pendidikan Indonesia
Bandung, Indonesia

Trio Adiono[3]
Electrical Engineering Department
School of Electrical Engineering and Informatics
Institut Teknologi Bandung, Bandung, Indonesia

Aris Agung Pribadi[2]
University Center of Excellence on Microelectronics Insitutut Teknologi Bandung, Bandung, Indonesia

Tengku Ahmad Madya[4]
Xirka Silicon Technology Ltd,
Bandung, Indonesia

Abstract*—With the rapid growth and widespread implementation of Internet-of-Things (IoT) technology, Radio Frequency Identification (RFID) has become a vital supporting technology to enable it. Various researchers have studied the design of digital or analog blocks for RFID readers. However, most of these works did not provide a comprehensive design methodology. Hence, the motivation of this study is to full fill the research gap. This paper proposes a comprehensive design and testing methodology for the Ultrahigh Frequency (UHF) RFID passive tag baseband processor at the register transfer (RTL). A complete design procedure of each block from state diagram to schematic level is presented; it comprises several blocks, *i.e.*, transmitter, receiver, Cyclic Redundancy Check (CRC), command processing, and Pseudorandom Number Generator (PRNG). Each block produces low latency (<400 ns). Two CRCs were applied to this system for different purpose: CRC-5 and CRC-16. To perform multi-parameter combinations of as many as 1344 combinations (including timing parameter, query respond, state transition, and BLF), a Universal Verification Methodology (UVM)-based test is conducted. The simulation results reveal that the proposed RFID baseband processor passes all the testing scenarios using UVM (version 1.1d). Moreover, we also implemented the proposed design on the FPGA board (ALTERA DE2-115). The system consumes 976 logic elements and 173.14 mW of total power dissipation (*i.e.*, 0.13 mW of dynamic power dissipation, 98.6 mW of static power dissipation, and 74.34 mW of I/O dissipation), which is reasonably low. This demonstrates that our design is synthesizable and ready to be processed further. All system design and test criteria were conducted following the EPC Gen-2 standard. The developed chip can be a solution for various kinds of RFID chip-based IoT applications.

Keywords*—*UHF RFID passive tag; baseband processor; register transfer level; universal verification methodology; Internet-of-things enabler; FPGA

I. INTRODUCTION

Within the last few decades, automatic identification system has become an essential part of industrial applications, such as logistics, retails, and manufacturing [1]–[4]. Now, with the rise of the Internet of Things (IoT), the use of Radio Frequency Identification (RFID) technology has a crucial role [5]. The RFID has been applied in various fields, *e.g.*, smart agriculture, smart homes, health care, medicines, transportation sector, payment, environment monitoring, disaster warning, or even telepresence for distant objects monitoring [6]–[13]. RFID is a communications technology that uses electromagnetic fields to detect tags attached to objects wirelessly. RFID technology is preferred over other identification or wireless communication technologies due to its various advantages, including bigger information storage capacity, more comprehensive coverage, better security system, more affordable price, and a universal standard. RFID is usually designed under the EPC Gen-2 standard, then adopted into the ISO 18000-6C standard. Even though the typical reading distance of an RFID tag is relatively long, about 6 meters, most of the tags are passive, which means that they do not have a self-powered source [14]. Instead, it solely relies on a converted power obtained from the RF signal sent by the reader. The analog circuit part is responsible for this power conversion or generation. The tag will then use the generated power, especially the baseband processor, to run the whole system. The RFID tag design has been finalized and matures in the industry for many years. With the detailed technical description in the Gen-2 tag standard, it is effortless to come up with a Register-Transfer Level (RTL) design. However, further observation and exploration are still an exciting topic by many researchers worldwide to fulfill numerous requirements of the recent issues and use cases [15]–[18], such as IoT applications based on-chip technology.

Flourishing along with the emergence of IoT technology, various research projects have been conducted related to RFID tags or readers, including the analog and digital circuit building blocks. The baseband processor is part of the digital block. Most research on UHF RFID baseband processors explores low power/energy-efficient design. For example, Wei et al. [19] proposed a low-power baseband processor using several techniques, including low operating frequency clock, clock gating, and asynchronous design. Using these techniques, they can achieve a low power consumption baseband processor as low as 2.7 µW. Similarly, Lee et al. [20] also proposed a low-power baseband processor. To achieve a low power consumption, they use numerous techniques, including the lowest possible clock frequency, latch-based clock gating, variable clock frequencies, and resource combining and sharing between blocks. Their design achieves a rather high-power consumption, 29.2 µW. However, this is due to their circuit also including the analog part. Despite their advanced circuit design, they did not provide a comprehensive design and testing methodology for the baseband processor. Ismail et al. [21] presented a comprehensive baseband processor design methodology, including the Finite State Machine (FSM). However, they did not describe the command processor design and the system test. Moreover, the testing only showed the signal encoding and decoding test results. There is no other parameter for the tests was shown. Indeed, to perform many test parameters, an additional test technique or tool is needed so that a large number of test combinations can be efficiently performed. Su et al. [22] proposed a novel automatic verification strategy for a UHF RFID baseband processor. Using their verification, they could perform a set of 2700 commands testing in total. In detail, the verification or test included several command responses tests, e.g., Ready, Arbitrate, Reply, Acknowledge, and random command test. Li et al. proposed an RFID tag IC that is fabricated on 0.13 µm CMOS technology. They achieved low power consumption: 4.8 µW for reading operations and 11.5 µW for writing operations [23]. Bhanushali, et al., successfully realized a digital UHF RFID Tag IC compatible with EPC Gen2 standard applied on 55 nm CMOS technology [16]. However, in their work, no comprehensive design methodology was provided. Design, implementation, verification methodology, and testing techniques is one set discussion that essential to be reported [24]; this will gain benefit, such as it can provide insights to early the designers/engineers who want to follow the proposed design.

This work presents a comprehensive and systematic RTL design and testing methodology specifically for UHF RFID passive tag baseband processors to fill the research gap. The baseband processor design includes receiver, transmitter, and command processing block. Before designing the schematic, the state diagram of each block is investigated to ensure all possible states are included. Then, each block and the design of the integrated block is tested based on the EPC Gen-2 standard to verify and ensure the functionality of the design. The latency of each block is also calculated. For the testing, numerous tests are performed, including query command, timing parameter, state transition, and Backscattering Link Frequency (BLF) test. To conduct the tests efficiently, a Universal Verification Methodology (UVM) is used as the test methodology. Using this UVM, a set of tests, as many as 1344 test combinations, was conducted, and thus, a rigorous parameter test was able to be performed. Additionally, the RTL design is also implemented on the FPGA development board to show the practicability of the design for future real chip implementation. Having this comprehensive RTL design and methodology, a more reliable baseband processor design can be realized. The developed chip can further be a strong candidate for IoT applications.

Our main contribution is providing a comprehensive RTL design and testing methodology. It includes each block's state diagram and schematic design, comprising a receiver, transmitter, command processing, Pseudorandom Number Generator (PRNG), Cyclic Redundancy Check (CRC), and their corresponding functional tests. In addition, the integrated system test is performed using UVM to perform the tests efficiently. The rest of this paper is organized as follows. To provide a short technical background and point out the distinct contributions, a description of related works is elaborated in Section I. The details of system design, starting from each block to its integration, are provided in Section II. Section III shows the testing results of the design, including an integrated system test using UVM, the implementation design, and the test of the baseband processor RTL design on the FPGA board. Lastly, Section IV provides the conclusion of this work.

II. Methodology

The structure of the proposed RFID tag system is depicted in Fig. 1. The digital block has three inputs (i.e., data, clock signal, and voltage source) and one output (backscattering data). The digital block is the baseband processor itself; it comprises five blocks: receiver (PIE decoder and data buffer), command processor, transmitter (Miller encoder, FM0 encoder, and frame generator), CRC (CRC-5 and CRC-16), BLF generator, and PRNG block. Each block will be described in detail as follows.

A. Receiver Block

The receiver block comprises three main components, i.e., Pulse Interval Encoding (PIE) decoder, data buffer or command parser, and BLF generator. In Fig. 2, the PIE decoder and data buffer are combined and represented as a "logic block." When a data signal comes from an RF demodulator, the data will be processed and interpreted by the PIE decoder by using Counter and Logic blocks. Those interpreted data are then stored in the data buffer and transmitted to command processing. The command parser will then separate the reader commands stored in Memory. Besides storing the data, the data buffer will also separate the query command to obtain Miller index (M) and Divide Ratio (DR) parameters. These two parameters are essential for the RFID system since M is used to determine which encoder to be used; either FM0 or DR is used to determine the value of BLF [25].

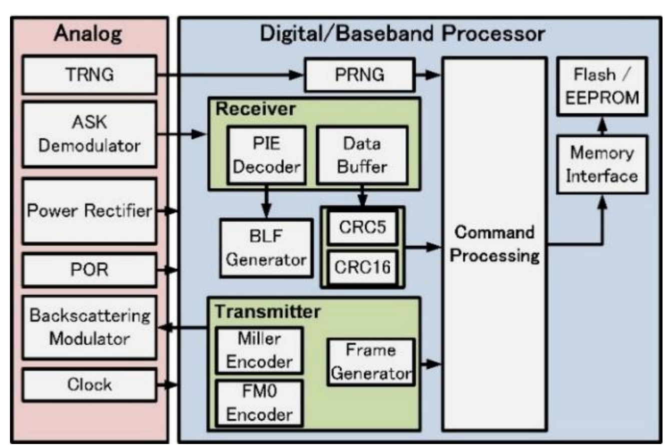


Fig. 1. Block Diagram of RFID Tag System.

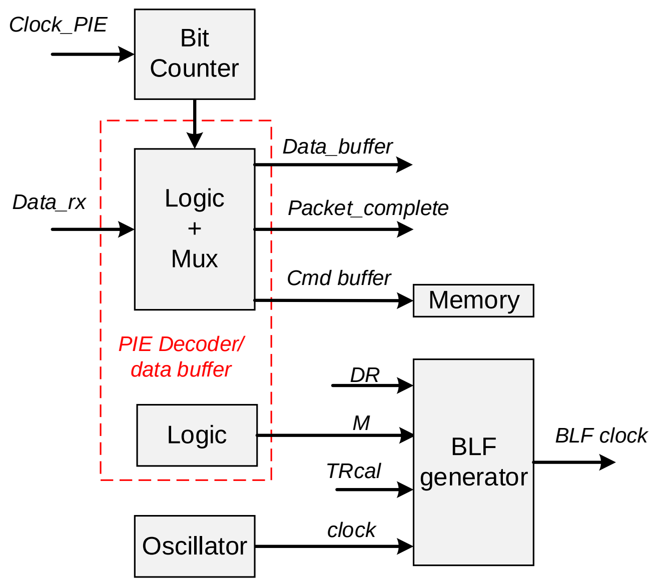


Fig. 2. Schematic of the Receiver Block.

Before designing the PIE decoder, the FSM should be determined first. PIE decoder consists of two state machines: PIE counter and data packet structure states. The PIE counter state machine detects the symbol of the bit logic state, either high or low, and then gives a corresponding instruction to the counter. As shown in Fig. 3(a), after the transition from logic low to logic high, the counter counts a full symbol until the end. Afterward, it begins to reset the counter when a new symbol comes. Meanwhile, the data packet structure state machine extracts the query parameters. The data packet structure sent by the reader includes delimiter, data-0, Reader-to-tag calibration (RTcal), Tag-to-reader calibration (TRcal), and data. In detail, the delimiter is a timing parameter with a constant value of 12.5 µs, ensuring that the tag can receive data from the reader even at the longest possible data-0. Data-0 is data that contains all binary 0. RTcal is a sum of the duration of symbols one and zero, used to determine the bit rate of the reader's transmitter. TRcal is a timing parameter used along with the DR parameter to determine the value of BLF. However, not every command has a TRcal. The TRcal will exit only if the symbol value is less than RTcal. This data packet structure state machine is illustrated in Fig. 3(b).

To decode the PIE symbol correctly, a 16-bit counter is used; this counter calculates the length of the PIE symbol. Afterward, the result is then compared to the calibration value in the preamble. According to the EPC Gen-2 standard, the counter requires at least 324 kHz of sampling clock since the minimum interval length between "0," and "1" symbols is 3.1 µs. In our design, the counter accommodates a clock input up to 16 MHz without causing the overflow to the longest possible TRcal symbol. The schematic structure of the PIE decoder is shown in Fig. 4. Data buffer and BLF generator are represented as "logic blocks." The logic and mux block are responsible for separating a command from its accompanying parameters, whereas another logic block is to extract DR and M values. These DR, M, and TRcal values are then used to determine the BLF value (Eq. 1),

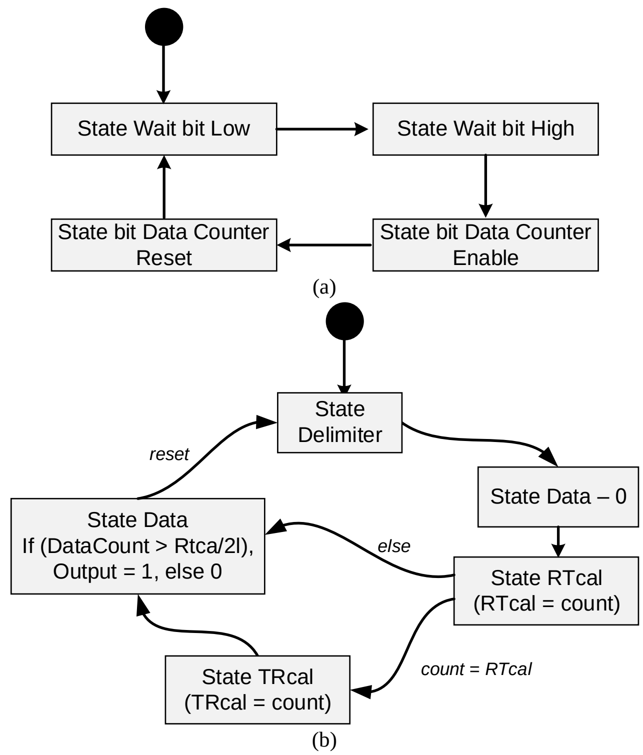


Fig. 3. (a) State Diagram of PIE Counter; (b) State Diagram of PIE Data Packet Structure.

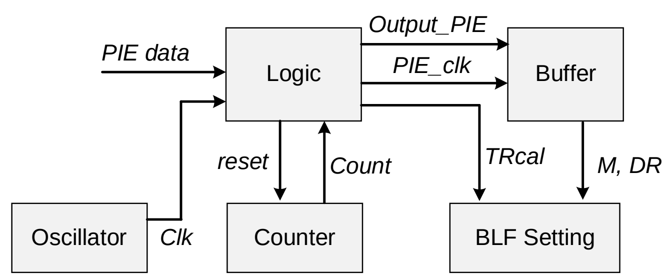


Fig. 4. Schematic Design of PIE Decoder.

$$BLF = \frac{DR}{TRcal} \quad (1)$$

The DR parameter has a value of either 8 or 64/3, depending on the parameter set in the query command. TRcal is calculated by the tag using oscillator frequency clock (f0) as a counter, and thus written as in Eq. 2, where TRcount defines the TRcal counts obtained from the receiver counter,

$$TRcount = TRcal \times f0 \quad (2)$$

B. Transmitter Block

The transmitter block (Fig. 5) comprises five inputs from the command processing block: (1) data input (PC_EPC); (2) M carrier signal, which is used to determine the encoding type; (3) TRext for deciding the use of pilot tone in the preamble; (4) CRC; and (5) Tx_enable signal. These inputs are processed in the frame generator, which is responsible for choosing the encoder type and inserting preamble and termination signals. In this study, the design of the transmitter block is focused on FM0 and Miller Encoder only.

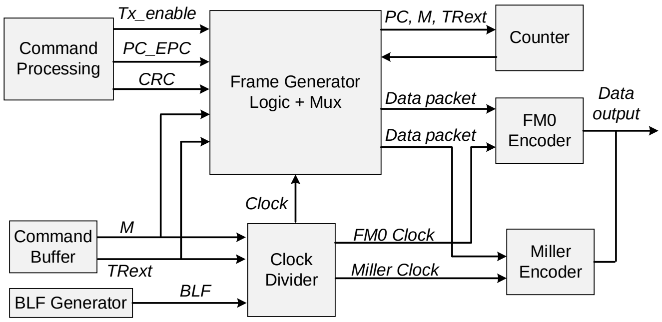
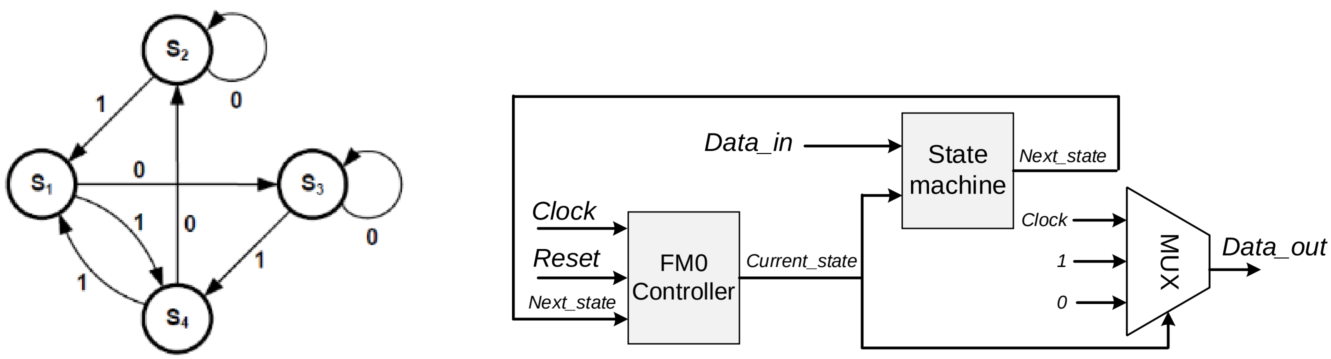


Fig. 5. Schematic of the Transmitter Block.

The default operation mode for data transmission from the tag uses FM0 Encoding. In transmitting the data, the tag changes the backscatter status in each symbol edge. Based on EPC Gen-2 standard, data transmission must be terminated with a dummy binary symbol "1" followed by a "low" state. The change of FM0 binary value determines the next sequential binary value, and therefore there are four states in total, as depicted in Fig. 6(a). The schematic design of this FM0 Encoding using the state machine model in Verilog is shown in Fig. 6(b). A Miller Modulated Signal (MMS) encoding is also added to the transmitted signal to avoid interference noise and be more flexible in managing the data rate. Any FM0 encoded signal will be encoded with an MMS subcarrier signal with a particular M value. The M value represents how many clock cycles MMS encodes the FM0 signal. The higher the M value, the slower the data rate. In the Miller encoder, the binary symbol "1" (a state transition) is inserted in the middle of the symbol. In contrast, the binary symbol "0" (no state transition) is in the middle of the symbol.

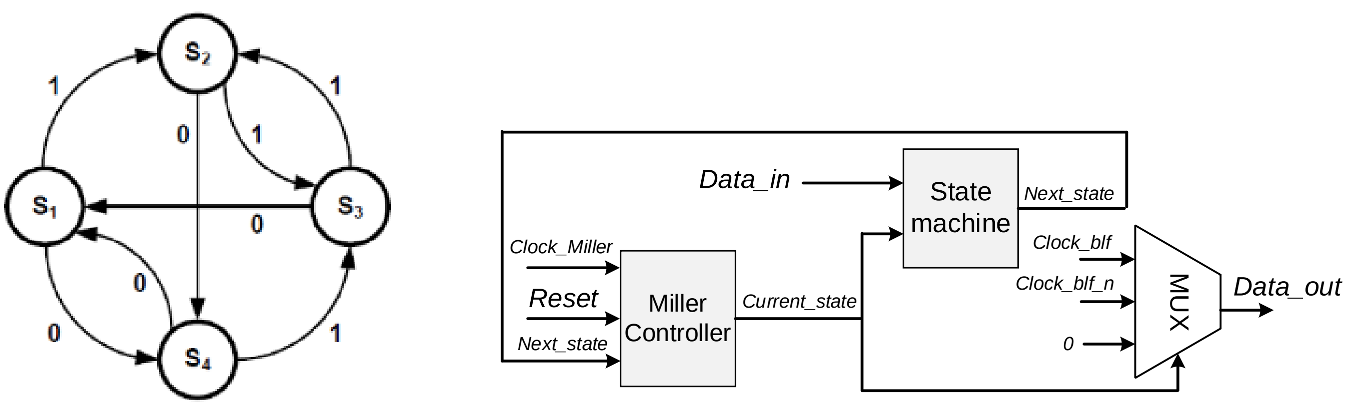
Moreover, there is no state transition in symbol transition, except when the binary symbols "0" meet consecutively. Furthermore, the encoding results are multiplied by the *M* value, where *M* is either 2, 4, or 8. The state diagram for this Miller encoding is represented in Fig. 7(a), while the Verilog schematic design is shown in Fig. 7(b).

The command processing is a block responsible for controlling the transition between EPC Gen-2 states and determining the tag response according to the reader commands. The command processor has two crucial logic functions to manage: slot logic as the anti-collision protocol and session logic. The schematic of the command processor block is depicted in Fig. 8(a). The state diagram for the command processing block is depicted in Fig. 8(b). The RFID transponder tag consists of seven states: Ready, Arbitrate, Reply, Acknowledged, Open, Secured, and Killed. When the tag is powered, it stands by in the Ready state until the Query command is received. The tag will then enter one of the two following states: 1) The tag changes to the slot state if a suitable session exists, or 2) The tag changes to the Arbitrate state if the slot value does not equal zero.



(a) (b)

Fig. 6. (a) State diagram of FM0 encoding [26]; (b) Schematic design of FM0 encoder



(a) (b)

Fig. 7. (a) State Diagram of Miller Encoding [26]; (b) Schematic Design of Miller Encoder.

The anti-collision protocol is applied within the Arbitrate state, and the tag's state will be held until the slot value equals zero. At this hold state, there are three command query options for the reader: (1) reduce slot number using QueryRep command; (2) change slot masking using QueryAdj, or (3) repeat the Query command. Once the slot has zero value, the tag will enter the Reply state and send a 16-bit random number (RN16). Afterward, the tag will wait for a response from the reader and enter one of the following states: 1) return to the Ready state if the received command query is an invalid session, or 2) return to the Arbitrate state if the received command query matches the valid session. If the reader correctly receives RN16, the reader will send an acknowledgment signal along with the previously received RN16. If the number matches the sent number, the tag will send the EPC value to the reader and change the session state from A to B or vice versa. Otherwise, the tag will return to the Arbitrate state. Additional blocks, i.e., CRC and PRNG, are required to make the system works correctly and adequately. For the CRC block, we employed two CRCs as suggested by the EPC Gen-2 standard: CRC-16 and CRC-5.

CRC-5 is used when inventory mode is started or when the reader gives the Query command. CRC-16 is used when the reader sends command select, Req-RN, NAK, Read, Kill, and Lock. The CRC-5 has a specification in the form of polynomial $x^5 + x^3 + 1$ [27], residue 000002, Preset 010012, and length 5 bits. The CRC-5 is implemented using a register flip-flop and XOR circuit in our design.

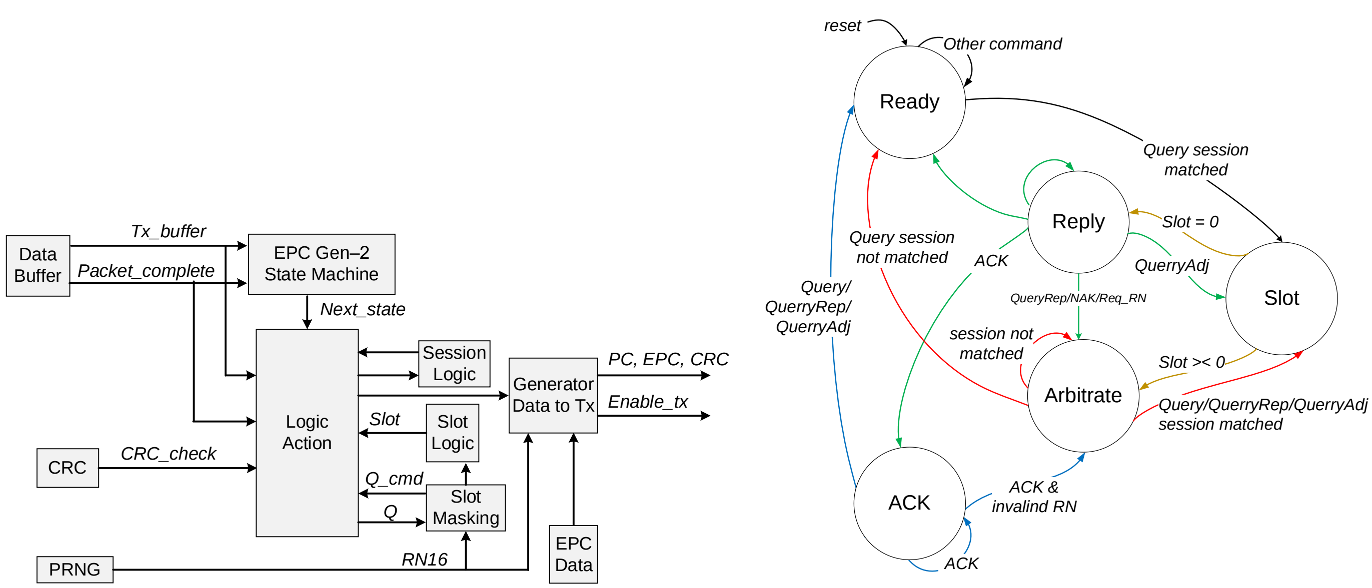
In this design, the value of preset 010012 is inserted into the register flip flop. Input data will then enter one by one according to the clock signal. CRC-5 check will return pass/successful if the value in the register output is 000002. On the other hand, CRC-16 has a specification in polynomial $x^{16} + x^{12} + x^5 + 1$ [28], residue 1D0Fh, Preset FFFFh, and length 16 bits. The same with the CRC-5 implementation, the CRC-16 uses a register flip flop and XOR circuit. The difference is that CRC-16 has more flip-flops compared to CRC-5. In this design, in the beginning, the FFFFh preset value is inserted into the register flip flop. Input data will then enter one by one according to the clock signal. The CRC-16 check will return pass/successful if the value in the register output is 1D0Fh.

C. *Pseudo Random Generator Number (PRNG)*

The PRNG block commonly generates a 16-bit random value. This random value will determine the slot value in the tag according to the Q value received by the tag. The most recent advancement of a hardware implementation for a security system employs the physical variation in an integrated circuit; it can be called Physically Unclonable Functions (PUF), as presented in [29], [30]. However, the system becomes complicated, so it is not preferred for RFID applications. Instead of PUF, this study employs the Linear Feedback Shift Register (LFSR) for the PRNG block to obtain a uniform output distribution. LFSR uses registers or flip-flops [31] whose input is based on linear functions from the previous state, which is implemented using the XOR function. The LFSR uses polynomial 1 + x11 + x13 + x14 + x16. This polynomial selection is based on the study by [32].

D. *Baseband Processor*

The overview of the baseband processor is depicted in Fig. 9. It works through the following steps: (1) the receiver sends input to command processing in the form of decoded data buffer along with a package complete flag as an indicator that input data has been received; (2) at the same time, the receiver also sends BLF and query parameters to the transmitter; (3) command processing then sends the reply data to the transmitter, transmitting the data back to the reader.



(a) (b)

Fig. 8. (a) Schematic of Command Processing; (b) State Diagram of Command Processing Block.

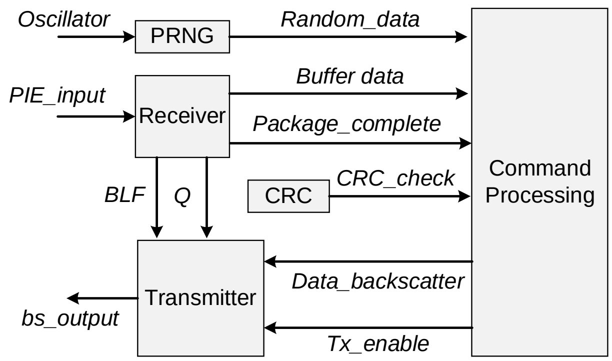


Fig. 9. Schematic of UHF RFID Tag baseband Processor.

III. Result and Discussion

In this section, we evaluate our design by testing each block that contains the receiver, transmitter, and command processing block. Also, we verify the entire system by implementing the UVM. In this test, we use a reader that fulfills the ISO8000-6C standard and evaluates our proposed design by confirming the RFID reader and tag data. The test was conducted on an RTL simulation.

A. Receiver Block

We first tested the PIE decoder block on the receiver side. The central controller is the PIE state machine. When the enable counter is active at the delimiter state, the state machine will check whether the delimiter value is 12.5 µs or not. Then, there are two possible states: data-0 and RTcal. Both are used as a comparison to determine the logic value “1” or “0”. Next is TRcal state, which is used to obtain the TRcal value. Finally, this TRcal will be used as a parameter to produce the desired BLF value. After the PIE block has been tested, the whole receiver block (including the PIE encoding, command buffer, and BLF generator) is evaluated. The value of PIE_output, originating from the PIE decoder, is stored in the data buffer module. It is then transmitted to command processing and the packet_complete signal, indicating the packet’s end. The data buffer module will also perform a command filter/parser. The command filer/parser will extract the query parameter used by the BLF generator module. The BLF module will only work to reduce dynamic power consumption if it receives a signal from a transmitter. The optimum throughput speed can be obtained using this receiver architecture, which only produces the latency of two-oscillator clock cycles. In our design, an internal 8 MHz oscillator is used, which causes a latency of 250 ns. However, this latency is insignificant compared to the whole system, and it can be neglected.

B. Transmitter Block

The FM0 and Miller encoder blocks are tested well on an RTL simulation. When the value of enable_FM0 is “1”, the FM0 encoder module becomes active. The input data from the frame generator will enter the state machine, and the next state will be determined. The FM0 controller module will provide input to the multiplexer in the form of data_selector to select the output of fm0_data. This FM0 encoder architecture design has a latency of one cycle fm0_clock that depends on the BLF value. Hence, the latency in the FM0 module is defined as Eq. 3,

$$FM0\_latency = \frac{TRcal}{DR} \tag{3}$$

In testing the Miller encoder block, the subcarrier encoding with M = 2 is conducted. The test result shows that the working principle of the circuit is similar to the FM0 encoder module, where the data_selector output from the controller determines the Miller encoder data output. The main difference is in the clock that it uses. In the FM0 encoder module, the FM0 clock has the same value as the BLF clock, whereas, in the Miller encoder, the Miller clock value depends on the M value. For instance, the Miller encoder latency depends on the BLF and M values, as defined in Eq. 4.

$$Miller\_latency = \frac{TRcal}{DR \times M} \tag{4}$$

We also tested the whole transmitter block. Input data from command processing in PC/EPC/RN16 data and tx_enable signals are processed by the frame generator and sent serially to the encoder module. Depending on the PC’s value, counters, preamble, data, and CRC will activate. In this test, the FM0 and TRext are set to “0”. Thus, the value of fm0_enable becomes “1”, activating the FM0 encoder module. Furthermore, clock for BLF and M value determine the clock value of each encoder module. Simulation reveals the latency of the transmitter block depends on the encoder type. The FM0 encoder has low latency. Nevertheless, when miller encoding with M = 8 is used, the latency becomes significant, as inferred in Eq. 5, where Tx_latency denotes the latency of the transmitter block and clk_osc defines the oscillator clock.

$$Tx\_latency = \left(2 \times clk\_osc + \frac{TRcal}{DR}\right) + \left(\frac{TRcal}{DR \times M}\right) \tag{5}$$

C. Command Processing Block

The complete testing sequence of the command processing module has been verified carefully. It is tested from the query command is received until the tag sends the EPC code. When the receiver module receives a packet, it will send a signal to indicate that the packet is complete, and thus the command processing will make a response accordingly. Finally, the CRC module checks whether the command query has a valid CRC value or not. The latency of this command processing is three oscillator clock cycles. 16 MHz of the internal oscillator is used in this design, which produces a latency of 375 ns; this latency is not significant.

D. Full System Test

The entire system is simulated and tested using UVM, specifically the Coverage Driven Verification (CDV) approach. The UVM is version 1.1d and implemented using Synopsys VCS 2014.10 on the EDA playground application. The testing diagram using UVM is illustrated in Fig 10(a). We used a reader that already fulfills the ISO 18000-6C standard requirement for the whole system test as a master. The reader has simple instructions and is intended only to identify the EPC value of the tag.

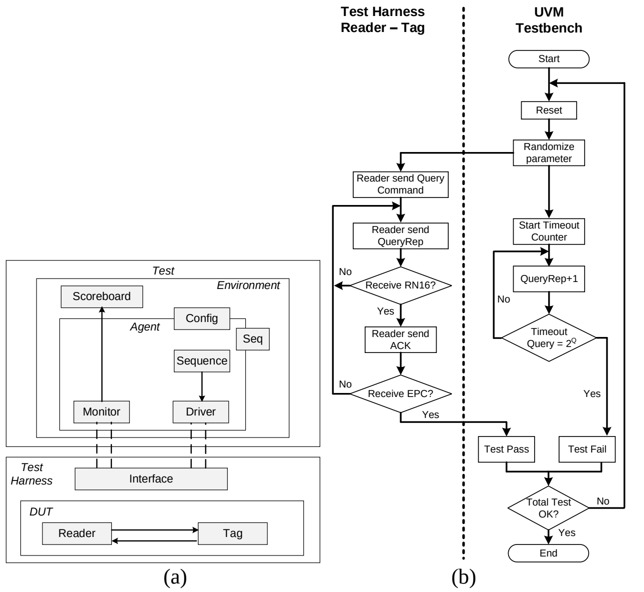


Fig. 10. (a) Full System Testing Diagram using UVM; (b) Flowchart of the System Testing using UVM.

In general, the test is conducted based on the following sequences: (1) after a reset signal is given to the reader, it sends the Query command to the tag; (2) the reader will continuously send QueryRep in order to have the tag reduces the slot value until zero and until the tag sends 16-bit random number (RN16) to the reader; (3) using “ACK” command, RN16 will be transmitted back to the tag, after which the reader will wait for EPC value from the tag. The test session is considered to be finished once the reader receives the EPC value from the tag. The flowchart of the testing sequences is illustrated in Fig. 10(b). The functional system test is intended to check both timing and query parameters to be used by the reader and the command response of the tag. Four available techniques are chosen depending on the dataset size, testing type, and specification. The techniques can obtain those parameters, i.e., constrained random test, direct test, coverage test, and edge corner test. A timeout program is also set up in the test to anticipate a situation when the tag does not send any response that makes the reader unable to obtain the EPC value. The timeout value depends on the Q value, where Q denotes the query parameter used to determine the tag’s random slot, and 2Q can find the total timeout of UVM (UVM_timeout).

1) Query parameter test: A query command is an inventory command that the reader first sends. The tag should be able to accept and process all possible combinations of query parameters (including Q, TRext, DR, and encoding). The query parameter test is conducted using constrained random and coverage tests. The coverage tests all possible combinations to generate the parameters, which amounts to 512 combinations. While a random test is conducted to identify possible errors caused by a state transition. All these possible combinations are summarized in Table I. Based on the tests results shown in Table II, the design100% passes all the test combinations.

2) Timing parameter test: The timing parameter is conducted using the edge corner and constrained random tests. The edge corner test is conducted only on the smallest and biggest possible value of Tari, PW, RTcal, and TRcal parameters, resulting in 16 combinations. Meanwhile, the constrained random test is conducted by setting the parameters within certain limits, i.e., 6.25 µs to 25 µs for Tari, then 2.5Tari to 3Tari for RTcal, and 1.1RTcal to 3RTcal for TRcal. The test results are shown in Table II. It shows that the design passes all the testing combinations of the timing parameter test.

3) State transition test: An RFID tag design must be able to respond according to specifications based on the command it receives from a reader. The RFID tag has four states in our design: ready, reply, arbitrate, and acknowledge. On the other hand, there are six commands from the reader that the tag must recognize: query, queryrep, queryadj, select, ACK, and NAK. The command response test is conducted using direct testing. The result is summarized in Table III.

4) BLF Test: The BLF test is conducted using 16 MHz of frequency sampling. Based on the specification regulated in ISO 18000-6C, the deviations in each parameter must be kept below 2.5%. Based on the test results shown in Table II and Table IV, the deviations of the parameters are all below 2.5%; thus, all tests are 100% passed. Further test will follow a procedure from [33].

TABLE I. QUERY COMMAND COMBINATIONS

	Command	***DR***	***M***	***TRext***	**Sel**	**Session**	**Target**	***Q***	**CRC-5**
#of bits	4	-	2	1	2	-	1	4	5
Description	1000	0 :DR=8 1:DR=64/3	00:M=1 01:M=2 10:M=4 11:M=8	00: No. Pilot 01:Use Pilot	00:All 01:All 10:~SL 11:SL	00:M=1 01:M=2 10:M=4 11:M=8	0:A 1:B	0-15	-

TABLE II. SUMMARY OF SYSTEM TESTING RESULTS USING UVM

Testing Parameter	**Testing technique**	**#of Combination**	**Result**
Query Parameter	Coverage Test	512	100% pass
	Random Test	500	100% pass
Timing Parameter	Edge Corner Test	16	100% pass
	Random Test	300	100% pass
BLF	Direct Test	16	100% pass

TABLE III. TEST RESULT OF COMMAND RESPONSE TEST

Command	**Condition**	**Transition State**				**Result**
		Ready	**Arbitrate**	**Reply**	**Acknowledge**	
Query	Slot=0	Reply	Reply	Reply	Reply	Pass
	Slot><0	Arbitrate	Arbitrate	Arbitrate	Arbitrate	Pass
QueryRep	Slot=0	Ready	Reply	Arbitrate	Ready	Pass
	Slot><0	Ready	Arbitrate	Arbitrate	Ready	Pass
QueryAdj	Slot=0	Ready	Reply	Reply	Ready	Pass
	Slot><0	Ready	Arbitrate	Arbitrate	Ready	Pass
ACK	-	Ready	Arbitrate	Acknowledge	Acknowledge	Pass
select	-	Ready	Arbitrate	Arbitrate	Ready	Pass
NAK	-	Ready	Arbitrate	Arbitrate	Arbitrate	Pass

TABLE IV. BLF PARAMETER TEST RESULTS

DR	***TRcal* (µs)**	**Expected BLF (kHz)**	**BLF Result (kHz)**	**Variance %**
64/3	33.3	640	640	0
	44.4	480	484.84	1.008
	66.7	320	320	0
	74.1	288	290.9	1.007
	83.3	256	258.06	0.805
	102.6	208	210.52	1.212
	159.8	133.5	134.45	0.712
	211.2	101	101.26	0.257
8	17.2	465	457.14	-1.690
	25	320	320	0
	27.8	288	285.71	-0.795
	31.25	256	253.96	-0.797
	38.46	208	207.79	-0.1009
	50	160	160	0
	59.9	133.5	133	-0.374
	200	40	40	0

A. *Implementation*

The baseband processor is implemented on the Altera DE2-115 FPGA development board. The board uses Cyclone EP4CE115 chip, which has 114,480 logic elements, 3,9 Mbits RAM, 266 multipliers [34], and a Nios II soft-processor. Moreover, the board is also equipped with several interfaces such as Ethernet, RS232, PS2, and USB. The implementation diagram of the baseband processor and the reader in the DE2-115 board is depicted in Fig. 11. The implemented system comprises two blocks: the processing system and programmable logic. The processing system comprises a Nios II soft-processor and memory-on-chip, while the programmable logic comprises the RFID tag baseband processor and the reader. The programmable logic is connected to the Avalon interface through the PIO. The Nios II soft-processor sends input in the form of EPC data to the tag and timing and query parameters to the reader. It will monitor the outputs in the form of protocol status, EPC data, and *TRcal*. The result is displayed on a host computer connected to Nios II through the UART JTAG interface. The RTL design synthesis is performed using Altera Quartus® software. Firstly, the implemented design is tested for its system communication. Then, the test process and results are monitored and displayed on the Nios II console. The DE2 communicates with the host PC via the JTAG interface.

As the test program is run, Nios II will initialize the parameters by giving the corresponding input to the reader. Afterward, Nios II will monitor any changes in the protocol status since it indicates the exchange data process or status between the reader and tag. Every communication from the reader is displayed on the Nios II console. There are three stages of communication between reader and tag. In the first stage, after Nios II initialization, the reader will send data in the form of the Query command, then followed by the QueryRep command. In the second stage, the reader receives a reply from the tag in the form of RN16. Then, the reader will return the RN16 along with the ACK command. Finally, the reader receives the EPC value from the tag in the last stage.

These three stages of communication and their testing result are shown in Fig. 12. It can be seen that the implemented design is successful in performing all stages of communication, and the reader can also successfully receive the EPC value from the tag. The synthesis result is presented in Table V and Table VI. The design uses 976 logic elements and 939 combination functions, consisting of 573 (4-input functions), 213 (3-input functions), and 153 (2-or-less-input functions). From 976 logic inputs available, 119 are used in arithmetic mode, and the other 373 are registers. As for the submodules, the command processor uses 346 logic elements, while the transmitter and receiver use 328 and 265 logic elements, respectively.

The power consumption is calculated using Powerplay analysis provided by Quartus, as shown in Fig. 13. A Powerplay is a standard tool used by various scholars to analyze the power consumption of the designed chip [35]–[37]. Based on the analysis result, the proposed chip consumes 173.14 mW with the following details: 0.13 mW from dynamic power dissipation, 98.6 mW from static power dissipation, and 74.34 mW from I/O dissipation. The power consumption is relatively low and it can also answer one of various challenges in RFID tag chip design, which is low-power allowing with low-cost [38]. This data is expected to transfer the design into a specific chip dedicated to IoT application, specially targeted for a low-power application case that requires RFID to operate.

The RFID tag design made is a digital block from a complete transponder tag. The phase for developing RFID tags is to integrate digital blocks with analog parts such as antennas, modulators, and voltage regulators. The design of the RFID tag made is an initial study that focuses on a digital block architecture that can accommodate the ISO 18000-6C standard. The development space for digital blocks is still vast, including design optimization so that a tag design can be obtained that has a smaller size and more efficient in consuming the power.

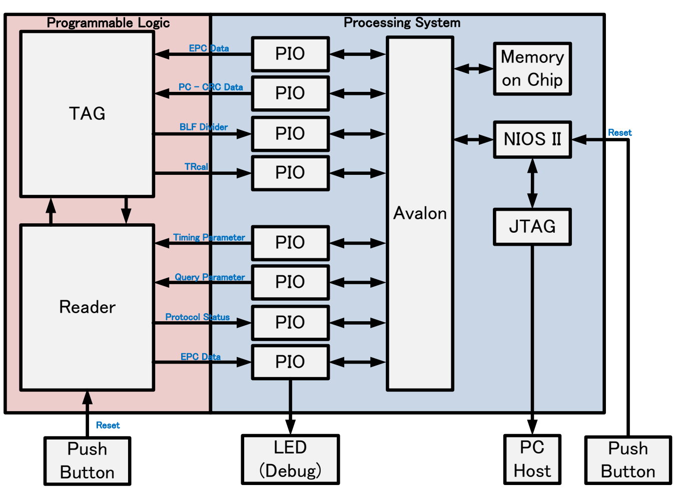


Fig. 11. Block Diagram of the Full RFID System as Implemented in FPGA Altera DE2-115 Board.

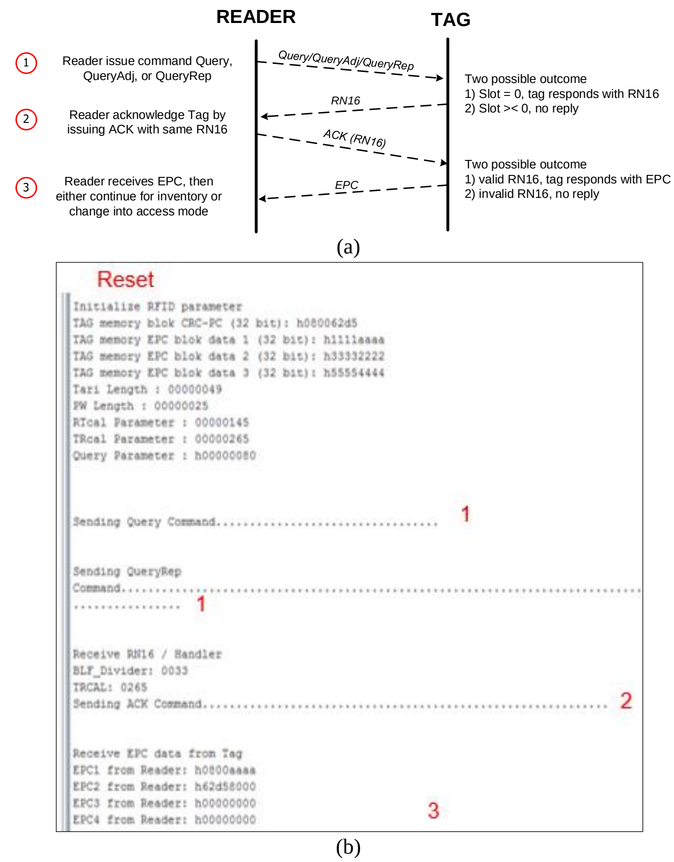


Fig. 12. (a) Communication Stages between Reader and Tag; (b) Its Corresponding Implemented System Result.

TABLE V. SYNTHESIS RESULT OF THE FULL RFID TAG BASEBAND PROCESSOR (MAIN MODULES)

Resource	Usage
Estimated total logic elements	976
Total combinational functions	939
Logic element usage by number of LUT inputs	
• 4 input functions	573
• 3 input functions	213
• ≤ 2 input functions	153
Logic elements by mode	
• Normal mode	820
• Arithmetic mode	119
Total registers	373
• Dedicated logic registers	373
• I/O registers	0
I/O pins	168
Embedded multiplier 9-bit elements	0
Maximum fan-out node	Clk-input
Maximum fan-out	268
Total fan-out	4641
Average fan-out	2.82

TABLE VI. SYNTHESIS RESULT OF THE FULL RFID TAG BASEBAND PROCESSOR (SUBMODULES)

Entity name	LC Combinations	LC Registers
Debug	939 (0)	373 (0)
Tag_transmitter	328 (138)	196 (145)
Miller_enc	25 (25)	13 (13)
Frame_generator	140 (140)	24 (24)
Fm0_enc	18 (18)	11 (11)
clock	7 (7)	3 (3)
Tag_receiver	265 (2)	126 (1)
Pnrg_module	11 (11)	16 (16)
Pie_demod	70 (55)	46 (35)
Counter_rx	15 (15)	11 (11)
Cmd_buffer	114 (114)	54 (54)
Blf_generator	68 (68)	9 (9)
Command processor	346 (346)	51 (51)

PowerPlay Power Analyzer Summary	
PowerPlay Power Analyzer Status	Successful - Fri Oct 27 15:02:47 2017
Quartus Prime Version	16.0.2 Build 222 07/20/2016 SJ Lite Edition
Revision Name	RFIDTag
Top-level Entity Name	tag_RFID
Family	Cyclone IV E
Device	EP4CE115F29C7
Power Models	Final
Total Thermal Power Dissipation	173.14 mW
Core Dynamic Thermal Power Dissipation	0.13 mW
Core Static Thermal Power Dissipation	98.67 mW
I/O Thermal Power Dissipation	74.34 mW

Fig. 13. Power Consumption Analysis Result from Power Play.

IV. CONCLUSION

This study has designed a baseband processor design for UHF RFID passive tag. The baseband processor consists of several blocks, *i.e.*, receiver, transmitter, command processing, CRC (CRC-5 and CRC-16), and PRNG. A detailed RTL design and testing methodology of the baseband processor was also presented. All the design and testing criteria were constructed compliant with the EPC Gen-2 standard. Each block showed a low latency (below 400 ns), which is negligible to the whole system. Using the UVM testing method, 1344 parameter combinations were conducted on the whole system design. The results reveal that it passed all the testing scenarios by the UVM, including query parameter, timing parameter, state transition, and BLF test. Moreover, the design was also successfully implemented on an FPGA board (Altera DE2-115). The implemented system used 976 logic elements and consumed 173.14 mW of power dissipation. This comprehensive RFID design and testing methodology allow a more reliable RFID baseband processor design to be realized, leading to a higher yield of other real chip fabrication processes. The results show that it can be used to enable a chip-based, low-power IoT Application. Future work will focus on the chipset design and tap-out the design.

ACKNOWLEDGMENT

The publication fee for this work is fully handled and sponsored by program peningkatan *Global Competitiveness* Perguruan Tinggi Indonesia Universitas Pendidikan Indonesia 2021 Batch II with No SK 1370/UN40/PT.01.02/2021.

REFERENCES

[1] Condea, F. Thiesse, and E. Fleisch, "RFID-enabled shelf replenishment with backroom monitoring in retail stores," Decision Support Systems, vol. 52, no. 4, pp. 839–849, Mar. 2012, doi: 10.1016/j.dss.2011.11.018.

[2] M. Liukkonen and T.-N. Tsai, "Toward decentralized intelligence in manufacturing: recent trends in automatic identification of things," Int J Adv Manuf Technol, vol. 87, no. 9, pp. 2509–2531, Dec. 2016, doi: 10.1007/s00170-016-8628-y.

[3] M. Kirch, O. Poenicke, and K. Richter, "RFID in Logistics and Production –Applications, Research and Visions for Smart Logistics Zones," Procedia Engineering, vol. 178, pp. 526–533, Jan. 2017, doi: 10.1016/j.proeng.2017.01.101.

[4] N. M. Sahar, M. T. Islam, and N. Misran, "Design of dualband antenna for RFID applications," International Journal of Electrical and Computer Engineering (IJECE), vol. 9, no. 4, Aug. 2019, doi: 10.11591/ijece.v9i4.pp3146-3152.

[5] E. E. Abel, A. L. M. Shafie, and W. H. Chan, "Deployment of internet of things-based cloudlet-cloud for surveillance operations," IAES International Journal of Artificial Intelligence (IJ-AI), vol. 10, no. 1, Mar. 2021, doi: 10.11591/ijai.v10.i1.pp24-34.

[6] R. Abdulla, A. Abdillahi, and M. K. Abbas, "Electronic Toll Collection System based on Radio Frequency Identification System," International Journal of Electrical and Computer Engineering (IJECE), vol. 8, no. 3, Jun. 2018, doi: 10.11591/ijece.v8i3.pp1602-1610.

[7] A. Abdulkareem, "Development and implementation of a miniature RFID system in a shopping mall environment," International Journal of Electrical and Computer Engineering (IJECE), vol. 9, no. 2, Apr. 2019, doi: 10.11591/ijece.v9i2.pp1374-1378.

[8] A. J. Samuel and S. Sebastian, "An algorithm for IoT based vehicle verification system using RFID," International Journal of Electrical and Computer Engineering (IJECE), vol. 9, no. 5, Oct. 2019, doi: 10.11591/ijece.v9i5.pp3751-3758.

[9] H. A. Khan, R. Abdulla, S. K. Selvaperumal, and A. Bathich, "IoT based on secure personal healthcare using RFID technology and steganography," International Journal of Electrical and Computer Engineering (IJECE), vol. 11, no. 4, Aug. 2021, doi: 10.11591/ijece.v11i4.pp3300-3309.

[10] M. O. Adebiyi, R. O. Ogundokun, A. I. Nathus, and E. A. Adeniyi, "Smart transit payment for university campus transportation using RFID card system," International Journal of Electrical and Computer Engineering (IJECE), vol. 11, no. 5, Oct. 2021, doi: 10.11591/ijece.v11i5.pp4353-4360.

[11] D. He, G. Mujica, G. Liang, J. Portilla, and T. Riesgo, "Radio propagation modeling and real test of ZigBee based indoor wireless sensor networks," Journal of Systems Architecture, vol. 60, no. 9, pp. 711–725, Oct. 2014, doi: 10.1016/j.sysarc.2014.08.002.

[12] D. Lautner, X. Hua, S. DeBates, and S. Ren, "WaaS (Wireless-as-a-Sensor): Conception, design and implementation on mobile platforms," Journal of Systems Architecture, vol. 88, pp. 65–73, Aug. 2018, doi: 10.1016/j.sysarc.2018.05.009.

[13] X. Jia, Q. Feng, T. Fan, and Q. Lei, "RFID technology and its applications in Internet of Things (IoT)," in 2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet), Yichang, China, Apr. 2012, pp. 1282–1285. doi: 10.1109/CECNet.2012.6201508.

[14] S.-Y. Wong and C. Chen, "Power efficient multi-stage CMOS rectifier design for UHF RFID tags," Integration, vol. 44, no. 3, pp. 242–255, Jun. 2011, doi: 10.1016/j.vlsi.2011.03.005.

[15] M. Škiljo, P. Šolić, Z. Blažević, L. D. Rodić, and T. Perković, "UHF RFID: Retail Store Performance," IEEE Journal of Radio Frequency Identification, pp. 1–1, 2021, doi: 10.1109/JRFID.2021.3129694.

[16] K. Bhanushali, W. Zhao, W. S. Pitts, and P. D. Franzon, "A 125 μm × 245 μm Mainly Digital UHF EPC Gen2 Compatible RFID Tag in 55 nm CMOS Process," IEEE Journal of Radio Frequency Identification, vol. 5, no. 3, pp. 317–323, Sep. 2021, doi: 10.1109/JRFID.2021.3087448.

[17] I. Galko, R. Kuffa, P. Magdolenová, J. Svetlík, and A. Veľas, "RFID tags at the operation of fire stations," Transportation Research Procedia, vol. 55, pp. 941–948, Jan. 2021, doi: 10.1016/j.trpro.2021.07.062.

[18] E. H. Hadj-Mihoub-Sidi-Moussa, R. Touhami, and S. Tedjini, "Design and Evaluation of an RFID Localization System based on Read Count," IETE Journal of Research, vol. 0, no. 0, pp. 1–10, Feb. 2021, doi: 10.1080/03772063.2021.1880341.

[19] D. Wei, C. Zhang, Y. Cui, H. Chen, and Z. Wang, "Design of a low-cost low-power baseband-processor for UHF RFID tag with asynchronous design technique," in 2012 IEEE International Symposium on Circuits and Systems (ISCAS), Seoul, Korea (South), May 2012, pp. 2789–2792. doi: 10.1109/ISCAS.2012.6271889.

[20] J.-W. Lee, N. D. Phan, D. H.-T. Vo, and V.-H. Duong, "A Fully Integrated EPC Gen-2 UHF-Band Passive Tag IC Using an Efficient Power Management Technique," IEEE Transactions on Industrial Electronics, vol. 61, no. 6, pp. 2922–2932, Jun. 2014, doi: 10.1109/TIE.2013.2278519.

[21] Smarani Ismail and A. Ibrahim, "Modelling and Simulation of Baseband Processor for UHF RFID Reader on FPGA," International Journal of Electrical and Electronic Systems Research, vol. 6, pp. 53–67, 2013.

[22] J.-T. Su, Z. Xie, X.-A. Wang, and Y. Cao, "Design and Automatic System Verification of Digital Baseband for UHF RFID Tag," International Journal of Electronics and Electrical Engineering, vol. 1, no. 3, pp. 130–134, 2013.

[23] S. Li et al., "A −20 dBm Passive UHF RFID Tag IC With MTP NVM in 0.13-μm Standard CMOS Process," IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 67, no. 12, pp. 4566–4579, Dec. 2020, doi: 10.1109/TCSI.2020.3007952.

[24] M. Yang, S. Wei, M. Liu, R. Huang, M. Yu, and F. Chen, "Design Verification and Test Techniques for UHF RFID Tag IC," in 2019 IEEE Sustainable Power and Energy Conference (iSPEC), Beijing, China, Nov. 2019, pp. 2500–2505. doi: 10.1109/iSPEC48194.2019.8975371.

[25] L. Xie, W. Nie, X. Yang, Y. Wang, and M. Zhou, "A BLF Generation Scheme with Clock Variance-Tolerance for Baseband Processor of EPC Gen2 UHF RFID Tag," in Cloud Computing and Security, Cham, 2018, pp. 543–552. doi: 10.1007/978-3-030-00015-8_47.

[26] K. Fyhn, R. M. Jacobsen, P. Popovski, A. Scaglione, and T. Larsen, "Multipacket Reception of Passive UHF RFID Tags: A Communication Theoretic Approach," IEEE Transactions on Signal Processing, vol. 59, no. 9, pp. 4225–4237, Sep. 2011, doi: 10.1109/TSP.2011.2159499.

[27] Y.-C. Hung and H.-J. Hung, "RFID design with CRC programmable capability, dual encode functions and dual modulation outputs," in 2010 International Symposium on Next Generation Electronics, Kaohsiung, Taiwan, Nov. 2010, pp. 219–222. doi: 10.1109/ISNE.2010.5669159.

[28] J. Zhang and H. Gao, "Implementation of CRC Algorithm in UHF RFID Test System Based on Labview," Jul. 2016, pp. 902–906. doi: 10.2991/iccia-17.2017.158.

[29] L. T. Clark, J. Adams, and K. E. Holbert, "Reliable techniques for integrated circuit identification and true random number generation using 1.5-transistor flash memory," Integration, vol. 65, pp. 263–272, Mar. 2019, doi: 10.1016/j.vlsi.2017.10.001.

[30] M. Soybali, B. Ors, and G. Saldamli, "Implementation of a PUF Circuit on a FPGA," in 2011 4th IFIP International Conference on New Technologies, Mobility and Security, Paris, France, Feb. 2011, pp. 1–5. doi: 10.1109/NTMS.2011.5720638.

[31] T. Khanom and F. Khanom, "Implementation of Pseudo-Random Number Generator Using LFSR," Publications and Research, Dec. 03, 2020. https://academicworks.cuny.edu/ny_pubs/642.

[32] Babitha K.P., Thushara T, and Dechakka M P, "FPGA based N-Bit LFSR to Generate Random Sequence Number," International Journal of Engineering Research and General Science, vol. 3, no. 3, pp. 6–10, 2015.

[33] D. G. Kuester, D. R. Novotny, J. R. Guerrieri, A. Ibrahim, and Z. B. Popovic, "Simple Test and Modeling of RFID Tag Backscatter," IEEE Transactions on Microwave Theory and Techniques, vol. 60, no. 7, pp. 2248–2258, Jul. 2012, doi: 10.1109/TMTT.2012.2195017.

[34] D. Tsiktsiris, D. Ziouzios, and M. Dasygenis, "A portable image processing accelerator using FPGA," in 2018 7th International Conference on Modern Circuits and Systems Technologies (MOCAST), Thessaloniki, Greece, May 2018, pp. 1–4. doi: 10.1109/MOCAST. 2018.8376566.

[35] A. Gidd, S. Ghasti, S. Jadhav, and K. Sivasankaran, "Performance Analysis of 32-Bit DADDA Multiplier Using 15–4 Compressor," in Microelectronic Devices, Circuits and Systems, Singapore, 2021, pp. 19–30. doi: 10.1007/978-981-16-5048-2_2.

[36] O. Drozd, G. Nowakowski, A. Sachenko, V. Antoniuk, V. Kochan, and M. Drozd, "Power-Oriented Monitoring of Clock Signals in FPGA Systems for Critical Application," Sensors, vol. 21, no. 3, Jan. 2021, doi: 10.3390/s21030792.

[37] M. Wagih and J. Shi, "Wireless Ice Detection and Monitoring Using Flexible UHF RFID Tags," IEEE Sensors Journal, vol. 21, no. 17, pp. 18715–18724, Sep. 2021, doi: 10.1109/JSEN.2021.3087326.

[38] J. Lu, D. Liu, H. Li, C. Zhang, and X. Zou, "A Fully Integrated HF RFID Tag Chip With LFSR-based Light-weight Tripling Mutual Authentication Protocol," IEEE Access, vol. 7, pp. 73285–73294, 2019, doi: 10.1109/ACCESS.2019.2920437.

IAGA: Interference Aware Genetic Algorithm based VM Allocation Policy for Cloud Systems

Tarannum Alimahmad Bloch[1]
RDIC C.U.Shah.University
Wadhwan, India

Sridaran Rajagopal[2]
FoCA, Marwadi University
Rajkot, India

Prashanth C. Ranga[3]
School of CS, University of Windsor
Windsor, Canada

***Abstract*—Diversified systems hosted on cloud infrastructure have to work increasingly on physical servers. Cloud applications running on physical machines require diverse resources. The resource requirements of cloud applications are fluctuating based on the resource intensity of the applications. The multi-tenancy of Cloud servers can be achieved based on effective resource utilization. The optimum resource utilization, maximum service level agreement, and minimization of interference are the major objectives to be achieved. Using live Virtual Machine (VM) migration techniques cloud resources can be utilized efficiently. But the migrated VMs can interfere with the ongoing applications on the targeted server which may lead to the service level agreement violation (SLAV) and performance degradation. To resolve this issue, understanding the current state of cloud hosts before the allocation of newly migrated VM is necessary. This paper presents Interference Attentive Genetic Algorithm (IAGA) based VM allocation strategy to achieve the aforementioned objectives. The proposed IAGA policy has outperformed existing policies for quantifiable performance metrics such as energy consumed by cloud systems, count of hosts shut down, average SLAV, and count of VM migrations.**

Keywords*—*Cloud computing; interference; VM allocation; SLA violation; resource utilization

I. Introduction

The term Cloud is buzzing technology that can be viewed as the provision of services over the Internet as per the demand of the users. The changeover of large organizations from the traditional Capital Expenditure (CapEx) model to the Operating expenses (OpEx) model supports the reality that the Cloud environment is one of the majority capable technologies in the dated digital era. The escalating number of cloud service consumers has amplified the challenges faced by the Cloud Service Providers (CSPs) to provide the requested services with high availability and reliability of the services. The virtualization technique provides the CSPs to meet these challenges. The main basis of the cloud environment is virtualization.

In virtualized scenarios, the hardware resources of every host also called Physical Machine (PM) are imitated to be independently running entities that are represented as a virtual machine (VM). In the host machine, the cloud service request from the user is managed by the VM to fulfill the computational resource demands such as the size of the memory, computing duration, CPU cycles, network bandwidth, etc. Various cloud systems may support such as the size of the memory, computing duration, CPU cycles, network bandwidth, etc. Various cloud systems may support such virtualizations, in different ways such as operating system-level virtualization, and virtualization based on type-I and type-II [1].

As and when the demand for any VMs' resource increases that particular set of VMs, in the running state, needs to be migrated and accommodated to the new host or physical machine. This entire procedure is known as live VM migration. The VM allocation is the sub-process of the full-stack live VM migration process. Once the migration decision is made by the virtual machine monitor of a particular host, the host needs to look for a new physical machine by keeping various constraints in an account. Such constraints include uninterrupted resource sharing of the existing VMs host with the migrated host after allocation.

The performance of the overall host should not be regretted because of the resource claim of newly migrated VMs and the SLA of the applications hosted on the migrated VMs should not be violated. An interference-aware technique of VM allocation is a must to achieve the aforementioned objectives. This paper discusses an Interference Attentive Genetic Algorithm-based VM allocation policy named IAGA. It is designed to achieve interference minimization by allocating the migratable VMs to the best-suited PM while maintaining high SLA and optimized resource utilization of cloud hosts.

In general, this research article aims to:

1) Discuss the state-of-art current trends in VM allocation policies.

2) The involvement of Genetic Algorithms in cloud service delivery.

3) Mathematical model, design constraints, and algorithm design of proposed technique for VM allocation using the genetic algorithm to address resource interferences.

4) Experimental result analysis of the proposed approach with the existing approaches.

The structure of the paper is as follows: Section 2 discusses the state-of-the-art and the details of genetic algorithm variants used by various researchers. Section 3 gives elaboration on the proposed system model along with the mathematical model, design constraints, and algorithm design. Section 4 talks about the experimental scenario and achieved results. Section 5 concludes the paper and the possible future extension of the proposed research.

II. STATE-OF-THE-ART

VM allocation is the major decision involved in the live VM migration procedure which raises many concerns such as proper resource utilization, increased throughput, SLA maintenance, energy consumption minimization, etc. Various optimization techniques have been applied by the researchers. According to Christina et al, VM allocation is a multi-objective constrained optimization NP-Hard problem [2].

Yuzhe et al proposed a system for optimizing VM allocation techniques in uncertain cloud environments based on the user requirements. The authors considered the optimization perspective of energy consumption for the data center for VMs with no special needs. For the remaining Virtual machines, the present throughput of the hosts and the service consumer's bandwidth requirements are considered in the allocation process of Virtual machines. They designed a VM allocation system to significantly improve multiple objectives such as proper resource utilization, minimization of PMs used, and minimization of energy consumption by taking into consideration the cost of data transmission between VMs [3].

As per Jenn-Wei et al, without considering the VM interference in the VM placement requirement, the Quality of Service (QoS) requirements of the cloud application executing in VMs may be violated. The authors have considered three factors in the proposed VM placement policy: (i) Resource demand of virtual machines (ii) The QoS of cloud applications (iii) The VM interference. As per the authors' point of view, it is difficult to accommodate all the aforementioned factors in the VM placement policy. Authors have named research problem as IAVMP – Interference Aware VM Placement problem. They have formulated an integer linear programming model to solve IAVMP as an NP-complete problem [4].

Sasmita et al considered a single parameter optimization technique to achieve the goal of cloud service consumers or cloud service providers. In realism, the user and service provider have opposing goals. This objective could be used to guide the selection of cloud hosts. They proposed an Efficient Multi-optimization Resource Allocation (eMRA) model using optimization techniques to achieve the goals of cloud service consumers and data centers in the proposed work. SGO (Social Group Optimization) technique is proposed to improve user requests by taking into account related parameters for allocation. Similarly, Particle Swarm Optimization (PSO) is being used to improve data center lists that are fitting for optimized user requests. To design the model that separates the proposed design model from other existing works, the eMRA considers distinct related parameters of cloud service consumer request, cloud host, and network. The eMRA technique is simulated using CloudAnalyst, and the authors researched ten distinct scenarios using three different CloudAnalyst broker policies [5].

S. Savitha et al. presented a perceptive priority-aware VM allocation strategy called the P-PAVA algorithm which determines an application's priority and also resource needs. Using an ML-based prediction model, the system assigns applications based on one's priority. Moreover, parallelization has been used before conveying various workloads to reduce the overhead of the allocation algorithm. To accomplish this, the algorithm uses the first fit technique as a baseline for user request allocation with a low priority norm. P-PAVA outshines the state-of-the-art algorithm for VM allocation for priority-aware applications on a variety of indicators such as average response time, execution time, and power consumption [6].

Garg R. et al. developed a virtual machine allocation policy to evaluate the behavior of commonly used heuristic allocation policies. To substantiate the logistic comparison, a policy that implements a bin-packing approach in Virtual machine allocation has been developed. The implemented research problem has been rigorously tested using a diverse range of workload data sets and experimental configurations. The performance of these algorithms is also evaluated with varying threshold and VM selection policies. Experimental results indicate that policies that take into account the server's power and computing capacity perform much better in almost all scenarios [7].

Jitendra et al. created an Intelligent SLA-aware and Energy Minimization VM allocation approach that uses the Emperor Penguin Optimization (EPO) algorithm. In a heterogeneous cloud environment, the system could indeed allocate virtual machines based on power usage. The proposed method has illustrated its appropriateness for virtual machines in the data center by making comparisons of it to the Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), and Binary Gravity Search Algorithm (BGSA). The outcomes of the proposed system have been evaluated by using the JAVA simulation platform. The investigation results reveal that the improved EPO-based system is very effective in limiting energy consumption, SLA Violations (SLAV), and the advancement of QoS requirements to provide desirable cloud services [8].

Rahimi Zadeh et al. [9] proposed an interference-aware and joint profit scheduling scheme (PIAS) to proficiently consolidate Virtual machines on the physical machine that hosts multi-tier application workloads in Infrastructure-as-a-Service (IaaS). For consolidating VMs, the PIAS policy took into account resource usage, costs and profits of power consumption, service level agreements (SLAs), and operational interference of VMs along with the number of transferred memory pages during live VM migration. The functioning behavior patterns of VMs are examined in this work. Furthermore, an optimization problem is presented to achieve maximum cloud service provider profit while reducing the overall cost of application workload executions.

From the various literature [11-15], it is clear that the impact of the interference is unavoidable to achieve high SLA. Recently also many researchers have developed VM migration techniques by keeping the different objectives in the center. CMIG – is a Concurrency-aware Migration System [10]. Rachael et al [11] used an intelligent approach to minimize interference and energy consumption. Anu et al [12] addressed interference minimization in the proposed system Interference Aware Live Migration (IALM). Babu et al [13] have designed an interference-aware system that has automatic scaling support to handle sudden load drift with precise prediction and minimum VM migration. Chao et al have addressed the interference awareness during the migration process to mitigate cache-based side challenge attacks in the cloud [14]. Yiling et al have proposed the live migration technique which applies to software-defined networks with awareness of interference and topology [15].

Neha et al [38], proposed an optimized VM allocation scheme called Resource Aware Provisioning (RAP). For the VM allocation criteria, the authors considered better energy efficiency and a fixed upper threshold value. RAP is working with the static upper threshold and the authors have not considered all types of required resources for different applications running inside VMs. The limitation of the RAP is that it has not taken the future resource type intensity into the consideration while choosing the PM for VM allocation. For example if one of the newly migrated VM starts grabbing more CPU cycles and if the PM is about to reach the upper threshold then any proactive steps are not taken by the authors. Only the energy consumption has been considered as a major parameter for the VM allocation which is not sufficient for the optimized VM allocation approach because interference minimization is an unavoidable parameter that has not been considered by RAP.

Genetic Algorithms have been applied significantly during the current era to achieve the optimized solution to various research problems. It has been observed from various works of literature that the involvement of genetic algorithms and variants has given improved results in the various cloud computing-related research problem solutions. The genetic algorithms have been applied to address the cloud security related issues [16-18], to achieve optimum solutions in the Internet of Things (IoT) service placement in fog computing environment cloud computing [19], and to leverage the fog computing frameworks the human activity recognition has been achieved through deep genetic algorithm [20].

Genetic algorithms are considerably being used for a better cloud service experience. The list of the Genetic algorithm and its variants applied in recent research trends by various researchers [21-32] for better cloud service is presented in Table I.

TABLE I. APPLICATION SUMMARY OF GENETIC ALGORITHM VARIANTS IN CLOUD SERVICE DELIVERY

Reference No.	**Authors**	**The variant of the Genetic Algorithm**	**Application**	**Experiment Environment**
[21]	Shilpa K., Sujata T.	Hybrid Max-Min Genetic Algorithm	Load Balancing	CloudSim Simulator
[22]	Jiawei L., Haotian Z., Wei Z., Jie Li, Gang ., Zhenbo C.	Improved Genetic Algorithm	Optimal VM Placement Strategy	Real Environment FEA - Finite Element Analysis
[23]	Zhang, B., Hao W., Xiao W.	Cluster-based Genetic Algorithm	VM Placement Strategy	Not mentioned
[24]	Mehran T., Mohammad I., Mostafa G.	Micro-genetic algorithm	VM Allocation Strategy	CloudSim Simulator
[25]	Huda I., Khaled E., Raafat O. A.,	Adaptive Genetic Algorithm	Dynamic Task Scheduling Strategy	CloudSim Simulator
[26]	Muhammad S., Muhammad T.	parallel multi-objective genetic algorithm	Task Scheduling for scientific workflow	CloudSim Simulator
[27]	Abbas A., Ahmad K., Seyed M.G.	Thermal Aware Genetic Algorithm	VM Allocation Strategy	CloudSim Simulator
[28]	Einollah J.G., Amir M. R, Nooruldeen N. Q.	Genetic Algorithm	Service Load Balancing	Matlab
[29]	Carlos G., Isaac L., Carlos J.	Non-dominated Sorting Genetic Algorithm – II	Micro Services Resource Allocation	Real Cloud Environment
[30]	Zhou Z., Fangmin L., Huaxi Z., Houliang X., Jemal H. A., Chowdhury, Morshed U	Improved Genetic Algorithm	Optimized Task Scheduling	CloudSim Simulator
[31]	Madhusudhan H S, Satish, K. T, Syed Mustapha.F. D., Punit G., and Raja P. T.	Genetic Algorithm	Resource Allocation	CloudSim Simulator
[32]	A. J. Miriam· R. Saminathan, S. Chakaravarthi	Non-dominated Sorting Genetic Algorithm – III	Resource Allocation	Hadoop Cluster

III. Proposed System

This section represents the proposed system architecture, mathematical model, design constraints, and algorithm design.

A Genetic Algorithm (GA) is a well-known meta-heuristic algorithm that is inspired by the biological evolution process [40]. In nature, GA mimics the Darwinian theory of the survival of the fittest. GA's fundamental components are chromosome representation, fitness, selection, and biologically inspired operators [40]. Chromosomes can be thought of as points in the solution space. These are processed by iteratively replacing the population with genetic operators. The fitness function is used to assign a value to each of the population's chromosomes [39].

In selection operation, chromosomes are chosen for further processing based on achieved fitness value. In the crossover operator, a random convergence point is chosen and the subsequences between chromosomes are changed to produce offspring [40].

For the proposed research problem the chromosomes are considered as the PMs to allocate the migrated VMs. The fitness function is designed by considering the total resource capacity of the server.

Crossover is performed to choose the best PMs by keeping the future resource demands of the VMs hosted on that PM. The following section discusses the proposed architecture, design constraints, mathematical model of the research problem, and ultimately the algorithm design presented thereof.

A. Proposed System Architecture

Fig. 1 depicts the system architecture of the proposed VM allocation system. In the cloud environment, all the PMs resource usage will be monitored especially at the time when that PM is the candidate to be chosen for the accommodation of a new set of migratable VMs. A genetic algorithm is applied in the proposed system to choose the best suitable PM to allocate migratable VMs.

In the proposed architecture, the major components are:

- Local Resource Monitor.
- Local Interference Monitor.
- Global Interference Monitor.
- Historical Resource Usage.
- Resource Usage-based VM Categorization System.

The local resource monitor keeps track of resource usage based on the resource intensity type and the local interference monitor observes the interference by considering SLA. From the study of existing systems, it is clear that during live VM migration, allocation of optimum resources and SLA management is the major objective. To achieve this research objective the proposed system must be clear about the available resource capacity of the chosen PM for the allocation of migratable VMs. The candidate PMs mostly has varying resource capabilities. The used resource amount of PMs depends on the hosted applications running on the VMs of that server. Based on this scenario the application need or VM resource intensity can be categorized as:

- CPU intensive VMs.
- Memory intensive VMs and.
- Network bandwidth-intensive VM.

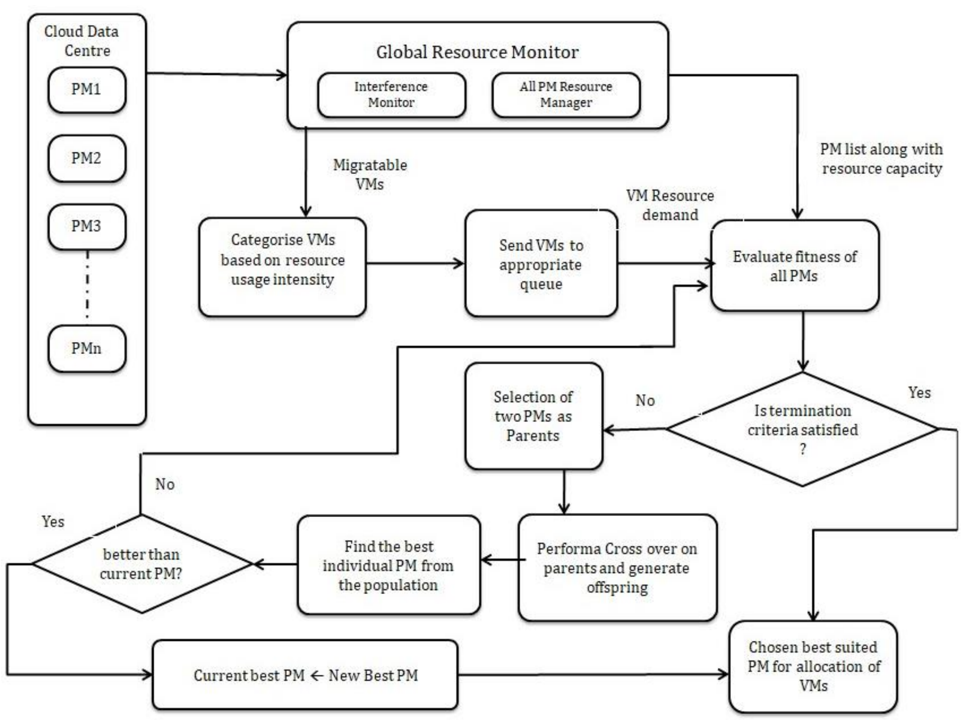


Fig. 1. Proposed System Architecture.

It is necessary to allocate the VMs based on the available resource type intensity of the PM. So in the proposed design the interference monitor categorizes VMs and aligns them to the appropriate queue before allocation using a resource usage-based VM categorization system. Categorization is performed based on the prediction carried out using historical resource usage. The global resource monitor keeps track of all the PM resource usage states and lists of migratable VMs selected by the overloaded PM in case of excessive resource usage. Ultimately the decision support system takes care of the migration process to the chosen destination PM. The list of chosen PMs and migratable VM list will be given as input to apply the genetic algorithm. The strength of the proposed solution is the fitness function evaluated for all the PMs. The fitness function is discussed in Section III C.

B. Proposed Mathematical Model

Let V be the set of n migratable VMs.

$$\mathrm{V} = \{\mathrm{VM}_1, \mathrm{VM}_2, \mathrm{VM}_3, \ldots \ldots \mathrm{VM}_\mathrm{n}\}$$

where n = number of migratable VM.

Let P be the set of PMs in the cloud environment.

$$\mathrm{P} = \{\mathrm{PM}_1, \mathrm{PM}_2, \mathrm{PM}_3, \ldots \ldots \mathrm{PM}_\mathrm{m}\}.$$

where m = number of PMs in set P.

Logically overall performance degradation due to the VM migration interferences on PM denoted by MI can be calculated as the total of co-location interference that occurred because of the resource share claim of newly migrated VMs known as co-location interference MIC and network interference is migration network time MNW.

So, for PM the total interference

$$MIC_i = CI_i + MNW_i \tag{1}$$

Based on the resource interference taxonomy provided in [33] the performance degradation due to the co-location interference is the total of CPU contention interference, memory consumption interference, and network bandwidth contention interference. Here co-location interference is denoted by CI. So, for i^{th} PM the co-location interference can be defined as:

$$CI_i = M_c + M_m + M_{nb} \tag{2}$$

Where

M_c is CPU contention time.

M_m is memory contention time, and

$M_{mb}$is network bandwidth contention time.

The CPU contention time $\mathrm{M_c}$ is the ratio of C_{dem} CPU cycles in demand C_{dem} and C_{avl} available CPU cycles that can be allotted. M_c can be represented as:

$$M_c = C_{dem} / C_{avl} \tag{3}$$

The severity of the performance degradation increases if the ratio increases. The demand for the CPU cycles C_{dem} can be defined as the sum of the CPU cycles being used by VMs in execution C_e and the waited CPU cycles by the VMs in the waiting queue C_{que}.

$$C_{dem} = C_e + C_{que} \tag{4}$$

Therefore, the overall performance degradation of i^{th} PM due to the network interference and co-location interference can be thought of as under:

$$I_i \approx f(CI_n, MNW_k)$$

So,

$$I_i = a \cdot CI_n + b \cdot MNW_k \tag{5}$$

where a and b are constants that are used to regulate the values.

C. Proposed System Design Constraints

At the time of VM allocation policy design, for k number of PM used from the P set of available physical machines and V set of migratable VMs there are certain design constraints (DC) that are mandatory to be considered:

1) Design Constraint 1: Each VM has to be allocated to one and only one physical machine

$$\mathrm{V} = \bigcup_{\mathrm{P_k} \in \mathrm{P}} \mathrm{V_k}$$

2) Design Constraint 2: To minimize the co-location interference, the requested resources such as CPU cycles, network bandwidth, and memory should not exceed the total capacity of the server.

$$\sum_{i=1}^{n} \mathrm{V_i^{CPU}} \leq \sum_{k=1}^{m} \mathrm{P_k^{CPU}}$$

$$\sum_{i=1}^{n} \mathrm{V_i^{mem}} \leq \sum_{k=1}^{m} \mathrm{P_k^{mem}}$$

$$\sum_{i=1}^{n} \mathrm{V_i^{bw}} \leq \sum_{k=1}^{m} \mathrm{P_k^{bw}}$$

Where,

V_i^{cpu}= CPU contention of i^{th} VM

P_i^{cpu}= CPU contention of i^{th}PM

V_i^{mem} = memory contention of i^{th} VM

P_i^{mem} = memory contention of i^{th} PM

V_i^{bw} = network bandwidth of i^{th} VM

P_i^{bw} = network contention of i^{th}PM

3) Design Constraint 3: The total allocation capacity of the server for i^{th}PM to proceed with the new allocation is defined by SC:

$$SC = \sum_{P_i \in P} u_i^{cpu} + \sum_{Pi \in P} u_i^{mem} + \sum_{P_i \in P} u_i^{bw} \tag{6}$$

where,

$u_i^{cpu}$is unused CPU cycles from the total CPU cycles of i^{th} PM consider it as x.

u_i^{mem} is unused memory from the total memory of i^{th}PM consider it as y.

u_i^{bw} is unused network bandwidth from the total network bandwidth of i^{th} PM consider it as z.

A genetic algorithm [34] is a method of an arbitrarily defined searching method with improved optimization and autonomic implied parallelism. GA can be used to automatically govern the search direction using probability, as well as to acquire and instruct the optimum searching space [35, 36]. Considering the advantages of the genetic algorithm mentioned in [34, 35, 36], the proposed algorithm employs an interference attentive genetic algorithm-based VM allocation strategy for live VM migration in a computing environment.

This method calculates resource usage in advance by taking historical data and current states into account, which will have an impact on the entire cloud system. By considering the above-described design constraints a genetic algorithm-based Interference Attentive VM allocation algorithm has been designed.

This research solution focuses on the total resource capacity of the server denoted as server capacity SC before allocating VMs to it.

As per design constraint number 3, there is a need to reduce the sum deviating from SC i.e. $|\ x + y + z - SC\ |$ should be zero. Hence the fitness can be considered as the inverse of $|\ x + y + z - SC\ |$.

So,

$$f(x) = 1\ \div\ |\ x + y + z - SC\ | \tag{7}$$

where

$f(x)\ \in\ [0, 1]$

In the proposed solution fitness function is the driving factor for VM allocation. The fitness function is designed by considering total server capacity which considers unused CPU cycles, memory, and bandwidth. Before allocation, the migrated VM demand and available server capacity are being evaluated for choosing the optimal solution to the VM allocation problem. The interconnections of the applied solution mathematically prove that the proposed fitness function will reduce the interference raised by migratable VMs.

D. Proposed Algorithm

The section discusses the proposed algorithm design. The VM allocation to the PM should be based on the probability of fitness value of a particular PM's resource utilization described in equation (7).

The design of the IAGA – Interference Attentive Genetic Algorithm-based VM allocation policy is in the direction of achieving the aforementioned research objective while simultaneously keeping discussed design constraints in mind. There are various strategies for finding the best genes for step 6.

The fitness ratio-based selection algorithm is used in this research. The calculation of the fitness value for each PM in the present PM population has been derived first and then kept for the individual PM with the highest score in the next generation. After that for allocation, every PM probability of the accommodation based on the fitness ratio is calculated.

Algorithm: Proposed VM allocation policy

Input PM: PM list, VM: Migratable VM list
Output P: best suitable PM for allocation of migratable VMs

Procedure:

```
 For all the PMs in the list initialize the random population
   Categorize VMs into different categories based on the resource intensity as well the prior history of resource utilization
   Send VM to the appropriate queue
 Evaluate fitness based on utilization of resources for each PM based on eq. (7)
 While termination condition (! = generation count) do
    Parents ← select two parent individuals from the PM list according to the fitness value.
    For each parent1, parent2 do
         Offspring1, offspring2 ← Crossover (parent1, parent2)
         Apply mutation on generated offspring
         Find the best individual in the PM population
         If the best individual in the population is better then
              current best individual PM ← new best individual PM
         End If
  End For
End While
Return best-suited PM
```

E. Performance Metrics

Quantitative parameters such as energy consumption, amount of VM migrations, amount of host shutdowns, and average service level agreement violations are all properly considered. The following are explanations of the aforementioned performance matrices:

- Energy Consumption: The sum of energy consumed by each host during the overloading, underloading, and migration procedures.
- The number of VM migrations: The number of VMs migrated from one host to another during the allocation procedure while migrating.
- The number of host shutdowns: The number of host shutdowns performed during the migration process to minimize energy consumption as much as possible.
- Average Service Level Agreement Violation (SLAV): The average value of quantitative service level agreement violation due to migration and/or interference across all VMs.

IV. Experimental Setup and Result Discussion

The live VM migration can be considered as the whole process of host overload or underload detection, choosing the VMs for the migration and placing these VM in the target host. The term VM allocation policies are interchangeably used with PM selection policies. From the various literature discussed here, it is clear that while allocating the host to the migratable VMs different researchers have kept different research goals into consideration. Some of the common goals are energy minimization and SLA maximization. The major diverse effect of the live VM migration is the interference which has been considered by very few researchers. For the VM allocation strategy, one of the major goals should be the proactive strategy about the co-location interferences that may occur after the allocation VMs to the targeted PM along with optimum resource allocation to the ongoing VMs and newly migrated VMs. High SLA can be achieved if the resource allocation is as per the promised amount. The resource allocation could be proper if the newly migrated VMs are not exceeding their share by creating trouble for ongoing VMs. The interference awareness is the novel approach applied to the VM allocation policy to achieve multiple aforementioned research objectives.

To implement and test the proposed system the CloudSim simulator has been used. The CloudSim toolkit, developed by Rajkumar Buyya et al., was used for the experimental implementation [37], and then after the CloudSim is being used as a low-cost simulation environment for cloud-based research projects. It is an open-source simulation tool that most researchers use to simulate the cloud environment. The workload dataset compiled by PlanetLab has been used in the experimentation. The workloads used in this investigation are 20110303, 20110325, and 20110420. PlanetLab workloads are also included in the CloudSim simulator package.

Certain algorithms for VM allocation and VM selection approaches are also available in the CloudSim tool kit. Virtual machine allocation policies such as Static Threshold – THR, Median Absolute Deviation – MAD, Inter-Quartile Range – IQR, Local Regression – LR, and Local Robust Regression – LRR are given in the simulator. There are certain policies for VM selection, which include Maximum Correlation – MC, Minimum Migration Time – MMT, and Random Selection – RS. These are the benchmark algorithms used by many researchers for the comparisons of proposed systems.

LR and MAD VM allocation policies have been considered in the experiment. VM selection policies, MMT, MU, and RS have been used. The validation of the proposed system Interference Attentive Genetic Algorithm IAGA has been carried out against these policies. Workloads have been applied in various configurations of VM allocation and host overload detection policies. The evaluation has incorporated the IAGA policy and studied the imperial results of the aforementioned experiment scenarios. The proposed algorithm was also compared to Neha et al [38] RAP VM allocation policy.

The VM allocation approach applied in RAP is based on the static upper threshold whereas in the modern cloud application the workload fluctuates based on the high and the low number of user requests for a particular cloud application. Due to this, the static threshold will not lead to the appropriate resource utilization. The proposed VM allocation system is developed using a genetic algorithm.

A genetic algorithm is an adaptive heuristic search-based solution that presents the intelligent utilization of search space to solve the optimization problem. From the graphs, it is proved that for the dynamic workload-based cloud systems, the adaptive solution performs better than the static threshold-based policy for the VM allocation problem.

So the given experiments study the effect of the applied genetic algorithm for searching for the best suitable PM for the available list of candidate PMs. So the existing VM allocation policies have been modified to proceed further with the Genetic algorithm principles. The behavior of the system under study has been observed by various combinations specified in the result graphs.

Fig. 2, Fig. 3, and Fig. 4 show the number of hosts shut down for mentioned workloads. The resultant graph indicates that the count of hosts shutdowns is less in IAGA. Less number of host shutdowns will reduce the downtime for the ongoing application on the cloud host. This will lead toward the goal of SLA maximization. It can be observed in Fig. 5, Fig. 6, and Fig. 7, that the IAGA reduces energy consumption by the PMs. Fig. 8, Fig. 9, and Fig. 10 show the number of VM migrations. Average SAL violations achieved through experimentation have been presented in Fig. 11, Fig. 12, and Fig. 13. The results show that the SLA violation is lower for the proposed approach IAGA to ensure more effective service to cloud service consumers compared to the existing policies.

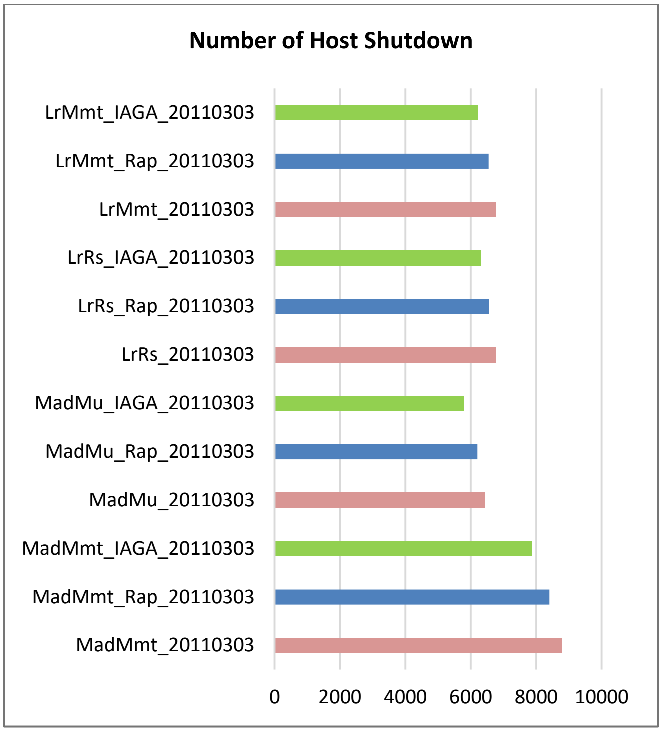


Fig. 2. Number of Host Shutdown for Workload 201100303.

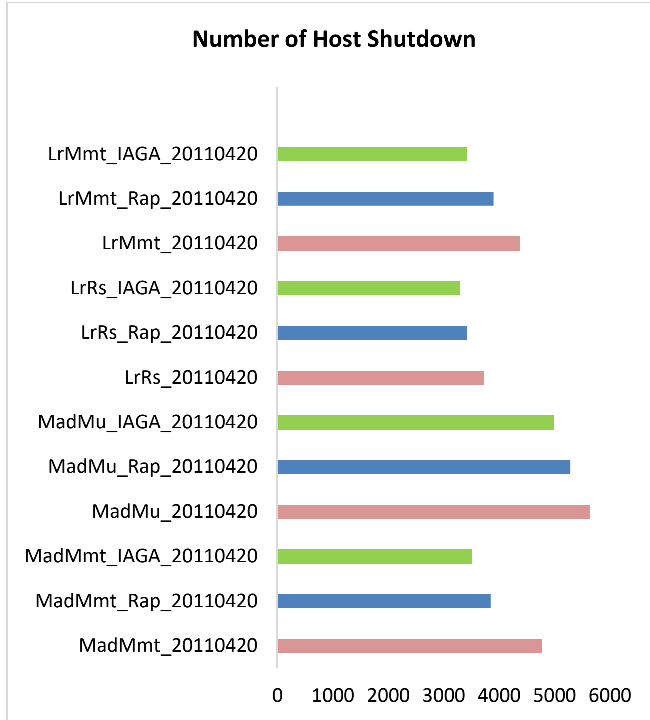


Fig. 4. Number of Host Shutdown for Workload 20110420.

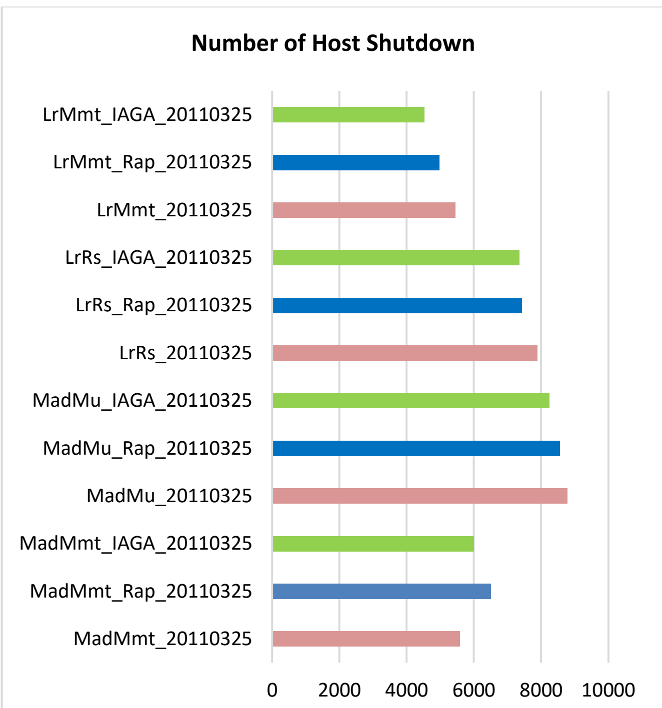


Fig. 3. Number of Host Shutdown for Workload 20110325.

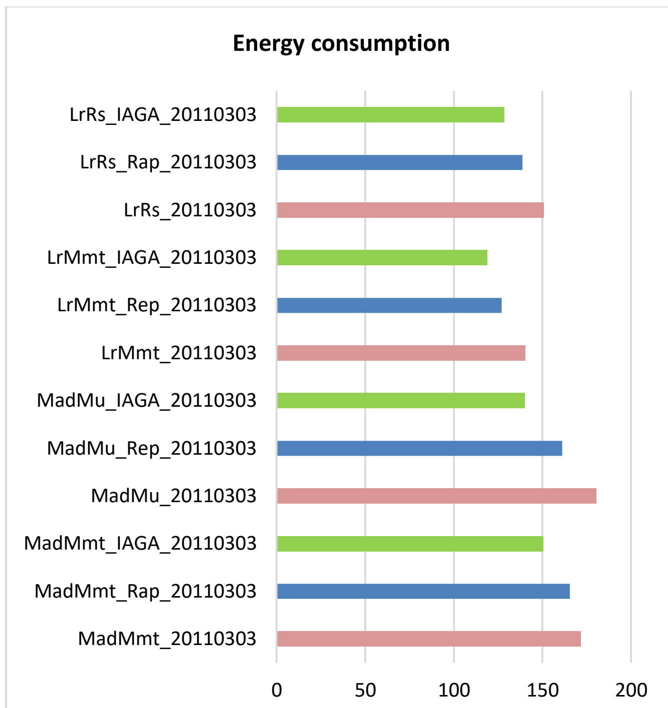


Fig. 5. Energy Consumption for Workload 20110303.

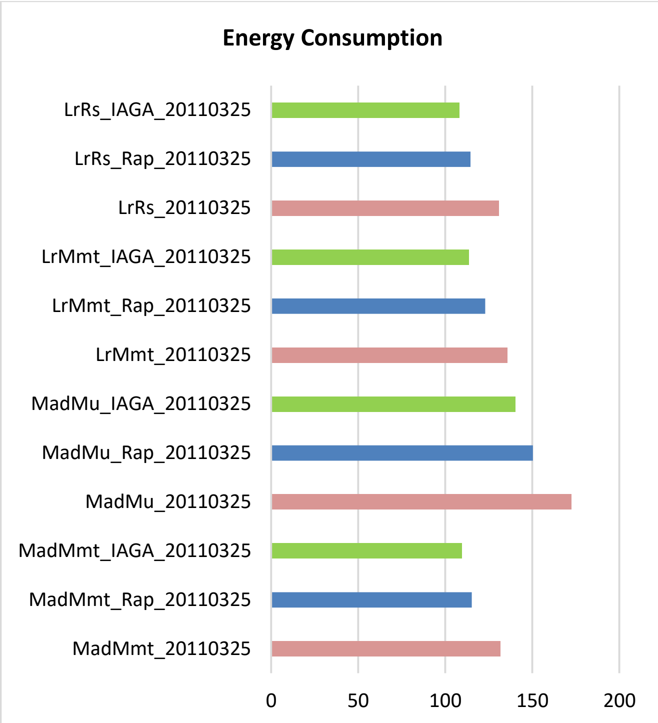


Fig. 6. Energy Consumption for Workload 20110325.

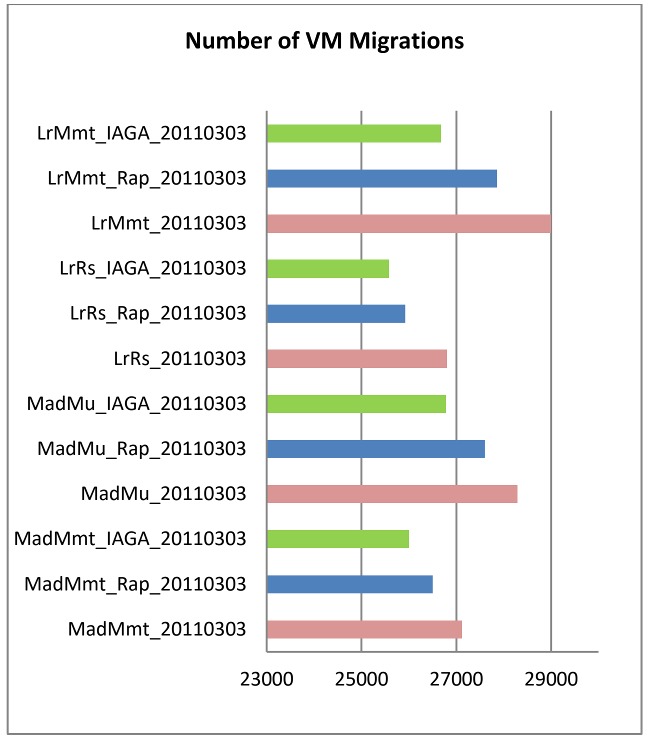


Fig. 8. Number of VM Migrations for Workload 20110303.

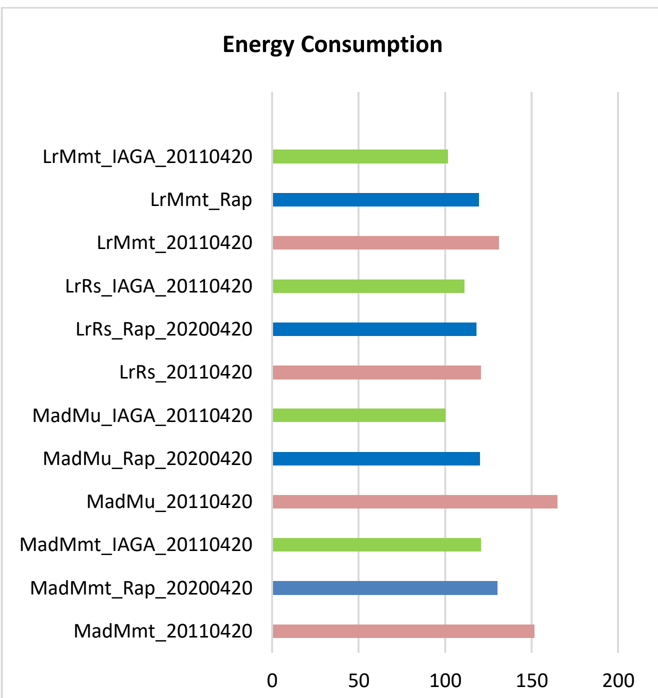


Fig. 7. Energy Consumption for Workload 20110420.

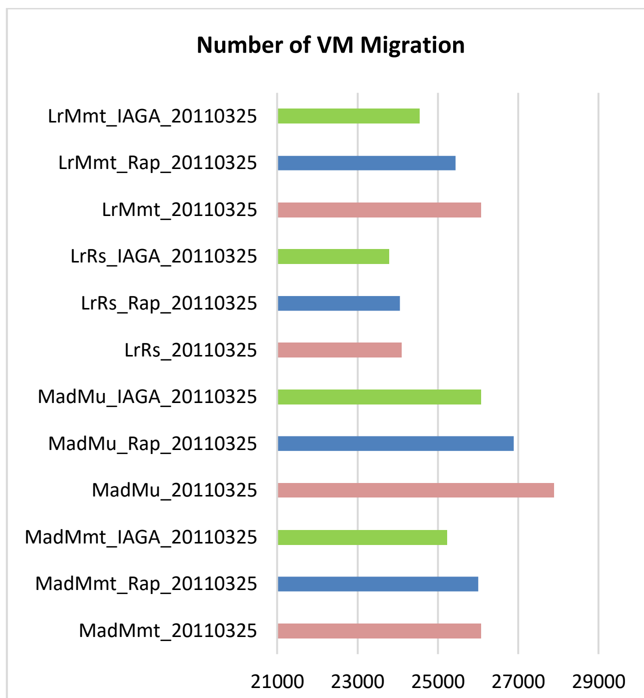


Fig. 9. Number of VM Migrations for Workload 20110325.

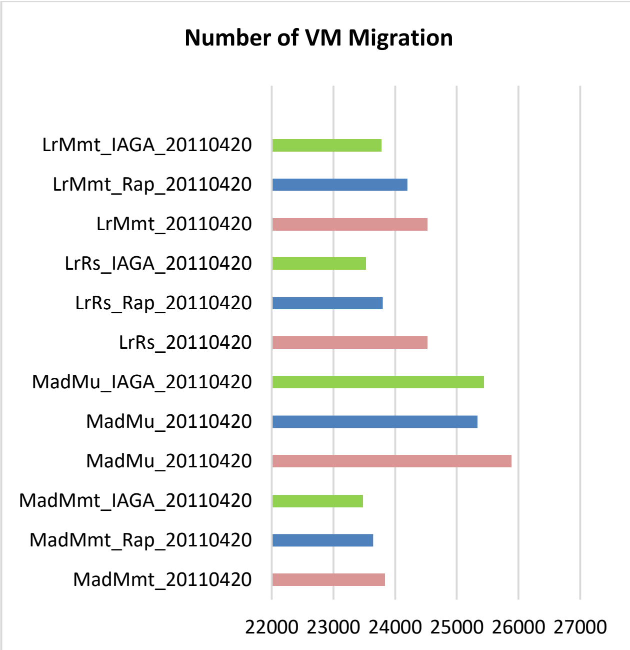


Fig. 10. Number of VM Migrations for Workload 20110420.

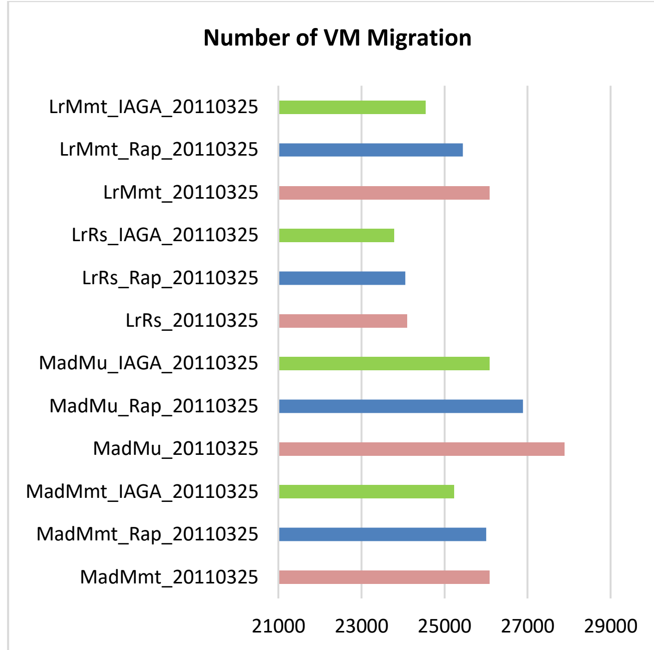


Fig. 12. Average SLA Violation for Workload 20110325.

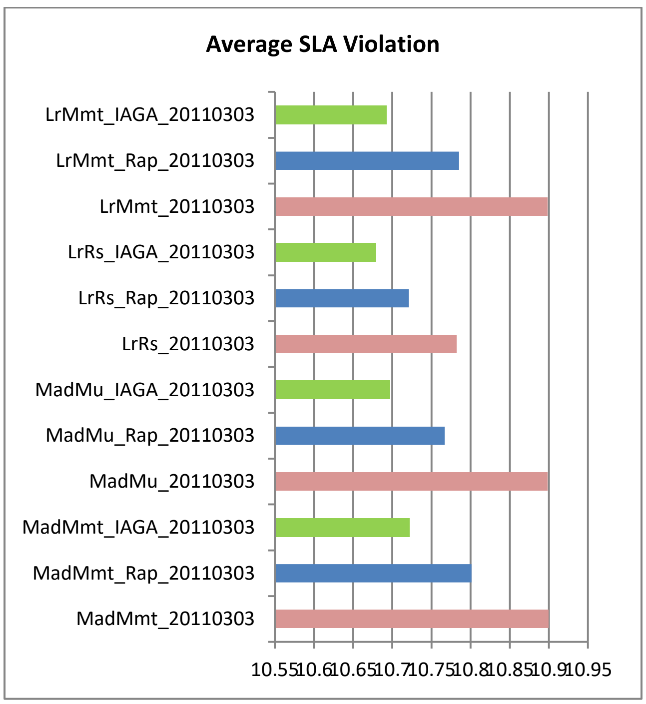


Fig. 11. Average SLA Violation for Workload 20110303.

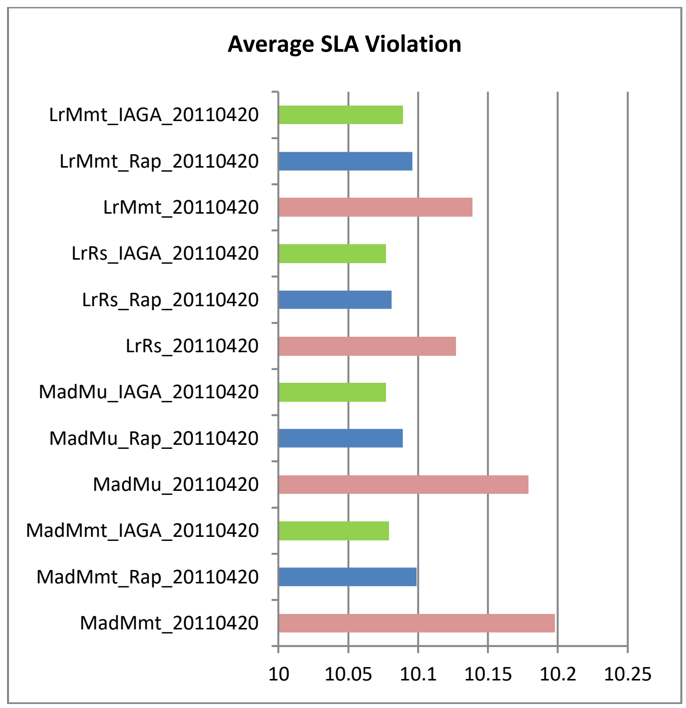


Fig. 13. Average SLA Violation for Workload 20110420.

V. Conclusion and Future Work

The movement of VMs across the system for proper resource utilization is one of the solutions that may result in outlays and service interruptions, resulting in a drop in service quality by affecting SLA. The data centers in a cloud environment require a large number of VMs to run continuously with a high SLA demand. Proposed research experimentation focuses on interference attentive VM allocation. The main challenge of the proposed research work was to strike a balance between effective server resource utilization and the preservation of cloud resources while adhering to SLA constraints with minimized interference effects. The fitness function is a critical component of the proposed method for reducing interferences in Infrastructure as a Service (IaaS) Cloud data centers as it helps the proposed algorithm to find the best optimum PM from all the candidate PMs for the allocation of migratable VMs. One of the most inescapable principles of cloud computing systems that should be pursued is the principle of service quality. As a result, in proposed work concentrated on cloud computing systems competence characteristics such as the current state of PM computing resources specifically resource usage and demand queue to ensure the expected quality of service with minimal interferences from newly moved VMs during live VM migration. To improve the efficiency of the traditional algorithms, a Genetic Algorithm has been applied for VM allocation; the paper proposes a VM allocation policy to address the reduced number of VM migrations, reduced energy consumption, and low SLA violations. The promising results obtained from the proposed approach reveal that the genetic algorithm can be efficiently applied in real data centers to achieve interference minimizations. Furthermore, as the energy consumption decreases, it can also be used in green data centers. The proposed system has achieved the research objectives with a noticeable improvement compared to the existing approaches. The proposed research solution is useful to the cloud service provider to enhance the SLA and achieve satisfactory resource allocation by using the cloud server capacity at the optimum level. As of now, the proposed approach has been tested in the Cloud simulation environment. In the future, this approach could be extended by implanting it in a real cloud environment, the behavior of which could be studied and could be validated for other workloads and/or live streaming applications.

References

[1] Tzenetopoulos, Achilleas, et al. "Interference-aware workload placement for improving latency distribution of converged HPC/Big Data cloud infrastructures."

[2] Christina Terese Joseph, K. Chandrasekaran, Robin Cyriac, “A Novel Family Genetic Approach for Virtual Machine Allocation,” Procedia Computer Science, Volume 46,2015, Pages 558-565, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2015.02.090.

[3] Y. Huang, H. Xu, H. Gao, X. Ma, and W. Hussain, "SSUR: An Approach to Optimizing Virtual Machine Allocation Strategy Based on User Requirements for Cloud Data Center," in IEEE Transactions on Green Communications and Networking, vol. 5, no. 2, pp. 670-681, June 2021, DOI: 10.1109/TGCN.2021.3067374.

[4] Jenn-Wei Lin and Chien-Hung Chen, "Interference-aware virtual machine placement in cloud computing systems," 2012 International Conference on Computer & Information Science (ICCIS), 2012, pp. 598-603, DOI: 10.1109/ICCISci.2012.6297100.

[5] Parida, S., Pati, B., Nayak, S.C. et al. “eMRA: an efficient multi-optimization based resource allocation technique for infrastructure cloud” J Ambient Intell Human Comput (2022). https://doi.org/10.1007/s12652-021-03598-8.

[6] S. Savitha and S. Salvi, "Perceptive VM Allocation in Cloud Data Centers for Effective Resource Management," 2021 6th International Conference for Convergence in Technology (I2CT), 2021, pp. 1-5, DOI: 10.1109/I2CT51068.2021.9417960.

[7] Garg R., Arora I., Gupta A. (2021) “Performance Evaluation of VM Allocation Strategies on Heterogeneous Environments in Cloud Data Center,” In Panigrahi C.R., Pati B., Pattanayak B.K., Amic S., Li KC. (eds) Progress in Advanced Computing and Intelligent Engineering. Advances in Intelligent Systems and Computing, vol 1299. Springer, Singapore. https://doi.org/10.1007/978-981-33-4299-6_43.

[8] Jitendra Kumar Samriya, Subhash Chandra Patel, Manju Khurana, Pradeep Kumar Tiwari, Omar Cheikhrouhou, "Intelligent SLA-Aware VM Allocation and Energy Minimization Approach with EPO Algorithm for Cloud Computing Environment," Mathematical Problems in Engineering, vol. 2021, Article ID 9949995, 13 pages, 2021. https://doi.org/10.1155/2021/9949995.

[9] Rahimi Zadeh, K., Dehghani, “A. Design and evaluation of a joint profit and interference-aware VMs consolidation in IaaS cloud datacenter,” Cluster Comput 24, 3249–3275 (2021). https://doi.org/10.1007/s10586-021-03310-7.

[10] He T, Toosi AN, Buyya R. “CAMIG: Concurrency-Aware Live Migration Management of Multiple Virtual Machines in SDN-enabled Clouds,” IEEE Transactions on Parallel and Distributed Systems. 2021 Dec 28.

[11] Shaw, Rachael, Enda Howley, and Enda Barrett. "An intelligent ensemble learning approach for energy-efficient and interference aware dynamic virtual machine consolidation," Simulation Modelling Practice and Theory 102 (2020): 101992.

[12] Anu, V. R., and Sherly Elizabeth. "IALM: Interference Aware Live Migration Strategy for Virtual Machines in Cloud Data Centres," Data Management, Analytics, and Innovation. Springer, Singapore, 2019. 499-511.

[13] Babu, KR Ramesh, and Philip Samuel. "Interference aware prediction mechanism for auto-scaling in the cloud," Computers & Electrical Engineering 69 (2018): 351-363.

[14] Yang, Chao, et al. "Interference-based VM Migration to Mitigate Cache-based Side-channel Attacks in Cloud." 2018 IEEE 4th International Conference on Computer and Communications (ICCC). IEEE, 2018.

[15] Qin, Yiling, et al. "Interference and topology-aware VM live migrations in software-defined networks," 2019 IEEE 21st International Conference on High-Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS). IEEE, 2019.

[16] Tahir, Muhammad, et al. "CryptoGA: a cryptosystem based on genetic algorithm for cloud data security," Cluster Computing 24.2 (2021): 739-752.

[17] Denis, R., Madhubala, P. “Hybrid data encryption model integrating multi-objective adaptive genetic algorithm for secure medical data communication over cloud-based healthcare systems,” Multimed Tools Appl 80, 21165–21202 (2021). https://doi.org/10.1007/s11042-021-10723-4.

[18] Balakrishna, S.K., Shetty, S.M., Martis, J.E., Ramasamy, B. (2022). “Genetic Algorithm-Based Pseudo-Random Number Generation for Cloud Security.,” In Misra, S., Kumar Tyagi, A., Piuri, V., Garg, L. (eds) Artificial Intelligence for Cloud and Edge Computing. Internet of Things. Springer, Cham. https://doi.org/10.1007/978-3-030-80821-1_10.

[19] B.V. Natasha, Ram Mohana Reddy Guddeti, “Adopting elitism-based Genetic Algorithm for minimizing multi-objective problems of IoT service placement in the fog computing environment,” Journal of Network and Computer Applications, Volume 178, 2021, 102972, ISSN 1084-8045, https://doi.org/10.1016/j.jnca.2020.102972.

[20] R. Raja Subramanian, V. Vasudevan, “A deep genetic algorithm for human activity recognition leveraging fog computing frameworks,” Journal of Visual Communication and Image Representation, Volume

77, 2021, 103132, ISSN 1047-3203, https://doi.org/10.1016/j.jvcir.2021.103132.

[21] Kodli, Shilpa, and Sujata Terdal. "Hybrid max-min genetic algorithm for load balancing and task scheduling in a cloud environment," Int J Intell Eng Syst. 14.1 (2021): 63-71.

[22] Lu, Jiawei, et al. "Optimal machine placement based on improved genetic algorithm in cloud computing," The Journal of Supercomputing 78.3 (2022): 3448-3476.

[23] Zhang, Binbin, Xiao Wang, and Hao Wang. "Virtual machine placement strategy using cluster-based genetic algorithm," Neurocomputing 428 (2021): 310-316.

[24] Xue, Xingsi, et al. "Optimizing ontology alignment through linkage learning on entity correspondences," Complexity 2021 (2021).

[25] Huda Ibrahim, Raafat O. Aburukba, Khaled El-Fakih, “An Integer Linear Programming model and Adaptive Genetic Algorithm approach to minimize the energy consumption of Cloud computing data centers,” Computers & Electrical Engineering, Volume 67,2018, Pages 551-565, ISSN 0045-7906, https://doi.org/10.1016/j.compeleceng.2018.02.028.

[26] Sarfaraz M, Tahir M. “A parallel multi-objective genetic algorithm for scheduling scientific workflows in cloud computing,” International Journal of Distributed Sensor Networks. August 2020. doi:10.1177/1550147720949142.

[27] Akbari, A.; Khonsari, A.; Ghoreyshi, S.M. “Thermal-Aware Virtual Machine Allocation for Heterogeneous Cloud Data Centers.” Energies 2020, 13, 2880. https://doi.org/10.3390/en13112880.

[28] Jafarnejad Ghomi, Einollah, Amir Masood Rahmani, and Nooruldeen Nasih Qader. "Service load balancing, scheduling, and logistics optimization in cloud manufacturing by using genetic algorithm," Concurrency and Computation: Practice and Experience 31.20 (2019): e5329.

[29] Guerrero, C., Lera, I. & Juiz, C. “Genetic Algorithm for Multi-Objective Optimization of Container Allocation in Cloud Architecture,” J Grid Computing 16, 113–135 (2018). https://doi.org/10.1007/s10723-017-9419-x.

[30] Zhou, Z., Li, F., Zhu, H. et al. “An improved genetic algorithm using greedy strategy toward task scheduling optimization in cloud environments,” Neural Comput & Applic 32, 1531–1541 (2020). https://doi.org/10.1007/s00521-019-04119-7.

[31] Madhusudhan H S, Satish Kumar T, S.M.F D Syed Mustapha, Punit Gupta, Rajan Prasad Tripathi, "Hybrid Approach for Resource Allocation in Cloud Infrastructure Using Random Forest and Genetic Algorithm", Scientific Programming, vol. 2021, Article ID 4924708, 10 pages, 2021. https://doi.org/10.1155/2021/4924708.

[32] Miriam, A.J., Saminathan, R. & Chakaravarthi, S. Non-dominated Sorting Genetic Algorithm (NSGA-III) for effective resource allocation in the cloud. Evol. Intel. 14, 759–765 (2021). https://doi.org/10.1007/s12065-020-00436-2.

[33] Bloch, T., Sridaran, R., Prashanth, C. S. R. (2018) Understanding Live Migration Techniques Intended for Resource Interference Minimization in Virtualized Cloud Environment. Big Data Analytics. 487-497.

[34] Introduction to Genetic algorithm http://www.obitko.com/tutorials/genetic-algorithms/crossover-mutation.php.

[35] Chen Ming; Li Mengkun; Cai Fuqin; "A Model of Scheduling Optimizing for Cloud Computing Resource Services Based on Buffer-pool Agent," Granular Computing (GrC), 2010IEEE International Conference, Aug. 2010.

[36] Rajkumar Buyya, “explaining Dynamic Resource Allocation for Efficient Parallel Data Processing in the Cloud,“ http://dsl.cs.uchicago.edu/TPDS_MTC/papers/TPDSSI-2010-01-0012.pdf.

[37] Calheiros, R.N., Ranjan, R., Beloglazov, A., De Rose, C.A., Buyya, R.. “Cloudsim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms,” Software: Practice and Experience 2011;41(1):23–50.

[38] Neha Garg, Damanpreet Singh, Major Singh Goraya, “VM Selection and Allocation Policy to Optimize VM Migration in Cloud Environment,” International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-2, July 2019.

[39] Michalewicz Z, Schoenauer M (1996) Evolutionary algorithms for constrained parameter optimization problems. Evol Comput 4(1):1–32.

[40] Katoch, Sourabh, Sumit Singh Chauhan, and Vijay Kumar. "A review on genetic algorithm: past, present, and future." Multimedia Tools and Applications 80.5 (2021): 8091-8126.

A Hybrid Deep Learning Approach for Freezing of Gait Prediction in Patients with Parkinson's Disease

Hadeer El-ziaat[1], Ramadan Moawad[3]
Department of Computer Science
Future University in Egypt (FUE)
Cairo, Egypt

Nashwa El-Bendary[2]
College of Computing and Information Technology
Arab Academy for Science, Technology and Maritime Transport (AASTMT), Aswan, Egypt

***Abstract*—The main objective of this work is to enhance the prediction of the Freezing of Gait (FoG) episodes for patients with Parkinson's Disease (PD). Thus, this paper proposes a hybrid deep learning approach that considers FoG prediction as an unsupervised multiclass classification problem with 3 classes: namely, normal walking, pre-FoG, and FoG events. The proposed hybrid approach Deep Conv-LSTM is based on the use of Convolutional Neural Network layers (CNN) and Long Short-Term Memory (LSTM) units with spectrogram images generated based on angular axes features instead of the normal principle-axes features as the model input. Experimental results showed that the proposed approach achieved an average accuracy of 94.55% for FoG episodes early detection using Daphnet and Opportunity publicly available benchmark datasets. Furthermore, the proposed approach achieved an accuracy of 93.5% for FoG events prediction using the Daphnet dataset with the subject independent mode. Thus, the significance of this study is to investigate and validate the impact of using hybrid deep learning method for improving FoG episodes prediction.**

Keywords*—Freezing of Gait (FoG); Parkinson's disease (PD); angular axes features; spectrogram; convolutional neural network (CNN); long short-term memory (LSTM)

I. Introduction

Parkinson's Disease (PD) is a disorder that affects the patient's nerves, which are recognized by low levels in the brain's dopamine. Also, PD is the second most common symptom after Alzheimer's disease [1]. Low levels of dopamine lead the patients to the inability to control their body motion or activity. PD can affect the patients by two kinds of symptoms, motor, and non-motor symptoms, the motor symptoms, or cardinal symptoms, which are related to the movement in general, including resting tremor, stepping slowness (bradykinesia), postural instability (issues in balance), and Freezing of Gait (FoG). Non-motor or dopamine-non-responsive including cognitive weakness, sleeping behavior problems, sense of smell loss, difficulty defecation, talking and swallowing problems, and other related side effects to the human's sensory [2].

FoG is a side effect and one of the main symptoms for patients with both advanced and early PD stages. It is one of the most debilitating motor symptoms in patients with PD as it may lead to falls and a loss of independence. FoG can occur in the arms (which affects the writing ability), face (which affects the patient's vision), and leg (which results during gait initiation by a series of short steps with tremors on lower limbs, turning, or walking towards a particular goal) [2].

FoG could be seen in different forms like 1) A complete FoG cycle 2) Freezing with knee-trembling 3) Walking with very short steps. For some patients, there is a brief trembling of the feet in place followed by short steps, while others experience total immobility and are unable to move it all for a few moments.

The main objective of this study is to predict FoG episodes based on a hybrid deep learning approach using time-series episodes windowing and relay on the use of angular axes feature to help in predicting FoG episodes with their different occurrence cases as mentioned previously. The baseline model has experimented with the proposed one (Conv-LSTM), a model based on the use of a deep convolutional neural network (CNN). This model is also implemented for detecting/predicting FoG episodes. The developed models in this paper have been tested with both Daphnet and Opportunity benchmark datasets.

Accordingly, the main contributions of this paper are summarized in the following points:

- Updating the publicly available Daphnet dataset through adding a new label (3) that specifies the pre-FoG episodes prediction.
- Developing a deep learning (CNN-LSTM) hybrid approach for handling the problem of FoG episodes prediction in patients with PD.
- Investigating subject-dependent spectrograms for predicting FoG episodes.
- Testing and validating the performance of the proposed approach through implementing multiple experiments using the benchmark datasets.

The remainder of this paper is organized as follows. Section II presents a review of the state-of-the-art related work employing different deep learning algorithms for predicting/detecting FoG episodes based on different approaches. Section III presents an overview of the main structure of the proposed approach. Results and discussion are presented in Section IV. Conclusions and future work are highlighted in Section V.

II. Related Work

Relevant state-of-the-art studies addressing FoG detection and prediction are described in this section.

The authors of [4] developed an approach to predict FoG for patients with PD using a deep learning model based on the Recurrent Neural Network (RNN) and Long Short-Term Memory networks (LSTMs). The achieved results were 94.7% with 1-second prediction, 82.9% with 3 seconds prediction, and 68.1% with 5 seconds prediction.

Also, the authors in [5] aimed to detect FoG based on the use of the DL approach, the data was collected from one wrist based on linear and angular acceleration. The collected data was fed into a CNN model with 10-fold cross-validation and Leave-one-out-subject-out (LOSO) Cross-Validation. The achieved results were 83% and 86% for sensitivity and specificity, respectively when using LOSO-CV. When using the 10-fold CV the achieved results were 88% and 90% for sensitivity and specificity, respectively.

For capturing long-range dependencies in variable-length input sequences, the authors in [6] used a Deep Recurrent Neural Network (DRNN) for building recognition models that can fulfill the paper's purpose. The proposed algorithms like cascaded, unidirectional, and bidirectional architectures are based on LSTM DRNNs. By using Unidirectional DRNN for the UCI dataset, the results were 96.7%, 96.8%, 96.7%, and 0.96. By using Unidirectional DRNN for the USC-HAD dataset, the results were 97.8%, 97.4.0%, 97.4%, and 0.97. By using Bidirectional DRNN for Opportunity dataset, the results were 92.5%, 86.7%, 83.5% and 0.92. Using the Cascaded DRNN for the Daphnet dataset, the results were 94.1%, 84.7%, 78.9%, and 0.93. Using the Cascaded DRNN for the Skoda dataset, the results were 92.6%, 93.0%, 92.6%, and 0.92. All datasets were measured with accuracy, average precision, average recall, and f1 score, respectively.

The author in [7] sought to detect FoG episodes from the data signal for the subject in/dependent with a 3D accelerometer. The proposed model implemented on patients in the daphnet dataset was based on the RNN LSTM model with three accelerometer sensors. The overall results were 80% and 79% for specificity and sensitivity, respectively when using all features (statistical and frequency domain features) with sensor and subject independent. When using only the statistical features the results were 89% and 34% for specificity and sensitivity, respectively with sensor and subject independent. The achieved results when using the frequency domain features were 80% and 77% for specificity and sensitivity, respectively with sensor and subject independent. They also implement a baseline model using the Random Forest classifier for comparing the achievements of each model. The results showed that the proposed RNN LSTM model outperforms the baseline model (RF classifier).

In [8], the authors developed a system for FoG detection that is based on CNN from a 2D acceleration signal. The average results for the subject independent were 80.7%, 69.29%, and 90.6% for accuracy, precision, and specificity, respectively. Moreover, the author in [9] proposed a hybrid deep learning algorithm for detecting FoG episodes. Their first model was based on the use of convolutional layers with LSTM units. The achieved results were 87.8% accuracy, 88.1% sensitivity, 89.1% specificity, and 88.4% geometric mean. The second model was based on convolutional layers with GRU (Gated Recurrent Unit). The results achieved were 85.4% accuracy, 94.8% sensitivity, 84.7% specificity, and 89.5% geometric mean.

The goal of the authors in [10] was to detect FoG using four types of feature sets with a DNN with 4-sec windowing. A model of CNN with MLP layers was introduced. It contains two convolutional layers with a max-pooling layer for feature extraction and with three fully connected layers for classification. The achieved sensitivity was 93.1% and 75% for specificity.

Furthermore, the authors in [19] used deep learning approaches and image processing techniques to detect FOG. The presented approach was based on the use of 1D-ConvNet. The results were 88.6% for sensitivity and 78% for specificity.

The aim of the study proposed in [20] was to detect FOG episodes in PD patients with a proposed model that consists of eight layered of 1D-ConvNet using two activation functions (sigmoid and hyperbolic). Down-sampling and low pass filter are used for data preprocessing. The achieved results were 89% for accuracy, 91.9% for sensitivity, and 89.5% for, specificity, respectively.

The purpose of the approach proposed in [21] was to detect FOG episodes using CNN based on two types of features, namely time and frequency domain features with 2.5 s windowing. The features were extracted from a tri-axil accelerometer and gyroscope sensors from a smartphone located in the patient's trouser pocket. Results were 91.8%, 93.8%, and 90.1% for F1-score, sensitivity, and specificity, respectively.

The main objective of this study is to investigate the impact of using deep feature learning based on angular axes spectrogram images and different windowing mechanisms through a hybrid deep learning approach for improving FoG episodes prediction.

III. Proposed Approach

The proposed approach section is divided into four stages, starting with dataset acquisition, data preparation, going throw deep learning features, and ending with the classification, as illustrated in Fig. 1.

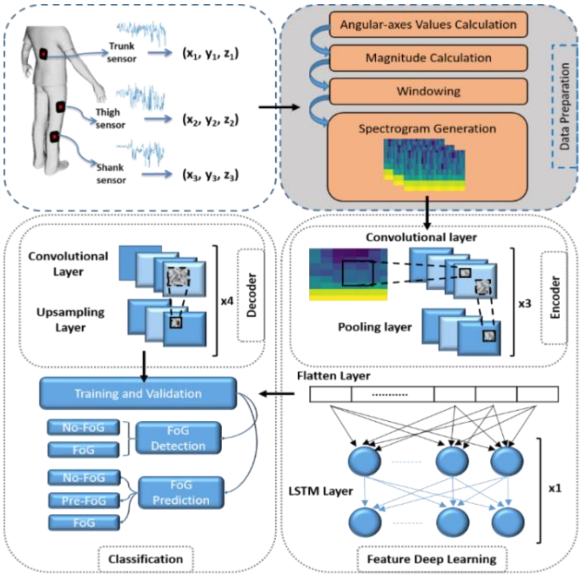


Fig. 1. Structure of the Proposed Hybrid FoG Prediction Model.

A. Dataset Acquisition

To train and validate the proposed models, two benchmark datasets are considered; namely the Daphnet and the Opportunity datasets of FoG time-series data.

1) The Daphnet dataset, which is available publicly from the UCI Machine Learning Repository [11], has been used for validating the proposed approach. The dataset was collected from a total of 10 PD volunteers patients, where 8 subjects have experienced FoG while performing several walking tasks in the lab. The Shank (ankle), the thigh (above the knee), and the trunk (lower back) of each subject were used to place a three-wearable tri-axial accelerometer for data recording [3]. From Fig. 3, which presents the same portions for ankle, knee, and trunk sensors of the walking, pre-FoG, and FoG signals it can be observed that the existence of the signal spacing when there are pre-FoG or FoG samples. The original data samples have been labeled as 0 for describing out-of-experiment events, 1 for describing no-FoG events, and 2 for describing FoG events. Sessions with a duration between 20 and 30 minutes are accomplished by the subjects in the Daphnet dataset to represent different characteristics of daily walking [11].

2) The opportunity dataset, which contains recordings that are collected from 12 subjects using 15 networked sensor systems with 72 sensors [12]. For each subject, 6 different runs were recorded 5 of them were the Activity of Daily Living (ADL) the remaining was a drill run. For this study, the activity recognition subset of the opportunity dataset was used. Only 4 subjects with ADL runs were used, which corresponds to recordings of three tri-axial accelerometer placements namely, hip, above the right knee (RKN^), and below the right knee (RKN_).

B. Data Preparation

Originally the validation Daphnet dataset doesn't contain samples with label 3 corresponding to pre-FoG events. For featuring the pre-FOG episodes with a new label as shown in Fig. 2, all samples with the same label (unified labels) for a specific window time before label 2 have been converted into label 3. Therefore, for the FoG prediction problem, a multi-class classification has been handled with the proposed approach. This method is used only for prediction implementations, but for the detection implementations, labels 1 and 2 only are used. Subjects four and ten have not experienced any Freezing of Gait episodes while performing several walking tasks. As previously mentioned in the data description section, the Daphnet dataset consists of ten PD subjects; the data of the two previous subjects and all records with zero labeling have been neglected. The data preparation phase has consisted of four steps, as follows:

1) Angular axes values calculation: The angular axes features [13], or axes of rotation, provide a combination of three rotations about different axes to represent the 3D orientation of an object. Because of the need to focus on the patient's rotation and orientation the reliance on the accelerometer features will not be efficient. Motions are dominated by rotations; therefore, it is needed to avoid the use of accelerometers and use a gyroscope sensor [14].

For the angular-axes calculation x, y, and z of the principle-axes from the three sensors have been used to calculate the angular axes features (Roll, Pitch, and Yaw). As shown in equations (1), (2), and (3) details of calculating the angular values, Roll (r), Pitch (p), and Yaw (y), about the x, y, and z axes, respectively, where $\Pi = 3.14$ is constant.

$$Roll = 180 * \arctan(y/\sqrt{x^2 + z^2}))/\pi \quad (1)$$

$$Pitch = 180 * \arctan(x/\sqrt{y^2 + z^2}))/\pi \quad (2)$$

$$Yaw = 180 * \arctan(z/\sqrt{x^2 + y^2}))/\pi \quad (3)$$

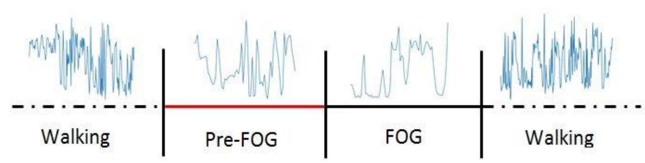


Fig. 2. An Example Signal Pattern for Accelerometer Signal Categories.

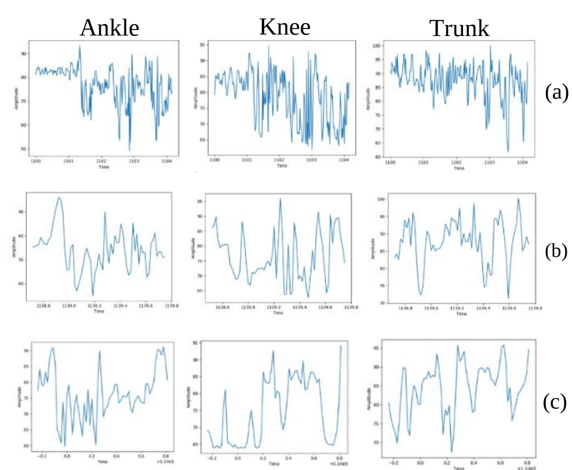


Fig. 3. Portion of Walking (a), pre-FoG (b) and FoG (c) Signals for Ankle, Knee and Trunk Sensor.

2) Magnitude calculation: After converting the principle-axes into angular axes, the magnitude of the obtained values has been calculated from each record of the three angular axes features values, according to equation (4), where s refers to the used ankle's (shank), knee's (lower thigh), or trunk's (lower back) sensor.

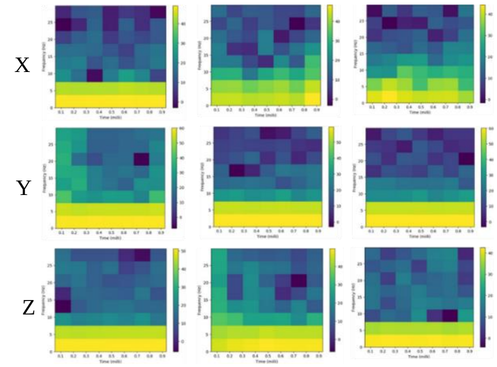
$$Magnitude = \sqrt{(r_s^2 + p_s^2 + y_s^2)} \quad (4)$$

3) Windowing: From each sensor, all the calculated magnitudes are divided into two overlapping windowing sizes, namely fixed and dynamic sizing based on the data labels. This paper presents two different schemes for selecting the best slicing technique and for testing various windowing mechanisms. Method 1 (scheme 1): The first method is adopting a 1 sec. window size, each window contains 76 samples, and each sample is 15 milliseconds [3]. Method 2 (scheme 2): On the other hand, a partially overlapping windowing method of dynamic-sized windows based on the data labels [3].

4) Spectrogram generation: Because other signals like nonstationary or non-periodic frequencies differ in time, in this step the information about the time domain and the frequency domain of real-life signals are needed, which is better computed from time-frequency analysis [15]. One of the basic visual tools for displaying the time-frequency analysis information is the spectrogram. The spectrogram is a 2D map, the vertical axis represents the frequency, and the horizontal axis represents the time of the signal. For spectrogram-based CNN, the amplitude provides a 2D array of successive segments used as an input feature for the spectrogram-based CNN. The spectrogram parameters are window size, noverlap, nfft (length of the FTT), and fs (sampling frequency). The input to the spectrogram is a signal data x and the output will be a matrix of 2D array q. The output q will have (nfft/2+1) rows if nfft is even and (nfft+1)/2 rows if nfft is odd. For columns, will be (the length of q – noverlap) / (length (window-noverlap)) [16].

Fig. 4 presents the 3D accelerometer axes spectrograms for FoG activity on the three sensors, namely the ankle, knee, and trunk. The time in milliseconds is presented on the x-axes and the frequency in Hz is presented on the y-axes, each row presents one of the axes, namely x, y, and z, respectively. The same portion of data is used for all sensors (ankle, knee, and trunk). It is visualized also from the figure areas with darker colors means that for a specific time and frequency point, the color of the image will be darker with a lower amplitude. Similarly, the color will be lighter with high magnitude.

It can be also observed that the most effective sensor placement is the trunk sensor, the readings from this sensor are better than the ankle and the knee sensors placements. It can be observed the lighter areas emitted from the trunk sensor, which means that those areas have a high frequency of FoG activity with high severity of FoG in the same area.



(a) (b) (c)

Fig. 4. Samples for each Axis (X, Y, and Z) for FoG Activity on the Ankle (a), Knee (b) and Trunk (c) Sensor.

C. Deep Learning Features

For the proposed approach (Conv-LSTM), with the use of three Convolutional layers, which receive and extract features from spectrogram images with intermediate three max-pooling layers (encoder) as shown in Fig. 1. The encoder in the network compresses or down-samples the input into a fewer number of bits. The space represented by these fewer numbers of bits is often called the latent space, at this point, the input has compressed to the maximum. A convolution layer tries to extract higher-level features by replacing data for each pixel with a value computed from the pixels. From Fig. 1 it can be observed how features are extracted from each Conv-LSTM layer, clarifying, and discussing the use of intermediate LSTM layers, which are used for the feature deep learning phase, and CNN layers that are used for classification.

It can be also observed how features are extracted from each Conv-LSTM layer after extracting features from spectrograms, by using an intermediate LSTM layer that is used to receive the output of the last convolutional layer. To do so, the flatten layer is used because convolutional layers output a shape of four dimensions whereas the LSTM layer needs a three-dimension input. Like the proposed approach, the baseline 2D CNN approach also used the encoder as a feature extractor with three convolutional layers and three max-pooling layers, but without using any LSTM layer.

CNN is a class of deep, feed-forward AI neural networks that are used in Image & video recognition/classification, Video to Text (seq. to seq.), and Image Question Answering. A CNN consists of an input and output layer as well as multiple hidden layers. Typically, a CNN's hidden layers consist of Convolutional layers, pooling layers, and fully connected layers. Typically, a CNN is used mostly for image and video recognition, classification, video to text, and image question answering applications. CNN learns useful features from data itself dispenses the use of the hand-crafted features and instead depends on learning features that the network automatically extracted. So, it is a combination of a feature extractor and a classifier. Mainly the CNN contains convolutional, pooling, fully connected layers, and soft-max [17], [18].

- Convolutional layer is responsible for extracting useful features automatically by learning and knowing the best features by selecting the highest weights of each neuron in each convolutional layer.
- Max-pooling layer is used to reduce the size of the feature map, which results in having a smaller number of parameters and computations.
- Up-sampling layer is a backward stride convolution and performed for end-to-end backpropagation learning from the pixel-wise loss
- Fully connected layer means that all the input neurons in each layer are connected to the previous layer.
- Soft-max layer acts like a classifier in which many probabilities were proposed from each class and the class with the highest probability was predicted.

LSTM layer, Fig. 5 magnifying the steps of how the LSTM network works. The first step is to decide what information will be thrown away from the cell state. This decision is made by the sigmoid layer (σ), which is called the "forget gate layer", it looks at xt (current input) and h(t-1) (output of the last LSTM unit) and outputs a number between 0 and 1 for each number in the cell state c(t-1) (memory of the last LSTM unit). 1 means "keep the previous" result and don't forget it while 0 means "don't keep the previous result" and forget it. For simplicity, if it is needed to predict the next word based on all the previous ones, the cell state might include the gender of the present subject so that the correct pronouns can be used. But when a new subject came up, the previous gender of the old subject needs to be forgotten.

In the next step, a decision as to what new information will be stored in the cell state is taken in two steps. First, a decision for which values will be updated is taken by a sigmoid layer called the "input gate layer". Next, a tanh layer creates a vector of new values ct (new updated memory). The next step is to combine the previous two outputs from each layer to create the updated cell state. For simplicity going back to the example, here adding the gender of the new subject to the cell state by replacing it with the old one that was forgotten is needed.

Finally, the ht (output) is based on the value stored in the cell state. First running a sigmoid layer that decides, which parts of the cell state will be going to be output is needed. Then, go through the tanh layer and multiply (x) the output of each layer.

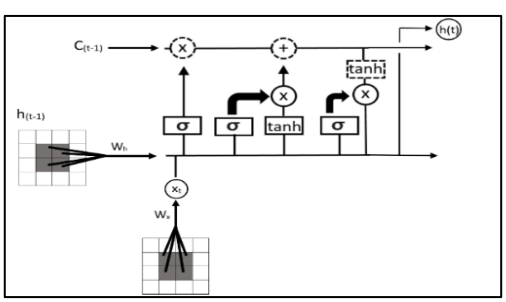


Fig. 5. Presents an Illustration of the Proposed Hybrid Model.

As the problem is based on unsupervised learning, the use of an Auto-Encoder mechanism is taking place as it has the option to not use dense layers, it can use the convolutional layers itself to learn, which is better for video, image, and series data. The encoder compresses or down-samples the input into a fewer number of bits, the space represented by these fewer bits is often called the latent space or bottleneck. At a maximum level, the input at a particular point is compressed that's why it can be also called the "maximum point of compression". The decoder is the reverse of the encoder, and it is importance lies in rebuilding the original image with the highest possible quality.

D. Classification

For the proposed model, both datasets are divided into sub-datasets, one for training and the other for validation using a k-fold cross-validation method applied for both training and validation. A hybrid deep learning structure is used as shown in Table I, the hybridization here is based on three convolutional layers with intermediate max-pooling layers, a flatten layer, three LSTM layers for feature deep learning, and four convolutional layers with intermediate up-sampling layers for classification. Convolutional layers as mentioned previously learn features automatically in each layer, so it doesn't need any hand-crafted features, and the reason is that the deep architecture, which includes multiple layers allows those layers to be stacked. So, this deep architecture can characterize the prominence of signals on different scales. In each convolutional layer, each neuron carries the maximum weight (the maximum weight means a better result) from the input layer and if a neuron was for example in the third layer it takes the output of the previous layer as its input. The convolutional layer's network has its output as a convolution fully connected neural network instead of matrix multiplication.

The structure of the baseline model (Deep 2D CNN) model was fed with spectrogram images as model input. The same as the proposed model, a baseline model consists of three convolutional layers with intermediate max-pooling layers for feature deep learning and four convolutional layers with intermediate up-sampling but without flatten and LSTM layers.

IV. Result and Discussion

This section discusses different models' results used in this paper to reach the best approach that effectively influences predicting FoG episodes. Also presents and discusses the experimental outcomes of using the proposed hybrid deep learning scheme for FoG episodes prediction. Both datasets are divided 80% for training with a random dividing and 20% for validation. The implementation consists of two phases using the Daphnet dataset:

- Working on the original dataset applying all sensors and each sensor apart (Ankle, Knee, and Trunk).
- Working on the original dataset after applying the proposed angular axes features for in/dependent subjects using all sensors and each sensor apart (Ankle, Knee, and Trunk).

The performance measurements as shown in equations (5), (6), (7), (8), and (9) are the accuracy, precision, recall, F-measure, and Specificity.

$$Accuracy = \frac{tp+tn}{tp+fn+fp+tn} \quad (5)$$

$$Precision = \frac{tp}{tp+fp} \quad (6)$$

$$Recall = \frac{tp}{tp+fn} \quad (7)$$

$$F - measure = 2\frac{precision*recall}{precision+recall} \quad (8)$$

$$\text{Specificity} = \frac{TN}{(TN+FP)} \quad (9)$$

A. Daphnet Dataset Results

In subject-independent FoG prediction, from Table II and Table III it can be observed that the results derived from sensor independent are almost the same as the results when using only data from the trunk sensor (sensor dependent). But for overall performance, using sensor independent is better than using sensor dependent. It was also observed that using a windowing with 1-second achieves a better performance than using a windowing with 4-seconds. Another enhancement is for using angular axes features over principle-axes. For the 2D CNN model, same as the hybrid Conv-LSTM model; the trunk sensor outperforms other sensors, and angular axes features achieve better performance. Using 1-second windowing records an enhancement against using 4-seconds windowing and better overall performance is achieved when using sensors independent over-dependent sensors. In addition, for the 2D CNN model, from Table IV and Table V it can be observed that using only CNN without LSTM the model achieves better results than using a hybrid structure of CNN with LSTM.

The results show that the implemented deep 2D CNN model achieves the highest accuracy against the proposed one, it can be observed in Fig. 7. It also emphasizes that the use of the angular axes features is much better than the principle-axes. For subject-independent and sensor dependent with 1-second (67 samples) windowing the two implemented models namely, 2D CNN and hybrid deep Conv-LSTM achieves an increase in the performance using angular axes against principle-axes by 2.7%, and 2.8% using ankle sensor for the two models respectively. And for the knee sensor the performance increased by 2.6%, and 2% for the two models, respectively. The enhancement in the trunk sensor is 3.3%, and 2.5% for the two models, respectively. When using sensor independent the angular axes features achieve an enhancement of 3.8%, and 2.2% for the two models, respectively, over the principle-axes.

Implementing subject independent with sensor dependent for FoG prediction with a 4-seconds (268 samples) windowing technique, the results as shown in Fig. 7 clarifies that using a bigger windowing technique negatively affects the performance. When using 1-second windowing against 4-seconds windowing with angular axes features from Fig. 6 and Fig. 7, the performance increases by 9% and 4.3% for the ankle sensor with 2D CNN, hybrid deep Conv-LSTM, respectively. On the knee sensor, the enhancement is 6.5% using the 2D CNN model and 4.4% using the hybrid Conv-LSTM model. The enhancement using trunk sensor is 4.8% and 3.7 for the two models, respectively. Finally, using sensor independent achieves increasing in the performance by 5.6% and 3.8% for the two models, respectively.

Applying the 4-seconds windowing technique using angular axes features outperforms using principle-axes as shown in Fig. 7. The enhancement on the ankle sensor is 2.3% and 5.6% for 2D CNN, and hybrid deep Conv-LSTM respectively. On the knee sensor, the enhancement increased by 3.1% and 3.6% for the two models, respectively. The performance using the trunk sensor increases by 2.6% and 2.8%. When applying sensor independent with angular axes features the performance against the use of principle-axes increased by 2.8% and 3.1% for the two models, respectively.

The implementation of Subject-independent FoG detection for deep learning models with sensor dependent from Table VI emphasized that the trunk sensor outperforms other sensors, and angular axes features achieve better performance against principle-axes. Also, using 1-second windowing achieves an enhancement over using the 4-second windowing. It has been observed from Table VI and Table VII that a better overall performance is achieved when using sensors independent over-dependent sensors. When comparing results of prediction with detection, it can be observed that FoG prediction outperforms FoG detection for both 2D CNN and hybrid deep Conv-LSTM models. In addition, from Table VIII and Table IX the results ensure that the 2D CNN model structure is better than the hybrid Conv-LSTM model structure.

Detecting FoG episodes for subject-independent and sensor dependent with 1-second (67 samples) windowing as shown in Fig. 8 when applying angular axes features outperforms using principle-axes. The enhancement on the ankle sensor is 2.6% and 3% for the 2D CNN and hybrid deep Conv-LSTM models, respectively. Furthermore, the enhancement on the knee sensor is 0.5% and 2.5% for the two models, respectively. The performance increased also when using the trunk sensor, the enhancement is 1% and 2.9%. An enhancement using angular axes features with sensor independent by 1.6% and 1.3%. Applying the 4-seconds windowing technique using angular axes features outperforms using principle-axes as shown in Fig. 9. The enhancement on the ankle sensor is 3.7% and 4.9% for 2D CNN, and hybrid deep Conv-LSTM, respectively. On the knee sensor, the enhancement increased by 3.3% and 3.5% for the two models, respectively. The performance using the trunk sensor increased by 4% and 3.2%. An enhancement is achieved when applying sensor independent with angular axes features by 2.9% and 3.4% for the two models, respectively against the use of principle-axes.

An enhancement can be observed in Fig. 8 and 9 when using angular axes features with 1-second, which equals 67 samples against 4-seconds (268 samples) windowing. The enhancement on the ankle sensor increased by 4.6% using the 2D CNN model and 4.3% using the hybrid deep Conv-LSTM model. On the knee sensor, the enhancement is 3.5% using the 2D CNN model and 4.9% using the hybrid deep Conv-LSTM model. The enhancement on the trunk sensor achieves 5% using the 2D CNN model and 3.2% using the hybrid deep Conv- LSTM model. The implementation of the sensor independent has also increased the performance by 4.7% using the 2D CNN model and 4.3% using the hybrid deep Conv-LSTM model.

TABLE I. HYBRID CONV-LSTM MODEL STRUCTURE

Layer	**Kernel size**	**Output map size**	**Activation**	**Optimizer**
Conv2D	3*3	28*28*16		
Max-pooling2D	2*2	14*14*16		
Conv2D	3*3	14*14*8		
Max-pooling2D	2*2	7*7*8		
Conv2D	3*3	7*7*8		
Max-pooling2D	2*2	4*4*8		
Flatten	-	128	tanh	
LSTM	-	1*16		adadelta
Conv2D	3*3	4*4*8		
Up-sampling2D	2*2	8*8*8		
Conv2D	3*3	8*8*8		
Up-sampling2D	2*2	16*16*8		
Conv2D	3*3	14*14*16		
Up-sampling2D	2*2	28*28*16		
Conv2D	3*3	28*28*3	Sigmoid	

TABLE II. HYBRID DEEP CONV-LSTM 1 SEC. AND 4 SEC. WINDOWING SENSOR DEPENDENT

1-second windowing				**4-second windowing**	
Sensor	**Subject**	***Accuracy***			
		Angular axes	***Principle axes***	***Angular axes***	***Principle axes***
Ankle	1	89.5%	86.7%	84.4%	78.1%
	2	91.7%	86.6%	79.1%	77.9%
	3	91.3%	90.5%	80.4%	80.4%
	5	91.9%	86.8%	82.4%	75.9%
	6	90.6%	85.7%	87.6%	78.6%
	7	90.8%	84.8%	81.3%	75.1%
	8	**92.3%**	78.9%	**91.1%**	**83.2%**
	9	92.8%	**88.8%**	74.6%	75.1%
Knee	1	90.1%	87.1%	78.9%	80.1%
	2	91.8%	86.6%	87.1%	77.5%
	3	91.1%	89.8%	85.3%	78.2%
	5	90.6%	86.7%	87.3%	79.4%
	6	91.9%	87.1%	85.2%	84.5%
	7	89.2%	89.6%	82.4%	77.7%
	8	**92.3%**	**91.3%**	**90.8%**	**87.4%**
	9	90.5%	89.7%	81.4%	83.1%
Trunk	1	91.7%	88.2%	81.5%	84.7%
	2	93.5%	88.1%	89.6%	81.4%
	3	89.1%	92.1%	82.2%	83.2%
	5	93.1%	88.8%	87.1%	84.6%
	6	94.6%	90.6%	92.5%	85.3%
	7	94.1%	90.9%	92.1%	85.1%
	8	**94.7%**	**92.6%**	**91.7%**	**89.8%**
	9	92.6%	91.3%	90.9%	82.9%

TABLE III. HYBRID DEEP CONV-LSTM 1 SEC. AND 4 SEC. WINDOWING USING SENSOR INDEPENDENT

1-second windowing			**1-second windowing**		
Sensor	**Subject**	***Accuracy***			
		Angular axes	***Principle axes***	***Angular axes***	***Principle axes***
All sensors	1	90.9%	89.1%	88.1%	84.1%
	2	93.5%	89.1%	89.8%	83.9%
	3	93.5%	93.1%	85.9%	82.9%
	5	91.3%	91.8%	86.7%	83.9%
	6	94.1%	92.2%	89.1%	87.9%
	7	92.2%	92.2%	84.9%	81.5%
	8	**95.2%**	**92.3%**	**93.1%**	**89.2%**
	9	94.1%	91.5%	89.2%	80.2%

TABLE IV. 2D CNN 1 SEC. AND 4 SEC. WINDOWING USING EACH SENSOR DEPENDENT

1-second windowing			**4-second windowing**		
Sensor	**Subject**	***Accuracy***			
		Angular axes	***Principle axes***	***Angular axes***	***Principle axes***
Ankle	1	91.3%	86.1%	79.3%	79.2%
	2	91.1%	89.3%	78.5%	74.7%
	3	90.5%	88.5%	79.4%	71.6%
	5	93.1%	90.2%	78.6%	76.4%
	6	86.6%	89.1%	79.8%	70.9%
	7	92.1%	88.6%	86.1%	77.3%
	8	91.5%	**90.5%**	**83.3%**	**75.4%**
	9	**91.6%**	90.1%	70.4%	63.8%
Knee	1	88.6%	87.5%	**84.1%**	57.5%
	2	92.1%	83.9%	60.1%	74.3%
	3	91.8%	90.6%	82.1%	75.1%
	5	82.7%	87.3%	80.8%	78.7%
	6	89.1%	84.2%	76.8%	50.2%
	7	90.4%	88.6%	81.3%	51.2%
	8	**92.8%**	**90.5%**	82.1%	**77.9%**
	9	91.8%	89.4%	74.6%	77.9%
Trunk	1	90.8%	84.3%	83.4%	76.1%
	2	89.1%	90.9%	81.9%	69.8%
	3	92.8%	89.5%	82.1%	78.4%
	5	92.4%	83.7%	80.3%	74.5%
	6	91.9%	86.5%	80.9%	70.9%
	7	93.3%	90.9%	81.3%	86.1%
	8	92.5%	86.5%	**87.1%**	**87.2%**
	9	**96.1%**	**93.1%**	79.9%	84.3%

TABLE V. 2D CNN 1 SEC. AND 4 SEC. WINDOWING USING SENSOR INDEPENDENT

1-second windowing				4-second windowing	
Sensor	**Subject**	***Accuracy***			
		Angular axes	***Principle axes***	***Angular axes***	***Principle axes***
All sensors	1	92.6%	92.1%	87.7%	81.7%
	2	**95.1%**	92.5%	**89.8%**	80.9%
	3	93.5%	92.9%	80.1%	79.2%
	5	92.1%	93.1%	75.7%	79.8%
	6	93.6%	93.1%	79.8%	80.2%
	7	89.2%	**93.6%**	81.6%	82.1%
	8	94.6%	91.8%	87.8%	**82.6%**
	9	94.5%	92.4%	81.1%	82.4%

TABLE VI. HYBRID DEEP CONV-LSTM 1 SEC. AND 4 SEC. WINDOWING USING SENSOR DEPENDENT

1-second windowing				4-second windowing	
Sensor	**Subject**	***Accuracy***			
		Angular axes	***Principle axes***	***Angular axes***	***Principle axes***
Ankle	1	89.4%	85.3%	80.8%	74.1%
	2	90.3%	85.6%	85.1%	79.4%
	3	89.1%	84.5%	85.9%	81.2%
	5	91.9%	89.1%	85.5%	77.2%
	6	91.9%	85.7%	85.2%	77.2%
	7	88.8%	84.4%	79.1%	76.2%
	8	**94.3%**	**92.6%**	**88.2%**	**84.8%**
	9	91.7%	90.5%	84.1%	78.4%
Knee	1	89.5%	86.6%	81.4%	75.4%
	2	92.1%	86.1%	84.1%	75.5%
	3	92.1%	87.9%	83.3%	83.3%
	5	90.2%	89.1%	84.8%	80.7%
	6	92.6%	87.6%	77.6%	85.6%
	7	91.2%	89.5%	84.7%	78.4%
	8	**91.4%**	**90.3%**	**88.1%**	**83.1%**
	9	91.3%	89.3%	83.4%	79.3%
Trunk	1	91.2%	88.3%	82.8%	83.6%
	2	93.9%	88.7%	90.1%	86.7%
	3	94.2%	91.9%	86.4%	86.7%
	5	92.2%	90.1%	85.1%	82.7%
	6	93.1%	92.2%	87.9%	89.7%
	7	93.6%	86.4%	86.7%	87.2%
	8	**94.9%**	**92.9%**	**92.9%**	**91.3%**
	9	94.2%	91.4%	85.8%	84.5%

TABLE VII. HYBRID DEEP CONV-LSTM 1 SEC. AND 4 SEC. WINDOWING USING SENSOR INDEPENDENT

1-second windowing				4-second windowing	
Sensor	**Subject**	***Accuracy***			
		Angular axes	***Principle axes***	***Angular axes***	***Principle axes***
All sensors	1	92.1%	86.5%	86.2%	79.9%
	2	92.8%	86.9%	91.1%	83.1%
	3	93.1%	90.1%	88.8%	82.7%
	5	93.2%	90.5%	88.4%	83.6%
	6	93.9%	90.5%	89.2%	85.1%
	7	93.2%	89.9%	84.2%	82.3%
	8	**95.2%**	**91.9%**	**90.2%**	**90.5%**
	9	94.1%	91.8%	90.1%	85.7%

TABLE VIII. CNN WITH 1 SEC. AND 4 SEC. WINDOWING USING SENSOR DEPENDENT

1-second windowing				4-second windowing	
Sensor	**Subject**	***Accuracy***			
		Angular axes	***Principle axes***	***Angular axes***	***Principle axes***
Ankle	1	89.2%	85.6%	76.9%	73.2%
	2	83.4%	80.3%	73.6%	70.5%
	3	92.8%	89.6%	73.1%	72.4%
	5	91.6%	87.8%	81.5%	79.1%
	6	89.6%	83.4%	57.6%	50.1%
	7	91.5%	87.5%	83.1%	72.8%
	8	**92.7%**	**89.7%**	81.1%	74.8%
	9	86.7%	83.8%	**82.1%**	**75.2%**
Knee	1	88.3%	83.5%	79.8%	72.1%
	2	87.1%	84.9%	82.1%	75.8%
	3	91.1%	88.7%	82.4%	77.1%
	5	91.5%	89.1%	80.8%	76.3%
	6	89.1%	85.1%	80.2%	70.5%
	7	89.2%	85.4%	80.1%	77.8%
	8	**91.1%**	**86.3%**	**83.5%**	**80.4%**
	9	91.6%	88.5%	83.4%	80.2%
Trunk	1	89.4%	87.1%	86.1%	75.1%
	2	87.8%	86.4%	78.7%	72.9%
	3	90.2%	87.7%	87.6%	65.4%
	5	92.8%	88.6%	89.5%	67.5%
	6	93.6%	91.4%	84.5%	66.1%
	7	92.6%	88.1%	78.1%	72.3%
	8	**95.6%**	**91.6%**	78.5%	75.4%
	9	93.5%	91.1%	**78.6%**	**76.2%**

TABLE IX. CNN 1 SEC. AND 4 SEC. WINDOWING USING SENSOR INDEPENDENT

1-second windowing				**4-second windowing**	
Sensor	**Subject**	***Accuracy***			
		Angular axes	***Principle axes***	***Angular axes***	***Principle axes***
All sensors	1	92.1%	88.2%	88.4%	79.8%
	2	94.9%	90.5%	83.4%	80.5%
	3	94.8%	90.2%	65.8%	83.6%
	5	92.1%	85.6%	87.1%	82.8%
	6	**94.9%**	90.8%	**88.6%**	80.1%
	7	85.9%	80.9%	88.2%	78.9%
	8	93.5%	**91.1%**	87.2%	**87.1%**
	9	93.3%	88.9%	85.8%	83.7%

B. Opportunity Dataset Results

Results for sensor dependent, Table X shows that using angular axes features outperforms using principle-axes, as the Conv-LSTM model achieved an enhancement of 4.3%, 2.6%, and 2.4% for upper knee, hip, and lower knee sensors, respectively. Sensor-dependent results for 2D CNN, Table XI ensure also that the use of angular axes features results are better than principal axes. The results increased by 3.1%, 1.6%, and 1.2% for upper knee, hip, and lower knee sensors, respectively. For the independent sensor using angular axes features outperform the results from principle-axes with an enhancement by 3.5% and 4.1% for Conv-LSTM and 2D CNN models, respectively as shown in Table XII. From Table X and Table XI, it can be observed that the hip sensor outperforms other sensors and the performance of the 2D CNN model achieves results better than the Conv-LSTM model. Table XIII presents different deep learning models and features for the related work compared with the proposed models. Two different models are proposed for deep learning implementation, the first one is a combination of a convolutional neural network layer for 2D spectrograms images and LSTM recurrent units (Conv-LSTM). The second model used CNN also with 2D spectrogram images. Both models are implemented under the use of angular axes features instead of principle-axes. The best F-measure achieved by the proposed related work algorithms using the same Daphnet dataset was 80% on the other hand, the Conv-LSTM model achieves 95.1% F-measure and the CNN model achieves 98.1% F-measure. The enhancement here is 15.1% and 18.1% for Conv-LSTM and CNN models, respectively.

TABLE X. HYBRID DEEP CONV-LSTM USING SENSOR DEPENDENT

Sensor	**Accuracy**	
	Angular axes	***Principle axes***
Upper Knee	95.4%	91.1%
Hip	95.3%	92.7%
Lower Knee	95.5%	93.1%

TABLE XI. 2D CNN USING SENSOR DEPENDENT

Sensor	**Accuracy**	
	Angular axes	***Principle axes***
Upper Knee	96.3%	93.2%
Hip	96.2%	94.6%
Lower Knee	94.7%	93.5%

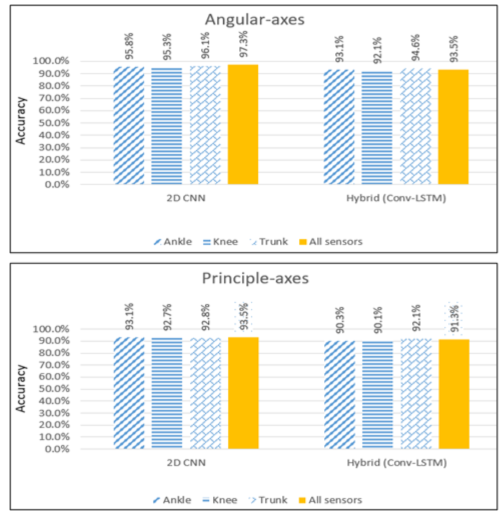


Fig. 6. Accuracy for FoG Prediction using Angular and Original Axes with 1-Second Windowing (67 Samples).

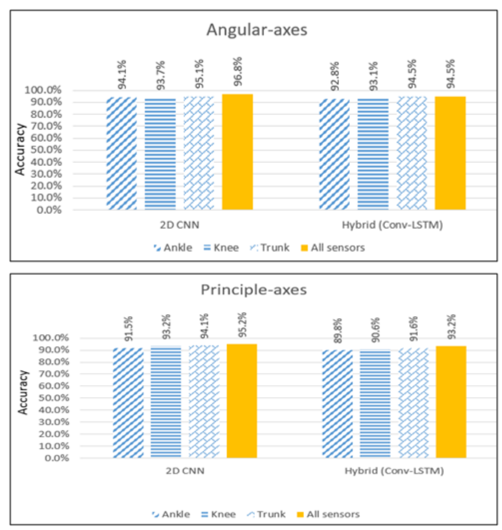


Fig. 7. Accuracy for FoG Prediction using Angular and Original Axes with 4-Seconds Windowing (268 Samples).

TABLE XII. HYBRID DEEP CONV-LSTM AND 2D CNN USING SENSOR INDEPENDENT

Model	**Accuracy**		**Pre.**	**Recall**	**F-measure**
	Angular axes	***Principle axes***	***Angular axes***		
Conv-LSTM	94.6%	91.1%	94.5%	93.1%	93.8%
CNN	97.6%	93.3%	96.5%	95.5%	95.1%

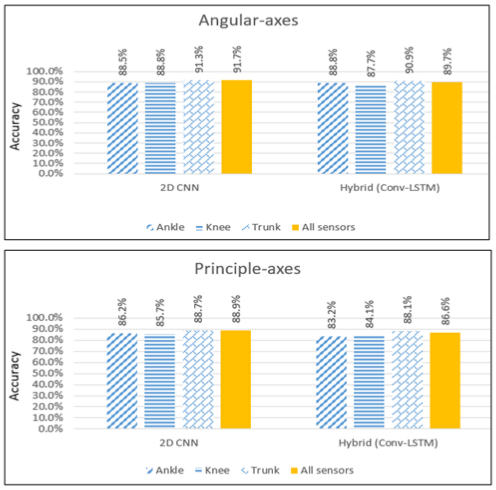


Fig. 8. Accuracy for FoG Detection using Angular and Original Axes with 1-Second Windowing (67 Samples).

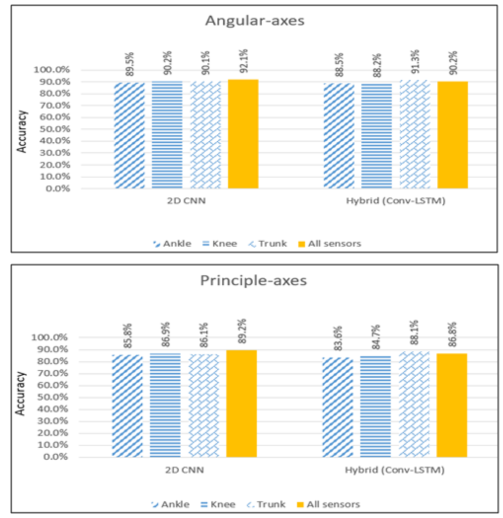


Fig. 9. Accuracy for FoG Detection using Angular and Original Axes with 4-Seconds Windowing (268 Samples).

TABLE XIII. DEEP LEARNING PERFORMANCE MEASUREMENTS COMPARED TO DIFFERENT RELATED WORK MEASUREMENTS

Ref.	Classifier	Accuracy	Precision	Recall	F-measure	Specificity
Murad & Pyun [6] 2017	unidirectional, bidirectional, and cascaded architectures based on long short-term memory (LSTM) DRNNs	94.1%	84.7%	78.9%	93%	-
Masiala [7] 2017	RNN with LSTM units with statistical features	-	-	79%	-	80%
Camps [9] 2017	Hybrid model based on convolutional layers with LSTM units	87.8%	-	88.1%	-	89.1%
Camps et al. [19] 2017	Four and five layers of 1D-ConvNet	-	-	88.6%	-	78%
Xia et al. [8] 2018	CNN	80.7%	69.29%	-	-	90.6%
Segundo et al. [10] 2018	CNN + MLP	-	-	93.1%	-	75%
Camps et al. [20] 2018	Eight 1D-ConvNet layers	89%	-	91.9%	-	89.5%
Kim et al. [21] 2018	CNN	-	-	93.8%	91.8%	90.1%
Yuan & Chakraborty [4] 2020	RNN with LSTMs and three different schemes.	94.7%	-	-	-	-
Bikias et al [5] 2021	CNN with two schemes (10-fold CV and LOSO-CV)	-	83% LOSO CV and 86% 10-CV	-	-	88% LOSO CV and 90% 10-CV
The proposed model (prediction)	**Conv-LSTM**	**93.5%**	**95.1%**	**93.5%**	**94.3%**	**95.1%**
The proposed model (detection)		**94.5%**	**94.6%**	**92.1**	**93.3%**	**94.1%**
Baseline model (prediction)	**2D CNN**	**97.3%**	**96.1%**	**97.1%**	**96.6%**	**96.5%**
Baseline model (detection)		**96.8%**	**96.3%**	**95.4%**	**95.8%**	**96.1%**

V. CONCLUSION AND FUTURE WORK

In this paper, a hybrid model of 2D convolutional layers with LSTM layers (Conv-LSTM) was proposed with the use of spectrograms as an input for the model and another baseline model was 2D CNN. The developed models are implemented using angular axes features and principle-axes sensor data readings. In this paper, two windowing methodologies are tested and applied for the Daphnet dataset. The two methodologies are a window of 67 samples, which is equivalent to 1-second, and a window of 268 samples, which is equivalent to 4 seconds. The paper aimed to predict the FoG episodes, the most approach that achieves this aim was the use of the CNN approach for both Daphnet and Opportunity datasets as shown in Fig. 6 and Table XII, followed by adopting the hybrid Conv-LSTM approach. Several observed enhancements are achieved. First, the enhancement was achieved when using angular axes features over principle-axes. Second, using 1-second windowing against 4 seconds achieves better performance. Deep learning algorithms for predicting pre-FoG episodes, results from Table II and Table III for the Conv-LSTM model, Table IV and Table V for the CNN model ensure on using angular axes features outperforms the use of principle-axes. Deep learning results for detecting FoG episodes from Tables VI and VIII for the Conv-LSTM model and Tables VIII and IX for the CNN model also clarify that the use of the angular axes features is better than the principle-axes. In addition, the prediction results for both Conv-LSTM and 2D CNN models are better than the detection results.

Deep learning algorithms are implemented with the use of an early stopping approach to avoid the model's overfitting. All training sets are implemented for 15 epochs and 15 folds in each epoch for both datasets. It also emphasizes that the maximum number of epochs the model reached was 14 epochs, the minimum number was 2 epochs and the average number of epochs lasting in each fold was 6 epochs. The best result achieved from deep learning models was 97.6% and 93.5% for 2D CNN and hybrid Conv-LSTM respectively for the subject and sensor independent. It can be also observed that patient 8 was the subject with almost the best accuracy in subject dependent, as the patient recorded a score of 4 out of 5 on the Hoehn and Yahr (H&Y) scale, which is considered the highest score among the other 10 patients in the Daphnet dataset.

For future work, various challenges could be considered in the domain of predicting the FoG episodes via pre-fog behavior detection as well as predicting FoG severity. Working on new feature fusion sets. Build our dataset to experiment with different algorithms and features and to be compared with other datasets.

References

[1] S. García, E. S. Castro, I. Expósito, T. De Deus, C. Tuñas, A. Aneiros, M. L. Fernández, D. N. Arias, and M. B. Torres, "Comorbid conditions associated with Parkinson's disease: a longitudinal and comparative study with Alzheimer disease and control subjects," Journal of the neurological sciences, vol. 373, pp. 210-215, 2017.

[2] A. H. V. Schapira, K. R. Chaudhuri, and P. Jenner, "Non-motor features of Parkinson disease," Nature Reviews Neuroscience, vol. 18, no. 7, pp. 435-450, 2017.

[3] H. Elziaat, N. El-Bendary, and R. Moawad, "Multi-Feature Fusion and Machine Learning: A Model for Early Detection of Freezing of Gait Events in Patients With Parkinson's Disease," Handbook of Research on Automated Feature Engineering and Advanced Applications in Data Science, edited by Mrutyunjaya Panda and Harekrishna Misra, IGI Global, pp. 95-118, 2021.

[4] A. M. Yuan and S. Chakraborty, "A Study of Deep Learning for Predicting Freeze of Gait in Patients with Parkinson's Disease," IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 1324-1331, 2020.

[5] T. Bikias, D. Iakovakis, S. Hadjidimitriou, V. Charisis, and L. J. Hadjileontiadis, "DeepFoG: An IMU-Based Detection of Freezing of Gait Episodes in Parkinson's Disease Patients via Deep Learning," Frontiers in robotics and AI, vol. 8, 2021.

[6] A. Murad and J. Y. Pyun, "Deep Recurrent Neural Networks for Human Activity Recognition," Sensors, vol. 17, no. 11, pp. 2556, November 2017.

[7] S. Masiala, "Detection of Freezing of Gait in Patients with Parkinson's Disease Using Deep Recurrent Neural Networks," PhD diss., Tilburg University, 2017.

[8] Y. Xia, J. Zhang, Q. Ye, N. Cheng, Y. Lu, and D. Zhang, "Evaluation of deep convolutional neural networks for detection of freezing of gait in Parkinson's disease patients," Biomedical Signal Processing and Control, vol. 46, pp. 221-230, 2018.

[9] J. Camps, "Applying deep-learning techniques to detect freezing of gait episodes in Parkinson's disease patients," Master's thesis, Universitat Politècnica de Catalunya, 2017.

[10] R. S. Segundo, H. N. Hellín, R. Torres, J. Hodgins, and F. D. L. Torre, "Increasing Robustness in the Detection of Freezing of Gait in Parkinson's Disease," Electronics, vol. 8, no. 2, pp. 119, 2019.

[11] M. Ba¨chlin, M. Plotnik, D. Roggen, I. Maidan, J. M. Hausdorff, N. Giladi, and G. T ro¨ster, "Wearable assistant for Parkinson's disease patients with the freezing of gait symptom," IEEE Transactions on Information Technology in Biomedicine, vol. 14, no. 2, pp. 436-446, 2009.

[12] R. Chavarriaga, H. Sagha, A. Calatroni, S. Digumarti, G. Tröster, J. D. R. Millán, and D. Roggen, "The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition", Pattern Recognition Letters, vol. 34, no. 15, pp. 2033-2042, 2013.

[13] H. A. Hashim, "Special orthogonal group SO (3), euler angles, angle-axis, rodriguez vector and unit-quaternion: Overview, mapping and challenges," arXiv preprint arXiv:1909.06669, 2019.

[14] I. A. Faisal, T. W. Purboyo, and A. S. Raharjo Ansori, "A Review of accelerometer sensor and gyroscope sensor in IMU sensors on motion capture," J. Eng. Appl. Sci, vol. 15, no. 3, pp. 826-829, 2019.

[15] M. Njirjak, E. Otović, D. Jozinović, J. Lerga, G. Mauša, A. Michelini, and I. Štajduhar, "The Choice of Time–Frequency Representations of Non-Stationary Signals Affects Machine Learning Model Accuracy: A Case Study on Earthquake Detection from LEN-DB Data," Mathematics, vol. 10, no. 6, pp. 965, 2022.

[16] L.Wang, C.Wang, and Y. A. Chen, "Fast three-dimensional display method for time-frequency spectrogram used in embedded fault diagnosis devices," Applied Sciences, vol. 8, no. 10, pp. 1930, 2018.

[17] M. Panwar, S. R. Dyuthi, K. C. Prakash, D. Biswas, A. Acharyya, K. Maharatna, A. Gautam, and G. R. Naik, "Cnn based approach for activity recognition using a wrist-worn accelerometer," In Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE, pp. 2438-2441, 2017.

[18] S. H. Lee, C. S. Chan, S. J. Mayo, and P. Remagnino, "How deep learning extracts and learns leaf features for plant classification," Pattern Recognition, vol. 71, pp. 1-13, 2017.

[19] J. Camps, A. Samà, M. Martín, D. Rodríguez-Martín, C. Pérez-López, S. Alcaine, B. Mestre, A. Prats, M. C. Crespo, J. Cabestany, À. Bayés, and A. Català, "Deep Learning for Detecting Freezing of Gait Episodes in Parkinson's Disease Based on Accelerometers," In International Work-Conference on Artificial Neural Networks, pp. 344-355, 2017.

[20] J. Camps, A. Sama, M. Martin, D. Rodriguez-Martin, C. Perez-Lopez, J. M. M. Arostegui, J. Cabestany, A. Català, S. Alcaine, B. Mestre, A. Prats, M. C. Crespo-Maraver, T. J. Counihan, P. Browne, L. R. Quinlan, G. Ó Laighin, D. Sweeney, H. Lewy, G. Vainstein, A. Costa, and R. Annicchiarico, "Deep learning for freezing of gait detection in Parkinson's disease patients in their homes using a waist-worn inertial measurement unit," Knowledge-Based Systems, vol. 139, pp. 119-131, 2018.

[21] H. B. Kim, H. J. Lee, W. W. Lee, S. K. Kim, H. S. Jeon, H. Y. Park, C. W. Shin, W. J. Yi, B. Jeon, and K. S. Park, "Validation of Freezing-of-Gait Monitoring Using Smartphone," Telemedicine and e-Health, vol. 24, no. 11, pp. 899-907, 2018.

Emotions Classification from Speech with Deep Learning

Andry Chowanda
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480

Yohan Muliono
Cyber Security Program
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480

***Abstract*—Emotions are the essential parts that convey meaning to the interlocutors during social interactions. Hence, recognising emotions is paramount in building a good and natural affective system that can naturally interact with the human interlocutors. However, recognising emotions from social interactions require temporal information in order to classify the emotions correctly. This research aims to propose an architecture that extracts temporal information using the Temporal model of Convolutional Neural Network (CNN) and combined with the Long Short Term Memory (LSTM) architecture from the Speech modality. Several combinations and settings of the architectures were explored and presented in the paper. The results show that the best classifier achieved by the model trained with four layers of CNN combined with one layer of Bidirectional LSTM. Furthermore, the model was trained with an augmented training dataset with seven times more data than the original training dataset. The best model resulted in 94.25%, 57.07%, 0.2577 and 1.1678 for training accuracy, validation accuracy, training loss and validation loss, respectively. Moreover, Neutral (Calm) and Happy are the easiest classes to be recognised, while Angry is the hardest to be classified.**

***Keywords*—Emotions recognition; speech modality; temporal information; affective system**

I. Introduction

Emotions are one of the essential communication factor during the social interactions. They provide additional meanings to verbal communication. Most of the conversation meaning can be captured mostly via non-verbal channels (e.g. speech prosody, body gestures and facial expressions) [1], [2], [3]. Hence, capturing emotions during social interactions between interlocutors is essential to building a system that can interact with humans effectively, efficiently, and naturally. Several efforts have been made to build models that can automatically classify emotions from non-verbal cues in the conversation. Some researchers aim to model the emotions classifier from image or video modality (e.g. Facial Expression Recognition and Hand and Body Gesture). The others use speech and text modality to recognise emotions from the conversation. Generally, the emotions are classified into six basic emotions plus neutral [4]. Recognising emotions from the conversation is a cumbersome task to a social ignorant computer [3]. Several problems exist in building good emotions classifier model from social conversation. First is the dataset; most datasets exist to model the emotions recognition are not balanced in the emotions class. This is due to not all emotions being expressed equally. The second problem is that not all the emotions recognition models have good performance to recognise emotions from the conversation. The results depend on the implemented machine or deep learning algorithms, the dataset used, pre-processing applied, and the modality used (video, image, text or speech). This research proposes and explores several deep learning architectures based on Temporal Convolutional Neural Networks (CNN) and Long Short Term Memory (LSTM) to extract features and classify emotions from speech. Most of the emotions recognition required temporal information to improve the model performance. Hence, this research proposes a combination of Temporal CNN to extract the features from the speech signals with LSTM to extract the features further and classify the emotions. The results have shown that MODEL-5 achieved the best model with the training accuracy score of 99.92%, validation accuracy of 78.22%, training loss of 0.0144 and validation loss of 0.8432. The rest of the sections in this paper are organised as follows: The next section illustrates the related work and state of art of emotions recognition from speech. The next section, Emotions Recognition from Speech, demonstrates the proposed framework to model emotions recognition from speech signals. The details of the experiment's settings are also shown in this section. The results are comprehensively presented and discussed in section Results and Discussion. Finally, the last section demonstrates the conclusion and future research direction of this research.

II. Related Work

A. Emotion Detection / Recognition

Emotions are one of the essential parts of social interactions. Emotions convey more than 80% meanings during the social interactions between interlocutors [2], [1]. Hence, detecting or recognising emotions is a paramount task to build a good and natural affective system. Emotion Detection / Recognition is a classification method that can bring up an important feature, namely the emotion contained in an input used for various uses [5]. The input used consists of various forms such as: speech [6], text [7] and visual [8], [9] cues. Most emotions detection/recognition tasks implement machine or deep learning (e.g. convolutional based, attention-based, recurrent based and transformer-based) to model the detector or recogniser. Analysing emotions can help in various fields, one of which is human and computer interaction which can later make computers better decisions for their users. some research regarding emotion detection has many variations, such as: Sarcasm Detection [10], Mood Prediction [11] and Personality Detection [12].

B. Speech Emotion Recognition

Speech Emotion Recognition (SER) is a method for mapping the features of a speech into the emotions contained in the speech. SER is not a new field of study [13]. However, along with the development of technologies, several methodological developments can be applied to SER. Thus, making research in the SER field more varied and complex to achieve more optimal results. SER usually utilises a classification algorithm to map input in a speech to output in the form of emotion classification. In general, the pipeline for SER is data pre-processing, features extraction and model training + evaluation. The data pre-processing generally involves data augmentation as well as data framing and windowing. Features extraction techniques are implemented to the data after the data is being pre-processed. The features can be extracted in the form of Spectral features, Prosodic features and the combination of both Spectral and Prosodic features. Finally, the features are then trained and evaluated using machine (or deep) learning algorithms.

C. Convolutional Neural Network in Speech Emotion Recognition

Although CNN is well-designed for Image Recognition it could be extended to Natural Language Processing and Speech Processing [14][15][16] Research regarding CNN for Speech Emotion Recognition conducted in 2016 [17] and 2018 [18] using RECOLA datasets [19]. by combining Convolutional Neural Network and Long Short-Term Memory which resulted in an outperformed model compared to traditional approaches on signal processing techniques. Then in 2017 [6] conducted research regarding Speech Emotion Recognition using Deep Convolutional Neural Network (DCNN) and Discriminant. Temporal Pyramid Matching (DTPM) to classify speaker's emotion resulting in a good model for automatic feature learning on speech emotion recognition tasks. The research concludes that DCNN is not only effective for image recognition but also in Speech Emotion Recognition. In 2020 [20] conducted research about CNN based Framework for Enhancing Audio Signal Processing for Speech Emotion Recognition proposing a framework that utilises a discriminative CNN using spectrogram which according to the author, the spectrogram has many features that texts or phonemes cannot represent.

III. Emotions Recognition from Speech

This research proposes the five best architectures by combining Temporal Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM) to extract and classify emotions from speech signals. The dataset used in this research is The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) [21]. The dataset contains more than 7000 audio files of emotional speech acted by twenty-four professional actors. This research only uses the emotional song dataset, where there are 920 audio data files and four basic emotions (i.e. Angry, Fear, Happy and Sad) plus Neutral. The RAVDESS dataset encodes the audio information (e.g. label and gender) in the filename. Hence, some text sub-string methods were applied to extract the label of the files. In this research, gender information is not used. Moreover, several pre-processing techniques were implemented to the dataset to enhance the quality of the dataset. First, the sample rate of the dataset was set to 16 KHz, and to normalise the speech time, the signal audio was padded to a maximum of 3 seconds. The dataset was split into two sets of data train and test with the ratio of 80%:20% (736:184). Moreover, the training data then were augmented to improve the quality of the data. This research proposed two settings of the data augmentation: seven times of the training data (5,152) and three times of the training data (2,208). Table I illustrates the emotions class distribution on each augmentation setting. The column Train Aug 1 denotes the augmentation with three times of the training data, while column Train Aug 2 refers to the augmentation with seven times of the training data. The dataset has an imbalanced dataset, where the Sad class is the majority class, and the Angry class is the minority class. Fig. 1 illustrates the example of the speech signal. The X-axis indicates the time is second (s), while the Y-axis indicates the amplitude of the signals in Decibel (dB). The left side of the image illustrates the original speech signal and the right side of the image demonstrates the augmented speech signal with noise.

TABLE I. Dataset Distribution

No	Label	Train Aug 1	Train Aug 2	Test
1	SAD	465	1120	46
2	NEUTRAL	450	1036	39
3	HAPPY	444	1015	39
4	FEAR	441	1015	36
5	ANGRY	408	966	24
6	**TOTAL**	**2,208**	**5,152**	**184**

Feature extraction methods using Short-time Fourier Transform (STFT) and Mel Frequencies were applied to generate the Mel-Spectrogram representation on each audio file. The features were extracted using several parameters, such as: the hop length of 512, the window of 256. To normalise the features, all the vector then padded with zeros up to 2,048 to match the Fast Fourier Transform input. The next step was to generate Mel-Spectrogram from the Mel frequencies generated from Mel bins of 128 and the maximum frequency of 4.0 KHz. Finally, the features were framed with a window step of 128 and a window size of 64. Fig. 1 and Fig. 2 illustrate the examples of the Mel Spectrogram features visualisation from a happy female audio (left side of the image) and a happy male audio (right side of the image). The features extracted are then used in training with the proposed architectures. There are five best architectures proposed in this research.

Fig. 3 demonstrates the blueprint of the proposed architectures. The architecture blueprint consists of four parts: the CNN block, The Flatten layer, the LSTM block and the Softmax Layer. The Temporal CNN block consists of one temporal convolutional layer (3x3 filter), one batch normalisation layer, one Exponential Linear Unit (ELU) activation layer, one Max Pooling layer (3x3 filter) and one Dropout layer. The CNN block can have two to four blocks of layers in the proposed architecture (see Table II. The Flatten layer aims to flatten all the extracted layers with temporal features. Moreover, the LSTM block consists of 128 units of LSTM layers. The LSTM block has one to two LSTM layers plus one bi-directional layer in the proposed architecture (see Table II. Finally, the

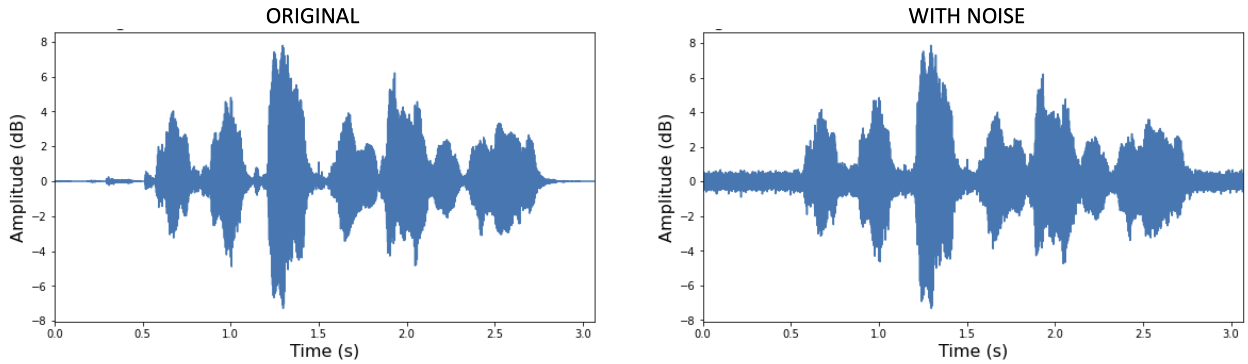


Fig. 1. Speech Signals: Original (left), Augmented with Noise (right).

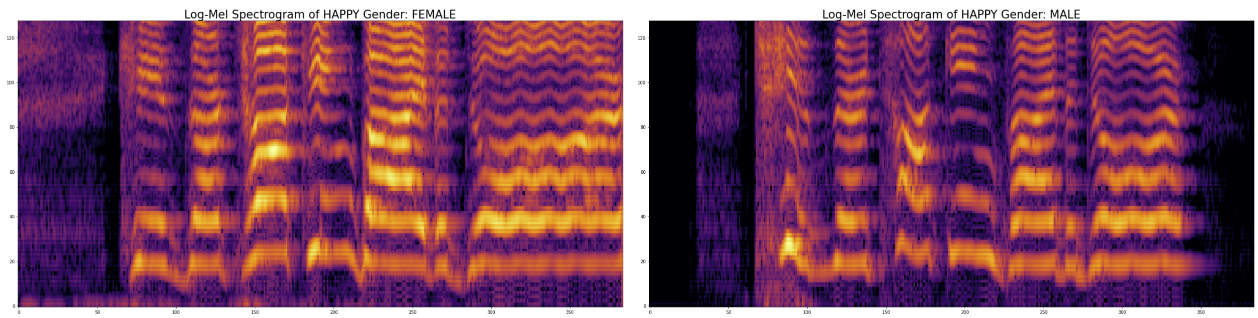


Fig. 2. Mel Spectogram Features: Happy Female (left), Happy Male (right).

features extracted then are classified in the Softmax activation layer. Table II shows the overall settings of the proposed architectures. The CNN column indicated the number of the CNN blocks (e.g. 2 indicates that there are two blocks of CNN layers in the model), while the LSTM column shows the number of the LSTM blocks, where the asterisk mark (*) indicates a Bidirectional LSTM layer. The Dropout implemented in this research is between 0.3 to 0.4. Finally, the column Aug indicates the augmented training dataset used, and three indicates three times the original training dataset and seven indicates seven times the original training dataset.

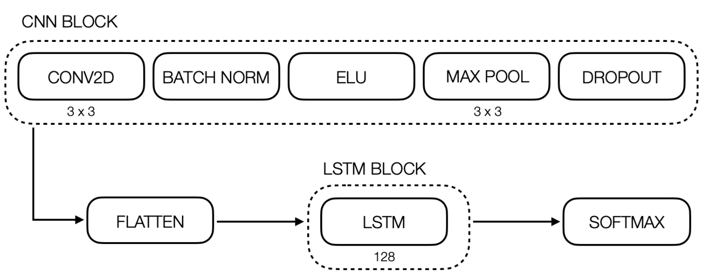


Fig. 3. Proposed Architecture.

IV. Results and Discussion

Five architectures with two settings of augmentation data were explored in this research. The architectures combine Temporal CNN and LSTM (or Bidirectional LSTM) to extract and classify the emotions. Fig. 3 illustrates the baseline of the proposed architectures and Table II demonstrates the proposed architecture settings. This research explores several settings of deep learning architectures and results in the five best architectures that provide the best results. The best result was achieved by the MODEL-5 that consisted of four CNN blocks combined with one bidirectional LSTM layer with a dropout value of 0.3. MODEL-5 also implemented data augmentation seven times bigger than the original training data. The MODEL-5 provides 99.92% and 78.22% of training accuracy and validation accuracy scores, respectively. Moreover, the MODEL-5 provides 0.0144 and 0.8432 of training loss and validation loss, respectively. The dropout value of 0.4 did not significantly improve the model compared to the dropout value of 0.3.

TABLE II. Proposed Architecture Settings

NO.	NAME	CNN	LSTM	DROPOUT	AUG
1	MODEL-1	2	1	0.3	3
2	MODEL-2	2	1	0.4	3
3	MODEL-3	2	2	0.3	3
4	MODEL-4	3	1*	0.3	7
5	MODEL-5	4	1*	0.3	7

TABLE III. Overall Results

NO.	NAME	Train Acc	Val Acc	Train Loss	Val Loss
1	MODEL-1	98.73%	58.15%	0.1396	1.1680
2	MODEL-2	94.25%	57.07%	0.2577	1.1678
3	MODEL-3	96.01%	55.98%	0.2098	1.2720
4	MODEL-4	98.84%	69.85%	0.0625	0.9312
5	MODEL-5	**99.92%**	**78.22%**	**0.0144**	**0.8432**

Table III illustrates the overall results of the experiments.

The results have shown that the models trained with seven times training data augmentation perform better than the models trained with three times training data augmentation. Overall, there are no significant differences in the training accuracy score of models trained with three times training data augmentation compared to the models trained with seven times training data augmentation. However, the seven times training data augmentation model provides higher validation accuracy and lower validation loss. Moreover, the models trained with three times training data augmentation suffer from over-fitting despite batch normalisation and dropout were applied to the CNN and LSTM architectures. MODEL-1 that implemented 2 CNN blocks and 1 LSTM block with 0.3 dropouts trained with three times training data augmentation resulted in 98.73%, 58.15%, 0.1396 and 1.1680 in training accuracy, validation accuracy, training loss and validation loss, respectively. Moreover, MODEL-4 implemented three blocks CNN, one bidirectional LSTM with dropout value of 0.3 and trained with seven times training data augmentation resulted in 98.84%, 69.85%, 0.0625 and 0.9312 in training accuracy, validation accuracy, training loss and validation loss, respectively. Training with two layers of LSTM did not improve the validation accuracy, albeit two layers of LSTM improved the training accuracy. The results show that the model trained with two layers CNN and one layer LSTM (MODEL-2) provides 94.25%, 57.07%, 0.2577 and 1.1678 for training accuracy, validation accuracy, training loss and validation loss, respectively. Moreover, the model trained with two layers CNN and two layers LSTM (MODEL-3) provides 96/01%, 55.98%, 0.2098 and 1.2720 for training accuracy, validation accuracy, training loss and validation loss, respectively. Fig. 4 illustrates the confusion matrix for each classes in the best model (i.e. MODEL-5). The result shows that Neutral (Calm) and Happy emotions are the easiest emotions to classify from the given speech dataset. Moreover, the Angry emotion is the hardest emotion to classify compared to the other classes. The Angry emotion is also mostly miss-classified as the false positive in the other classes. Most likely, it is due to the number of the Angry class in both the training and testing dataset. Finally, the Adam and SGD optimiser do not provide a significant difference to the training accuracy, validation accuracy, training loss and validation loss.

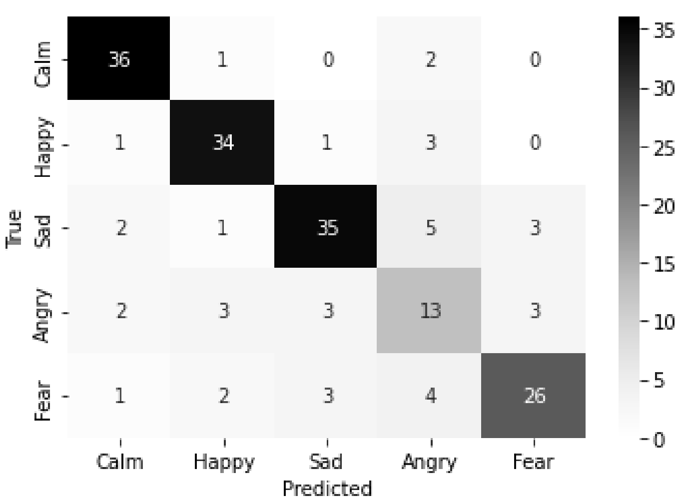


Fig. 4. Confusion Matrix for MODEL-5.

V. Conclusion and Future Work

Five settings of architectures with the combination of CNN and LSTM (or Bidirectional LSTM), number of dropouts and the data augmentation settings were explored in this research. The architectures were implemented to train the emotions recognition models using The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) dataset. The dataset was pre-processed and augmented with two data augmentation settings (i.e. three times and seven times of the original data). The results show that the best model was achieved by MODEL-2, which provides 94.25%, 57.07%, 0.2577 and 1.1678 for training accuracy, validation accuracy, training loss and validation loss, respectively. Moreover, Neutral (Calm) and Happy emotions are the easiest emotions to classify from the given speech dataset, while the Angry emotion is the hardest emotion to classify compared to the other classes. This is due to the number of data in the Angry class in both the training and testing dataset.

For future direction research, more combinations of the architectures, such as the attention architectures and Transformer based architectures, will be explored to increase the recogniser model performances. Moreover, the multi-modal features can also be explored to increase the accuracy and tackle the over-fitting problem. Furthermore, the features from videos (e.g. facial expressions and body gestures), speech and text, can be explored to build a better model for emotion recognition. Finally, the emotions recogniser model that has been trained can be implemented to the more complex affective system such as virtual humans, where recognising emotions can be one of the tools to extract non-verbal meanings from the human interlocutors.

References

[1] Alessandro Vinciarelli, Maja Pantic, Dirk Heylen, Catherine Pelachaud, Isabella Poggi, Francesca D'Errico, and Marc Schroeder. Bridging the gap between social animal and unsocial machine: A survey of social signal processing. *IEEE Transactions on Affective Computing*, 3(1):69–87, 2011.

[2] Alessandro Vinciarelli, Maja Pantic, and Hervé Bourlard. Social signal processing: Survey of an emerging domain. *Image and vision computing*, 27(12):1743–1759, 2009.

[3] Andry Chowanda and Alan Darmasaputra Chowanda. Recurrent neural network to deep learn conversation in indonesian. *Procedia computer science*, 116:579–586, 2017.

[4] Paul Ekman. Basic emotions. *Handbook of cognition and emotion*, 98(45-60):16, 1999.

[5] Yoones A Sekhavat, Milad Jafari Sisi, and Samad Roohi. Affective interaction: Using emotions as a user interface in games. *Multimedia Tools and Applications*, 80(4):5225–5253, 2021.

[6] Shiqing Zhang, Shiliang Zhang, Tiejun Huang, and Wen Gao. Speech emotion recognition using deep convolutional neural network and discriminant temporal pyramid matching. *IEEE Transactions on Multimedia*, 20(6):1576–1590, 2017.

[7] Boaz Shmueli and Lun-Wei Ku. Socialnlp emotionx 2019 challenge overview: Predicting emotions in spoken dialogues and chats. *arXiv preprint arXiv:1909.07734*, 2019.

[8] Kashfia Sailunaz, Manmeet Dhaliwal, Jon Rokne, and Reda Alhajj. Emotion detection from text and speech: a survey. *Social Network Analysis and Mining*, 8(1):1–26, 2018.

[9] Andry Chowanda. Separable convolutional neural networks for facial expressions recognition. *Journal of Big Data*, 8(1):1–17, 2021.

[10] Aditya Joshi, Pushpak Bhattacharyya, and Mark J Carman. Automatic sarcasm detection: A survey. *ACM Computing Surveys (CSUR)*, 50(5):1–22, 2017.

[11] Rémi Delbouys, Romain Hennequin, Francesco Piccoli, Jimena Royo-Letelier, and Manuel Moussallam. Music mood detection based on audio and lyrics with deep neural net. *arXiv preprint arXiv:1809.07276*, 2018.

[12] Yash Mehta, Navonil Majumder, Alexander Gelbukh, and Erik Cambria. Recent trends in deep learning based personality detection. *Artificial Intelligence Review*, 53(4):2313–2339, 2020.

[13] Mehmet Berkehan Akçay and Kaya Oğuz. Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. *Speech Communication*, 116:56–76, 2020.

[14] A Rakhlin. Convolutional neural networks for sentence classification. *GitHub*, 2016.

[15] WQ Zheng, JS Yu, and YX Zou. An experimental study of speech emotion recognition based on deep convolutional neural networks. In *2015 international conference on affective computing and intelligent interaction (ACII)*, pages 827–831. IEEE, 2015.

[16] Ruhul Amin Khalil, Edward Jones, Mohammad Inayatullah Babar, Tariqullah Jan, Mohammad Haseeb Zafar, and Thamer Alhussain. Speech emotion recognition using deep learning techniques: A review. *IEEE Access*, 7:117327–117345, 2019.

[17] George Trigeorgis, Fabien Ringeval, Raymond Brueckner, Erik Marchi, Mihalis A Nicolaou, Björn Schuller, and Stefanos Zafeiriou. Adieu features? end-to-end speech emotion recognition using a deep convolutional recurrent network. In *2016 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 5200–5204. IEEE, 2016.

[18] Panagiotis Tzirakis, Jiehao Zhang, and Bjorn W Schuller. End-to-end speech emotion recognition using deep neural networks. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5089–5093. IEEE, 2018.

[19] Fabien Ringeval, Andreas Sonderegger, Juergen Sauer, and Denis Lalanne. Introducing the recola multimodal corpus of remote collaborative and affective interactions. In *2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG)*, pages 1–8. IEEE, 2013.

[20] Soonil Kwon et al. A cnn-assisted enhanced audio signal processing for speech emotion recognition. *Sensors*, 20(1):183, 2020.

[21] Steven R Livingstone and Frank A Russo. The ryerson audio-visual database of emotional speech and song (ravdess): A dynamic, multimodal set of facial and vocal expressions in north american english. *PloS one*, 13(5):e0196391, 2018.

A Comprehensive Overview on Biometric Authentication Systems using Artificial Intelligence Techniques

Shoroog Albalawi[1], Lama Alshahrani[2], Nouf Albalawi[3], Reem Kilabi[4], A'aeshah Alhakamy[5]
Faculty of Computers and Information Technology,
Master of Artificial Intelligence at University of Tabuk, Saudi Arabia[1,2,3,4,5]
Industrial Innovation & Robotics Center (IIRC) and Faculty of Computers and Information Technology,
Department of Computer Science at University of Tabuk, Saudi Arabia[5]

***Abstract*—Biometric authentication is becoming more prevalent as it allows consumers to authenticate themselves without entering a physical address or a personal identification number. Thus, a simple finger gesture or a glance at a camera can still prove one's identity. In this review, we explain in detail how the concept of authentication and the various types of biometric techniques is used for user identification. Then, we discuss the various ways these techniques can be combined to create a truly multimodal authentication system. For a more organized approach, our overview is classified into two main categories based on human biometric traits. First, the physiological traits include fingerprint, facial, iris/retina, hand, and finger-vein. Second, the behavioral traits includes voice, signature, and keystroke recognition systems. Finally, we offer a comprehensive comparison of selected methods and techniques and focus on three criteria: algorithms, merits, and drawbacks. Based on this comparison, we provide insight into our future research in iris recognition, by which we combine several artificial intelligence algorithms to develop our system.**

Keywords*—Biometric authentication; physiological traits; behavioral traits; facial recognition; iris recognition; voice recognition; signature

I. INTRODUCTION

Authentication is a security function that involves providing and checking the proof of the person's identity, the message sender, the software, the logical server, or the device. Different identification methods have been proposed for exchanging safety-related information [1], [2], [3], [4], faced with the logical evolution of international regulations; new technological solutions are gradually being implemented. Among these technologies, biometrics is the most relevant technology to reliably and rapidly identify and authenticate a person, based on their unique biological characteristics.

Biometric authentication uses a person's biological characteristics to verify their identity and ensure secure access to an electronic system. The biometric technologies are based on how each person can be identified distinctly through one or more biological characteristics, such as fingerprint, hand morphology, retina and iris, voice, DNA, or signatures. In general, biometrics can be classified into two categories: physiological traits, and behavioral traits, as shown in Fig. 1. Biometric authentication is the application of these biometric technologies to identify a person as part of a user validation process to access a system.

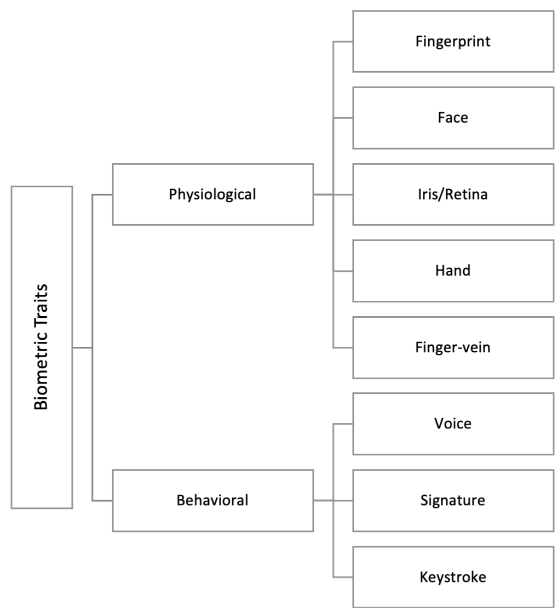


Fig. 1. Classification of Biometric Traits.

The motivations and potential benefits that drive our research are the convenient use of this approach in different systems around the world. Our future project involves using iris recognition methods in airports for fast and accurate authentication. The objective is to identify the human iris and link the data together by using artificial intelligence (AI) techniques to extract a passenger's information, such as flight number, dedicated gate, seat number, departure and arrival times, travel bag weight, how many kilos are reserved for travel bags, vaccination status, and chronic disease diagnosis (e.g., diabetes, high blood pressure).

Because different types of information can be extracted from the human iris, from our point of view, we believe the following methods better fit for our future work. (1) The edge detection algorithm is used to localize the iris and select the most important features. The system can then be segmented and localized by using the fuzzy algorithms and the edge detection. (2) statistical features extraction methods are then used to extract the characteristics of human iris. (3) The feature space is then divided into classes, and the extracted features are trained parallel to partition the features into these classes using Principal component analysis (PCA). (4) The Support vector machines (SVMs) algorithm is then used to classify the new images in the database according to their classes to pinpoint the traveler's identity.

In this work, we present a review of biometric authentication methods and techniques that ease the way people interact with systems and how their identity influences governmental and private systems. Section III introduces physiological biometrics, and the behavioural biometrics are presented in Section IV. In these sections, we briefly discuss and list the most common biometrics, such as the facial, iris, fingerprint, voice, and signature recognition techniques. We also depict, the typical design steps of each system and its applications to provide a thorough description.

II. Biometrics Recognition Systems

Biometric technologies are used to secure a wide range of electronic communications or to connect to a computer or smartphone. Biometric authentication systems compare the human biometric data to be authenticated with the biometric database. If the two samples are matched, the authentication is then confirmed and access is granted. This process is sometimes part of a multifactor authentication system. Thus, the smartphone user can connect using their secret code such as a personal identification number [PIN] and add an iris scan. Generally, biometric identifiers are classified into physiological or behavioral characteristics; see Fig. 1.

Human biopsies are related to the various physiological characteristics of the human body such as fingerprints, facial features, iris, retina recognition, and DNA. Behavioral biometrics are related to a behavioral pattern, like the rhythm of a person typing, or how they use their fingers or look at the camera.

III. Physiological Biometrics

Physiological or static biometrics use physical characteristics, such as fingerprint or facial recognition, etc. to unlock cell phones, log into bank accounts, or complete transactions. However, the main types of static biometrics used to verify a person's identity are fingerprint recognition, facial recognition, iris recognition, etc.

A. Fingerprint Recognition

The fingerprint is one of the oldest forms of biometric authentication, and mobile platforms use this technology widely. It was originally popularized by Apple's Touch ID. A fingerprint reader analyzes a person's fingerprint and compares it to the finger's stored digital pattern during authentication. Fingerprint recognition may change if the finger is wet or dirty. An attacker cannot replicate a person's fingerprint because of its vividness; however, it can be used to create a 3D model or a fake image.

Fingerprint authentication is based on the concordance between the registration or signature file obtained during enrollment and the file obtained during authentication. Several methods are used to recognize fingerprints, such as locating minutiae and processing textures. The process of extracting minutiae involves using the template comparing to the image digitization and minutiae extraction, as shown in Fig. 2.

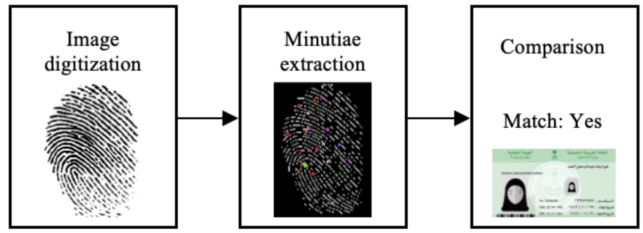


Fig. 2. Fingerprint Authentication Process: (1) Image Digitization, (2) Minutiae Extraction and (3) Comparison to the Templates.

The image digitization step consists of digitizing the fingerprint, filtering unnecessary features (e.g., scars), and determining information useful to the system. To detect the endpoints and crossing points of ridges, known as minutiae, a fingerprint skeleton is created using complex algorithms in order to make each line of the imprint, with a length from 5 to 8 pixels and a thickness of 1 pixel. Fig. 3 shows the creation of a fingerprint skeleton.

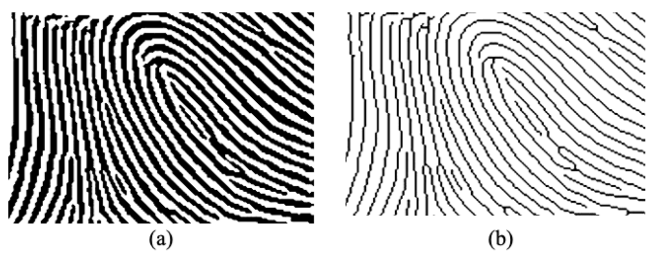


Fig. 3. Fingerprint Skeleton Creation.

The minutiae allows the extraction step to determine the signature of the fingerprint using different algorithms. The template used to characterize the fingerprint is based on a minimum number ranging from 12 to 14 minutiae, although this number can be higher as needed to establish a reliable comparison. In fact, it is possible to identify one fingerprint among several million fingerprints with this minimum of correctly identified and localized minutiae.

Finally, the comparison of two minutiae's, corresponding to two fingers to be compared, constitutes the identity verification system. To determine whether two minutiae's extracted from two images correspond to the same fingerprint, it is necessary to adopt a comparison system that is insensitive to any translations, rotations, or deformations, which systematically affect the fingerprints. From two extracted minutiae's, the system should be able to give a similarity or correspondence index of 0% if the fingerprints are totally different, and 100% if the fingerprints are from the same image.

Fingerprint recognition algorithms are sensitive to the image quality. The pretreatment step is therefore necessary before performing the following steps. The quality of fingerprint images depends on several factors, such as contact with the probe, quality of the probe, and depth of ridges/bifurcations. Generally, preprocessing consists of smoothing, contrast enhancement, spatial/frequency domain filtering. Today, several techniques are used to solve the problems associated with fingerprints recognition.

A comprehensive overview of the patterns and techniques used in fingerprint recognition depends on the minutiae-based technique. [5]. Peralta, et al. [6] described the various aspects of fingerprint authentication and identification with respect to the minutiae-based matching algorithm. Fingerprint authentication utilizing minutiae extraction technique are discussed by Sharma, et al. [7] who covered all related systems and processes. Unimodal and multi-modal biometrics techniques were summarised by Delac and Grgic [8] including pros and cons for each model.

B. Facial Recognition

facial recognition is a technique used to identify/verify human identity based on their facial features. facial recognition can be performed from photos or video recordings. To develop a system with robust facial recognition, four steps are taken under consideration: (1) facial detection, (2) feature extraction, and (3) feature classification, and (4) feature matching. The face detection step involves identifying the human face in the image, and the feature extraction step involves extracting the feature vectors for the identified facial. The feature extraction step is considered the most crucial step in the facial recognition process. The results can then be compared and classified according to a certain criterion to identify the features of the image. These steps are illustrated in Fig. 4.

There are three major approaches to automatic face recognition by computer: global feature approach (facial-based recognition), local feature approach (constituent-based recognition), and hybrid approaches.

1) Global methods: The global approach is commonly used to identify facials using the entire image without taking into account the a face's local physiological features, such as the eyes and mouth.

The global algorithms are based on statistical properties and are usually quick to implement. However, they are sensitive to various factors such as lighting conditions and facial expression. Some of these algorithms are principal component analysis (PCA), linear discriminant analysis (LDA), support vector machine (SVM), and neural network (NN).

In this context, Sugandi, et al. [9] have proposed a facial recognition method based on PCA and backpropagation NN. Each facial image in the training is represented exactly by a linear combination of eigenfaces. This method is performed in three stages. In the first step, facial detection is performed using Haar-like features. In the second step, the authors use the AdaBoost learning algorithm to select the most important features. Finally, the backpropagation NN is used for the recognition process. In their work, the authors demonstrated that using 5 data facial images with each data are taken 100 times than usual, the experimental result showed the satisfactory result with 87.5% recognition rate.

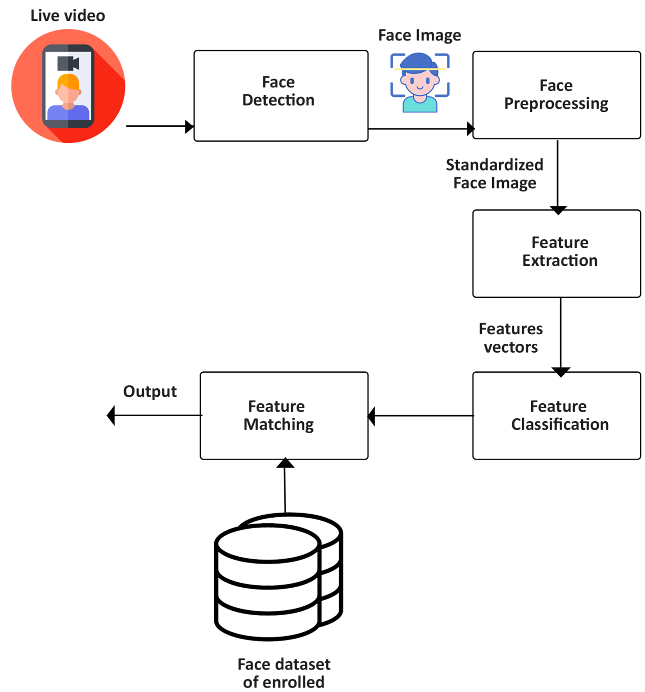


Fig. 4. Facial Recognition Structural Steps: (1) Facial Detection, (2) Feature Extraction and (3) Facial Recognition.

A study conducted by Anusha, et al. [10] showed that the weight of the image when compared with the test images can be used to identify the facials. With the same objective, Shen, et al. [11] pa method that uses both the, (1) PCA algorithm to get the feature space of the training set and then trains to identify the projected facial and (2) Fisher linear discriminant (FLD) algorithm to obtain the fusion feature space and then to train and recognize the projected facial in the feature space. Their study proved that the algorithm based on the FLD and PCA features can outperform the existing facial recognition methods.

Luabibi, et la. [12] proposed a four-phase approach for facial recognition. The first step is the localization of the facials, and the second involves extracting the features, and the third one is the classification, and the fourth phase is the back propagation algorithm to identify the facials.

The LDA method is a numerical technique that can be used to classify and improve the data representation. The resulting combinations are usually used as linear classifiers. This algorithm cuts each facial into a linear combination before classifying it. The resulting sets of new dimensions are called fisherfaces after the LDA method. Due to its complexity, the LDA algorithm is not as effective as other methods.

Bhattacharyya, et al. [13] proposed a method that involve grouping images of the same class and different classes. The resulting set of facials is then classified according to the closest training images. The study's results revealed that the proposed algorithm is significantly better than the existing methods. Lu, et al. [14] combined both the direct LDA and Fractional LDA techniques, and algorithms can be used for small sample sizes.

It can also classify the facials according to the closest training images. Results of the experiments on two databases supported the effectiveness of the proposed algorithm.

SVM's are commonly used in machine learning to solve problems related to discrimination and regression. Although they are not as efficient as some of the most popular algorithms, their potential is still very promising. In their study, Jin, et al. [15] proposed an algorithm based on a modified SVM learning scheme that uses the SVM and particle swarm optimization techniques. The results of the experiments showed that the proposed method has better accuracy than the existing methods.

In addition, Jose, et al. [16] presented a review of the techniques used for 2D facial recognition using the SVM technique. The authors analyzed the recognition results according to the techniques used to extract information such as facial features, pattern classifiers, and databases. A new method that uses the SVM algorithm to detect the facials in grayscale images was presented by Ignas et al. [17]. They then identifed the faces using the sizes and positions of the eyes and lips.

Also, Javed [18] used the PCA and SVM to build for building a facial recognition model. The authors [19] presented an overview of the effective use of machine learning, particularly regarding using SVMs in facial recognition. Therefore, the authors give an extensive survey of facial recognition and its applications.

Artificial Neural Networks (ANNs) are systems comprising of several interconnected processing units. They can perform various computational tasks based on the input data. In computer vision, an NN is composed of several processing units known as neurons. These components can learn and adapt to different tasks in order to classify the data.

Deep learning (DL) is an AI method that learns by itself. It is inspired by the human brain. The goal of DL is to learn and recognize different words and facials in an image. For example, DL can detect the letters in text before recognizing a face. In their work, Hassan, et al. [20] detailed the various approaches used in facial recognition and provided a thorough analysis of their results. They also introduced hybrid algorithms that can be used for extracting and classifying facial features.

Convolutional Neural Networks (CNNs) are commonly used in facial recognition. They are typically complex and require a high amount of processing power and storage space to perform their intended applications. In this area, Liu, et al. [21] tried to improve the performance of CNNs by introducing a block called the squeeze-and-excitation algorithm. The proposed algorithm has fewer parameters and can be more suitable for various applications.

2) Local methods: These are also called the geometrical, local characteristics, or analytical methods. This approach involves in applying transformations in specific places of the image, most often around the characteristic points (i.e., corners of the eyes, mouth, and nose). Attention is given to small local details, and the approach avoids the noise generated by aspects such as hair, eyeglasses, hats, and beards.

However, the difficulty of these methods arises when it taking into consideration several views of the facial and the lack of precision in the extraction phase of the points, which constitute their major drawback. Specifically, the methods start by extracting the local facial features such as the nose, eyes, and mouth; and the method then use their geometry and/or appearance as input to the classifier. Hence, we can distinguish two practices:

- The first practice is based on the extraction of entire regions of the facial; it is often implemented with a global facial recognition approach.
- The second practice extracts particular points from different characteristic regions of the facial, such as the corners of the eyes, mouth, and nose.

Among these approaches, we can list the hidden Markov model (HMM) and the elastic bunch graph matching algorithm (EBGM). An HMM is a statistical model in which the modeled system is assumed a Markov process with unknown parameters. Unlike a classic Markov chain, where the transitions have taken are unknown to the user but the states of execution are known, in an HMM, the states of execution are unknown to the user. HMMs are currently among the most widespread models of form recognition. They then were established in speech recognition and written recognition.. However, when using HMMs, the structure of the facial is considered as distinct regions, described by characteristic vectors.

In this context, Alhadi, et al. [22] studied three different methods to extract feature vectors from HMMs: discrete cosine transform, discrete wavelet transform, and PCA. The results of the experiments revealed that combining these methods improved the models' recognition performance. In addition, a state-of-art HMM model applied to facial recognition problems in the review [23]. In this work, the authors have presented the evolution of HMM use from the early 1990s to the present day; this makes it easier for new researchers to understand and adopt the HMMs more easily for facial recognition.

In 1997, Wiskott, et al. [24] proposed a graph matching algorithm commonly used in computer vision. This algorithm makes it possible to recognize objects in an image by using a graphic representation extracted from other images. This approach extracts a set of characteristics using a data structure called a packet graph [25]. Furthermore, Jaiswal [26] presented a method for recognizing human facials that uses the concept of a graph matching algorithm known as EBGM. The proposed method achieves high recognition rates for both the facial and the image graphs.

3) Hybrid methods: Hybrid methods pair the advantages of global and local methods by combining the detection of geometric (or structural) characteristics with the extraction of local characteristics. They make it possible to increase the stability of the recognition performance during changes in pose, lighting, and facial expressions. Local feature analysis and Gabor wavelet extracted features (such as EBGM) are typical hybrid algorithms [27], [28], [29].

C. Iris / Retina Recognition

The ability to identify the facial using the iris is one of the most accurate and secure methods for biometric identification. Unlike the hands and facial, the iris is a protected internal organ and is therefore less, and therefore less likely to be damaged. The user must fix a digital camera that scans the iris of a person from a distance of 30 to 60 cm and directly acquires iris drawing. Then, this drawing is compared to a computerized personal identification file to identify the person.

An iris recognition system is a type of biometric system that uses images of human irises to identify people. It features two main processes: localization and segmentation. Images of the iris are determine and analyzed to extract its biometric signature. However, the image processing operations are divided into four stages: localization, segmentation, encoding, and classification. The diagram in Fig. 5 shows the methodology steps and sequential processing of the proposed iris recognition system.

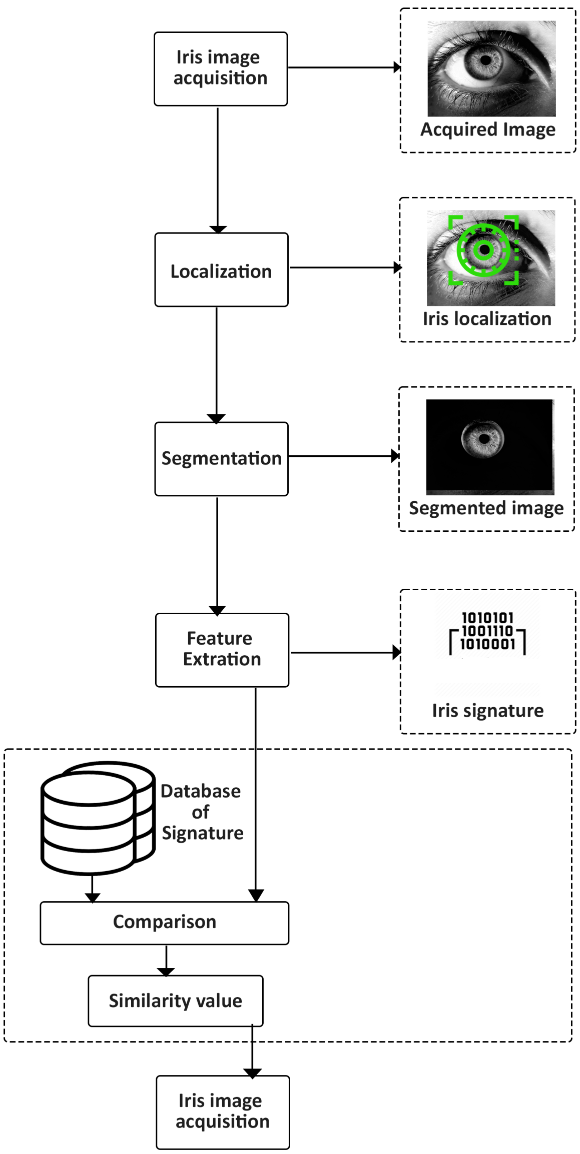


Fig. 5. The Architecture of the Iris Recognition System which Include; (1) Image Acquisition, (2) Iris Localization, (3) Segmentation and (4) Feature Extraction.

1) Iris Recognition: A new approach for the recognition of the human iris was developed by Fernando, et al. [30]. It uses the SIFT feature transformation to extract the characteristic features and then matches two images. Experiments with the BioSec database show that the SIFT approach achieves better performance than the existing matching methods.

A novel approach for deep learning that aims to improve the accuracy of the recognition of the iris using a more simplified framework was presented by wang and Kumar [31]. It utilizes residual network learning and dilated convolutional kernels to improve the training process. An unsupervised network approach can greatly simplify the network and provide better performance than state-of-the-art algorithms for iris recognition. It also eliminates the need for upsampling and downsampling layers. The results of our experiments demonstrate the applicability of our approach to improve the accuracy of iris recognition.

Furthermore, a new generalizability method for the recognition of the iris is introduced by Adamović, et al. [32], which performed on the two CASIA databases. The results of the experiments demonstrated the system performs as expected. The generalizability method can significantly reduce the computational costs of the system, which in turn makes the method suitable for practical applications. The goal is to achieve a classification accuracy of almost perfect. The method also eliminates the possibility of generating an image from a template.

2) Retina Recognition: The retina is a region of the eye that consists of four layers of cells. The arrangement of these cells is unique and provides a high level of recognition. This technology is well-suited for high-security applications because it can achieve a recognition rate of around 90% compared to other methods. The idea for retinal identification and unique vascular pattern was first introduced by two ophthalmologists, Dr Carleton Simon and Dr. Isodore Goldstein in 1935 [33].

A new method for retina recognition based on a fractal dimension was also presented by Sukumaran, et al. [34]. The authors of the study compared the accuracy of this technology with the commercially available ones. The experimental results of the method revealed that it produces high accuracy and low computational cost.

With the same objective, Tuama and George [35] proposed a personal identification system using the vascular diagram of the human retina. This system is composed of four stages. First, the prepocessing technique is used to extract the retinal image from the background and to remove noisy areas from the retinal image. Then, wavelet transforms 2D and adaptive thresholding were used to extract the blood vessels. Next, the system performs feature extraction and filtration. Finally, the matching step is used for the retina recognition. Experimental results on three publicly available databases (DRIVE, STARE, and VARIA) have demonstrated that the proposed method is better than several existing techniques.

IV. Behavioral Biometrics

Behavioral biometrics analyze a person's unique habits and movements to create a behavioral pattern. Like static biometrics, behavioral biometrics adds another layer of security to verify a person's identity. Accordingly, this technology uses motion sensors and AI to identify unique behaviors, such as how a phone is held. These technologies are is widely regarded as the last frontier in security [36], [37].

A. Keystroke Dynamics Recognition

Stroke rhythm analyzes a person's typing rhythm on digital devices (e.g., smartphones) to create a form of human digital footprint or signatures. Gaines, et al. [38] first proposed this technology when they created the first automated dynamic keystroke recognition systems [39].Major contributions have notably evaluated the fuzzy logic [40], NNs, [41], [42], and different pattern recognition techniques (e.g., Bayes classifier) [43], [44].

B. Signature Recognition

The recognition of a signature can be accomplished by analyzing a large number of discriminative variables: (i) global characteristics such as writing time or the number of touches on the tablet with the pen, and/or (ii) local characteristics such as the position of certain curvatures or the instantaneous speed. A pressure-sensitive pencil-shaped reader and a digital tablet usually acquire a signature.

Developing a robust facial recognition system involves three basic steps: (1) preprocessing, (2) feature extraction, and (3) signature matching and classification. Before the algorithm can be used for facial recognition, the preprocessing step involves removing background noise and refining on the signature. The signature recognition system architecture is presented in Fig. 6. Feature extraction is the next step in the recognition process. It involves extracting the various features of the human signature, such as the Walsh coefficient, grid, and texture. The signature recognition step matches the various features of the human signature to template signature databases.

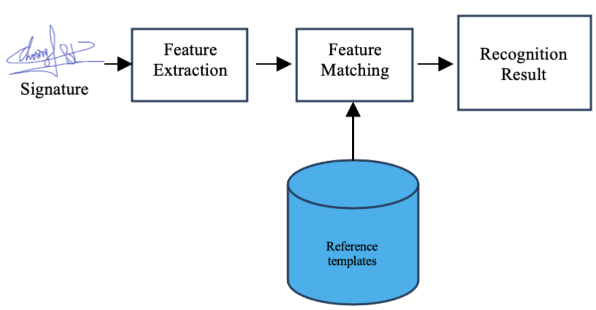


Fig. 6. Signature Recognition System Architecture that begins with Feature Extraction, then Matches these Features with a Reference Template and Finally Obtains the Recognition Result.

The best-known comparison techniques use a HMM [45], or a dynamic programming approach [46], [47].The International Graphonomic Society (IGS) research community, particularly the Scribens team, proposed the most important contribution to handwritten signature verification [48]. In this area, Jain, et al. [46] proposed to use a new measure of dissimilarity based on the alignment of characteristic vector sequences by dynamic time warping. This technique represents the most recent work in manual signature verification

C. Voice Recognition

Voice recognition is a process that uses the sounds produced by a person's vocal tract and the shape of their nose, mouth, and larynx to identify their voice. Analyzing a person's voice is a strong authentication method, but illness (e.g., the common cold, bronchitis) and background noise can distort the voice and disrupt authentication. The architecture of an automatic voice recognition system is represented in Fig. 7.

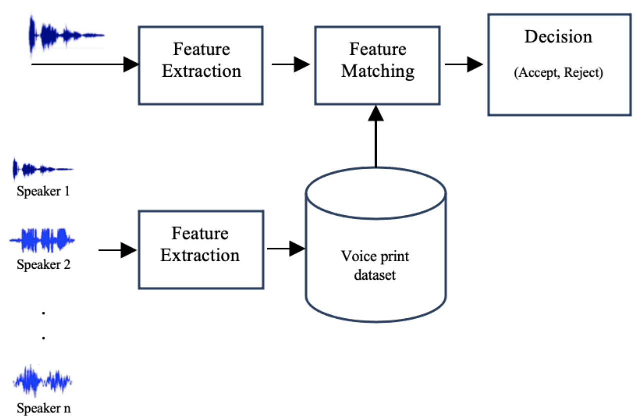


Fig. 7. Voice Recognition System Architecture that Includes Feature Extraction, Feature Matching and Decision-making based on the Voice Print Dataset.

However, the applications of speech recognition are diverse and each system has its own architecture and operation mode. In 1987, the Worlds of Wonder Company first marketed speech recognition via a doll named Julie [49]. For its voice recognition algorithm, Texas Instruments used a digital signal processing system that could recognize eight different sentences.

In addition, voice recognition has been studied for many years to help people with disabilities [50], [49], [51]. People without mobility can control an electric wheelchair through voice commands [52].

V. Discussion and Comparison

In this section, we provide a summary and discussion of the review and outline different methods of biometric authentication. We select and compare some of these methods and observe each method's algorithms, advantages and disadvantages. The study of limitations is key to improving in the future work, and we must highlight them to build our own methodology.

It is crucial to acknowledge that each method is not excluded from the others. These techniques can be combined with one another or used in different parts of the whole process to achieve the desired realism.

To explore each technique and compare these methods, we established criteria to recognize the major differences and how they affect the final outcome. The major differences in aim, accuracy, and robustness among methods make concluding the compression challenging.

Therefore, based on the information provided in each study we attempted to distill the most useful information to write this section. We believe this effort will be helpful as a future reference for our work and that of others regarding usability, performance and more. Thus, the criteria are as follows:

- **Method.** The fundamental calculations and algorithms essential for the system's methodology and structure.
- **Merit.** The advantages the system provides, including the accuracy of the result under the previous criteria.
- **Drawbacks.** The system's limitations based on the required data, assumptions, and outcomes. Future work usually starts by discussing and discovering these drawbacks and then attempting to solve them.

We could cover more criteria, but doing so requires insight into and evaluation of each system, which is outside the scope of the current review. We might approximate for the time and effort for the rendering and processing, but the result would be untrustworthy without any quantitative data.

An overview of these criteria to compare the previous methods from selected papers are presented in Table I. These papers were chosen based on two factors: (1) significance in their field, (2) ability to offer insights for our future work.

VI. Open Problem and Proposed Methodology

Biometric identification is a process that uses sensors to measure a person's biological characteristics. The data these devices collect then can be compared to information stored in a database. In addition to fingerprints and eyes, biometrics such as facial recognition and hand geometry have been studied and used. Eye biometrics offer the highest level of accuracy and individuality.

Iris recognition is a biometric technique that allows recognition of a person by observing their iris. Generally, the iris recognition system includes a series of steps: (i) image acquisition; (ii) iris preprocessing, including localization and segmentation; (iii) feature extraction; and (iv) matching and classification. The diagram in Fig. 8 shows the methodology steps and sequential processing of the proposed iris recognition system.

For passenger identification at an airport, we propose an iris recognition system that performs the feature extraction method and the PCA algorithm to extract and select the most important statistical features of the human iris. In addition, this system performs a supervised classification process using the SVM algorithm to identify a person's identity. The proposed system will be applied to a collection of human iris databases, such as the CASIA iris database.

This system consists of two main phases: the preprocessing phase and the classification phase. In the preprocessing phase, the system will acquire the required human iris image. The iris's features will be processed to obtain information using the statistical feature extraction method. Next, these features

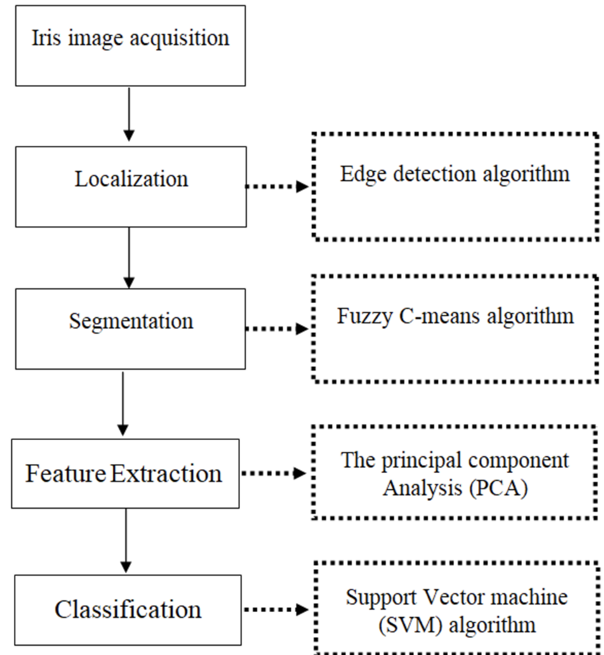


Fig. 8. Methodology Steps for Our Future Implementation in Iris Recognition System.

will be refined using the PCA reduction technique and will be the input to the SVM algorithm for classifiying human iris.

Once the iris's image is acquired, image-processing techniques are used to extract the iris, construct biometric signature, and finally find its identity. However, the acquired image must have a minimum resolution and quality to ensure the characteristics necessary for representation and identification have been obtained. Therefore, efficient segmentation of these regions is necessary.

Iris localization is one of the most important steps in an iris recognition system because it determines the matching accuracy. This step mainly localizes the boundaries of the iris, which are the inner and outer boundaries of the iris and upper and lower eyelids. To do this, edge detection techniques can be used to localize the iris in the original image. After acquiring the image, the iris should be isolated. As a first step, a series of image enhancement operations such as filtering, contrast enhancement, and histogram equalization type can be applied. The goal is to enhance the quality of the image and then apply iris segmentation operations.

The processing set necessary to extract the iris from its environment defines the iris segmentation: pupil, sclera, eyelids, eyelashes, and specular reflections. This is the most difficult step in the recognition system because it affects the system's performance.

Segmentation approaches are based on detection by fuzzy algorithms such as the fuzzy c-means algorithm. In this step, the iris is segmented in details. The encoding consists of extracting the iris's most discriminating and relevant characteristics, which are required for its identification.

To do this, the statistical features are extracted from the iris

TABLE I. COMPARISON OF PREVIOUS METHODS FROM THE SELECTED PAPERS THAT WE COVERED FOCUSING ON THREE CRITERIA: ALGORITHMS, MERIT AND DRAWBACKS

Research	**Method**	**Merit**	**Drawbacks**
		PHYSIOLOGICAL BIOMETRICS	
		Fingerprint Recognition Methods	
Borra, et al. [5]	-Minutiae based approach. -Pattern recognition. -Wavelet.	High accuracy rate.	-No noisy/encrypted images, -Slow performance due to three levels of split texture -Fails to determine real humans sometimes.
Peralta, et al. [6]	-Minutiae-based local matching -Correlation-based matching -Indexing	-Simplicity -Distortion tolerance.	Expensive computation, slow and depend on the skin situation.
Sharma, [7]	Minutiae based matching.	Widely used and familiar.	Affected with wet or dry skin.
Delac, et al. [8]	-Unimodal biometric systems -Multimodal biometric system.	Reliability due to use the combination of different biometric strength.	-Noisy Scanned data. Difficulty in the data gained from humans. -Biometric sign can expose to forgery.
		Facial Recognition	
Budiet, et al. [9]	-PCA -Back Propagation Neural Network.	-No Correlated Features, -High Performance, -Reduce Overfitting, -Improved Visualization, -No user action: not very intrusive, No physical contact.	Independent variables become less interpretable, -Data standardization is must before PCA, -Information Loss.
Anusha, et al. [10]	PCA	-Not intrusive, done from a distance, -Inexpensive technique. -Several characteristic features	-More suited for authentication than for identification purposes. -User perceptions and civil liberty.
Shen, et al. [11]	-PCA -FLD .	-Reduced Overfitting. -Improved Visualization.	-Data standardization before PCA. -Information Loss.
Morooj, et al. [12]	-PCA -Backpropagation NN	-Removes Correlated Features. -Improves Algorithm Performance	-Independent variables become less interpretable, -Information Loss
Bhattacharyya, et al. [13]	LDA	-Objective evaluation. -small set of features for classification purposes. -Overcomes the limitation of PCA by applying the linear discriminant criterion.	-Singular within-class scatter matrix due to small size sample. -Difficulty differentiating identical twins. Sensitive to changes such as beard and glasses.
Jin, et al. [15]	-PSO -SVM	Effective in high dimensional spaces.	-Not suitable for large data sets. -Premature convergence of PSO leading to stagnate in local optimum. -Environmentally sensitive technology.
Jose, et al. [16]	SVM	Effective in high dimensional spaces.	Not suitable for large data sets.
Ignas, et al. [17]	-SVM -Gaussian kernel for grayscale.	Effective in high dimensional spaces	-Not suitable for large data sets. -No Noisy data. -No Overlapped classes.
Mahnoor, [18]	-PCA -SVM	Effective in high dimensional spaces	-Not suitable for large data sets. -No Noisy data. -No Overlapped classes.
Jaiswal, et al. [26]	EBGM.	Insensitive to lighting variation -Rigid, and deformable matching.	More complicate procedure.
		Iris/Retina Recognition	
Fernando, et al. [53]	SIFT	- Stable in lighting and perspective variations. -Locality. -Resistant to occlusion and crowding -No prior segmentation. - Distinctiveness. -Quantity, generate many features.	-Complicated Mathematically. .High computation cost. .Not effective for low powered devices.
Poursaberi, et al. [54]	-Wavelet-based texture Hamming distance of minimum and harmonic mean	-Efficiency: close to real-time performance. -Extendable -Robust High accuracy. -Iris does not change over time	-Acquiring an image requires proper alignment and positioning. -Result affected by pupil size change.
Félix, et al. [55]	Facial key-point detection, Integro-differential operator (IDO) Mathematical morphology	-No intimate contact with the reader. -More robust than voices.	-Hard to use. -Difficulty integrating with other systems
Vahid, et al. [56]	Pattern Recognition Approach.	-Higher average for matching performance. -Convenient for people who wear glasses. -Low chances of a false positive.	-The eye position can be problematic. -Expensive specialized devices.
Sukumaran, et al. [57]	Fractal dimension using box counting Pattern Recognition Approach	-Most reliable biometric technology -Unique data points	-Serious health risk, infrared light beam. -Very large false rejection rate
Saba, et al. [58]	-Vascular diagram of the human retina -Pattern Recognition Approach	-Very accurate. -Impossible to forge a retina. -Low error rate. -Low false rejection rate -Low false acceptance rate.	-Inconvenient for people who wear eyeglasses. -Uncomfortable for some users. -Retina biometric devices.
		BEHAVIORAL BIOMETRICS	
		Keystroke Dynamics Recognition	
Gaines, et al. [38]	Examine the probability distribution of time to pressed/relaxed each key while typing on keyboard.	-Ergonomic. -Uniqueness. -Low Implementation and Deployment Cost. -Transparency and Noninvasiveness. -Increase Password Strength and Lifespan. -Replication Prevention -Additional Security. -Continuous Monitoring and Authentication	-Person's physical condition is important. -Lower Accuracy. -Lower Permanence .

Signature recognition			
Jain, et al.[46]	Behavioral biometrics based on person's handwriting	-Ergonomic. -Highly resistant to impostors. -Enrollment is intuitive and fast. -Fast response -Low storage needed. -Native language don't matter. -Little time of verification.	-Depend on the emotional state. -Difficult to use. -Large template. -High Cost -Vary over time.
Voice recognition			
Simpson, et al. [52]	Arduino, HM2007 Voice recognition module and Motors.	-More natural and easier- -Faster than typing -Can be done over the phone- No specialized equipment -Not intrusive, Low computing power	-Not reliable. Sensitivity to environment variations.

image, and the linear dimensionality reduction technique PCA is used to refine the features. Therefore, the iris encoding process constitutes representing the iris's signature. The encoding or feature extraction step is unique for each iris, regardless of dimensional variations or rotations created during the iris acquisition. This information then will be used to classify the iris.

VII. Conclusion and Future Work

Biometrics involves studying physical characteristics to identify human. With information technology and digitization, these techniques have undergone major development. Generally, biometrics can be applied to individuals who have voluntarily shared information with the biometric system, which must control them based on characteristics such as their fingerprint or voice. This step represents the enrollment procedure. This biometric information is then transformed into a digital file called a signature or template, using a specific algorithm that retains and encrypts the characteristic elements of the digitized image. This signature is then archived and compared with the person's characteristics during the control.

In this paper, we presented a detailed review of biometric systems and compared them to find the best methods and algorithms for our future work. We summarized the notion of authentication and authentication technologies and focused on biometric authentication methods. In addition, we briefly discussed the most common biometrics (i.e., face, iris, voice, signature, and fingerprint recognition techniques) and the different ways to combine them to obtain multimodal systems. Finally, we presented the typical design steps of each system and biometric applications. Consequently, the process automation, particularly the measure of characteristics such as the iris, the shape of the face or hand, the voice, the speed of typing or pressing keys, the dynamics of signature represent some challenges that can be develop in the authentication system.

After collecting all this information, we will present a hybrid approach that combines SVM-based classification and AI techniques to perform advanced iris recognition. The first step involves developing an edge detection algorithm that selects the most important features. The fuzzy and edge detection algorithms are then used to segment and localize the system. The extracted features are then divided into classes using the PCA algorithm. The proposed methodology will be used to identify passengers at an airport in our future work.

References

[1] C. Otti, "Comparison of biometric identification methods," in *2016 IEEE 11th International Symposium on Applied Computational Intelligence and Informatics (SACI)*. IEEE, 2016, pp. 339–344.

[2] A. Sumalatha and A. B. Rao, "Novel method of system identification," in *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*. IEEE, 2016, pp. 2323–2328.

[3] M. Ortega, M. G. Penedo, J. Rouco, N. Barreira, and M. J. Carreira, "Retinal verification using a feature points-based biometric pattern," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, pp. 1–13, 2009.

[4] C. D. Byron, A. M. Kiefer, J. Thomas, S. Patel, A. Jenkins, A. L. Fratino, and T. Anderson, "The authentication and repatriation of a ceremonial tsantsa to its country of origin (ecuador)," *Heritage Science*, vol. 9, no. 1, pp. 1–13, 2021.

[5] S. R. Borra, G. J. Reddy, and E. S. Reddy, "A broad survey on fingerprint recognition systems," in *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*. IEEE, 2016, pp. 1428–1434.

[6] D. Peralta, M. Galar, I. Triguero, D. Paternain, S. García, E. Barrenechea, J. M. Benítez, H. Bustince, and F. Herrera, "A survey on fingerprint minutiae-based local matching for verification and identification: Taxonomy and experimental evaluation," *Information Sciences*, vol. 315, pp. 67–87, 2015.

[7] M. Sharma, "Fingerprint biometric system: a survey," *International Journal of Computer Science & Engineering Technology (IJCSET)*, vol. 5, no. 7, pp. 743–747, 2014.

[8] K. Delac and M. Grgic, "A survey of biometric recognition methods," in *Proceedings. Elmar-2004. 46th International Symposium on Electronics in Marine*. IEEE, 2004, pp. 184–193.

[9] B. Sugandi, I. Dewita, and R. P. Hudjajanto, "Face recognition based on pca and neural network," in *2019 2nd International Conference on Applied Engineering (ICAE)*. IEEE, 2019, pp. 1–5.

[10] P. Anusha, K. L. Prasad, G. R. Kumar, E. L. Lydia, and V. S. Parvathy, "Facial detection implementation using principal component analysis (pca)," *Journal of Critical Reviews*, vol. 7, no. 10, pp. 1863–1872, 2020.

[11] S. Shen, C. Zhang, R. Xiao, W. He, and N. Zhang, "Research on face recognition based on pca and fld," in *2019 3rd International Conference on Data Science and Business Analytics (ICDSBA)*. IEEE, 2019, pp. 479–481.

[12] M. K. Luaibi and F. G. Mohammed, "Facial recognition based on dwt–hog–pca features with mlp classifier," *Journal of Southwest Jiaotong University*, vol. 54, no. 6, 2019.

[13] S. K. Bhattacharyya and K. Rahul, "Face recognition by linear discriminant analysis," *International Journal of Communication Network Security*, vol. 2, no. 2, pp. 31–35, 2013.

[14] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Face recognition using lda-based algorithms," *IEEE Transactions on Neural networks*, vol. 14, no. 1, pp. 195–200, 2003.

[15] J. Wei, Z. Jian-Qi, and Z. Xiang, "Face recognition method based on support vector machine and particle swarm optimization," *Expert Systems with Applications*, vol. 38, no. 4, pp. 4390–4393, 2011.

[16] J. A. C. Moreano, N. L. S. Palomino, and A. C. L. Casa, "Facial recognition techniques using svm: A comparative analysis," *Enfoque UTE*, vol. 10, no. 3, pp. 98–111, 2019.

[17] I. Kukenys and B. McCane, "Support vector machines for human face detection," in *Proceedings of the New Zealand computer science research student conference*. Citeseer, 2008, pp. 226–229.

[18] M. Javed, *Building a Facial Recognition Model using PCA & SVM Algorithms*. Medium, 2020.

[19] Y. M. Riyazuddin, S. MahaboobBasha, J. K. Reddy, and S. Naseera-Banu, "Effective usage of support vector machine in face detection," *International Journal of Engineering and Advanced Technology*, vol. 9, no. 3, pp. 1336–1340, 2020.

[20] R. J. Hassan, A. M. Abdulazeez *et al.*, "Deep learning convolutional neural network for face recognition: A review," *International Journal of Science and Business*, vol. 5, no. 2, pp. 114–127, 2021.

[21] W. Liu, L. Zhou, and J. Chen, "Face recognition based on lightweight convolutional neural networks," *Information*, vol. 12, no. 5, p. 191, 2021.

[22] M. W. F. F. H. Alhadi, "Support vector machines for human face detection," in *IASTED International Conference on Computational Intelligence*. Citeseer, 2005, pp. 4–6.

[23] P. Corcoran and C. Iancu, *Hidden Markov Models in automatic face recognition-A review*. IntechOpen, 2011.

[24] L. Wiskott, N. Krüger, N. Kuiger, and C. Von Der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 19, no. 7, pp. 775–779, 1997.

[25] ——, "Face recognition by elastic bunch graph matching," *Intelligent Biometric Techniques in Fingerprint and Face Recognition*, vol. Chapter 11, pp. 355–396, 1999.

[26] S. Jaiswal, "Comparison between face recognition algorithm-eigenfaces, fisherfaces and elastic bunch graph matching," *Journal of global research in computer science*, vol. 2, no. 7, pp. 187–193, 2011.

[27] H. Cho, R. Roberts, B. Jung, O. Choi, and S. Moon, "An efficient hybrid face recognition algorithm using pca and gabor wavelets," *International Journal of Advanced Robotic Systems*, vol. 11, no. 4, p. 59, 2014.

[28] S. Shan, P. Yang, X. Chen, and W. Gao, "Adaboost gabor fisher classifier for face recognition," in *International Workshop on Analysis and Modeling of Faces and Gestures*. Springer, 2005, pp. 279–292.

[29] H. Cho, R. Roberts, B. Jung, O. Choi, and S. Moon, "An efficient hybrid face recognition algorithm using pca and gabor wavelets," *International Journal of Advanced Robotic Systems*, vol. 11, no. 4, p. 59, 2014.

[30] V. R.-A. J. O.-G. Fernando Alonso-Fernandez, Pedro Tome-Gonzalez, "Iris recognition based on sift features," *International Conference on Biometrics, Identity and Security (BIdS)*, pp. 1–7, 2009.

[31] K. Wang and A. Kumar, "Toward more accurate iris recognition using dilated residual features," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 12, pp. 3233–3245, 2019.

[32] S. Adamović, V. Miškovic, N. Maček, M. Milosavljević, M. Šarac, M. Saračević, and M. Gnjatović, "An efficient novel approach for iris recognition based on stylometric features and machine learning techniques," *Future Generation Computer Systems*, vol. 107, pp. 144–157, 2020.

[33] C. Simon, "A new scientific method of identification," *New York state journal of medicine*, vol. 35, no. 18, pp. 901–906, 1935.

[34] M. P. S. Sukumaran, "Retina recognition based on fractal dimension," *International Journal of Computer Science and Network Security (IJCSNS)*, pp. 1–7, 2009.

[35] S. A. Tuama and L. E. George, "Retina recognition based on texture analysis: Building a system for individual recognition based on vascular retina pattern," *LAMBERT Academic Publishing (LAP)*, pp. 1–7, 2016.

[36] J. Liebers, M. Abdelaziz, L. Mecke, A. Saad, J. Auda, U. Gruenefeld, F. Alt, and S. Schneegass, "Understanding user identification in virtual reality through behavioral biometrics and the effect of body normalization," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–11.

[37] I. Stylios, S. Kokolakis, O. Thanou, and S. Chatzis, "Key factors driving the adoption of behavioral biometrics and continuous authentication technology: an empirical research," *Information & Computer Security*, 2022.

[38] R. S. Gaines, W. Lisowski, S. J. Press, and N. Shapiro, *Authentication by keystroke timing: Some preliminary results*. Rand Corp Santa Monica CA, 1980.

[39] S. Bleha, C. Slivinsky, and B. Hussien, "Computer-access security systems using keystroke dynamics," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 12, no. 12, pp. 1217–1222, 1990.

[40] W. G. De Ru and J. H. Eloff, "Enhanced password authentication through fuzzy logic," *IEEE Expert*, vol. 12, no. 6, pp. 38–45, 1997.

[41] T. Ord and S. M. Furnell, "User authentication for keypad-based devices using keystroke analysis," in *Proceedings of the second international network conference (INC-2000)*, 2000, pp. 263–272.

[42] M. S. Obaidat and B. Sadoun, "Verification of computer users using keystroke dynamics," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 27, no. 2, pp. 261–269, 1997.

[43] F. Monrose and A. D. Rubin, "Keystroke dynamics as a biometric for authentication," *Future Generation computer systems*, vol. 16, no. 4, pp. 351–359, 2000.

[44] J. A. Robinson, V. Liang, J. M. Chambers, and C. L. MacKenzie, "Computer user verification using login string keystroke dynamics," *IEEE transactions on systems, man, and cybernetics-part a: systems and humans*, vol. 28, no. 2, pp. 236–241, 1998.

[45] R. S. Kashi, W. Turin, and W. L. Nelson, "On-line handwritten signature verification using stroke direction coding," *Optical Engineering*, vol. 35, no. 9, pp. 2526–2533, 1996.

[46] A. K. Jain, F. D. Griess, and S. D. Connell, "On-line signature verification," *Pattern recognition*, vol. 35, no. 12, pp. 2963–2972, 2002.

[47] B. Wirtz, "Stroke-based time warping for signature verification," in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, vol. 1. IEEE, 1995, pp. 179–182.

[48] I. G. S. (IGS), "chemin de polytechnique montréal (québec)," http://www.scribens.polymtl.ca/, accessed: 2021-11-30.

[49] K. Tang, R. Kamoua, V. Sutan, O. Farooq, G. Eng, W. C. Chu, and G. Hou, "Speech recognition technology for disabilities education," *Journal of Educational Technology Systems*, vol. 33, pp. 173 – 184, 2004.

[50] J. Noyes and C. Frankish, "Speech recognition technology for individuals with disabilities," *Augmentative and Alternative Communication*, vol. 8, no. 4, pp. 297–303, 1992. [Online]. Available: https://doi.org/10.1080/07434619212331276333

[51] J. Schönbächler, "Le traitement de la parole pour les personnes handicapées," *Travail de séminaire*, 2003.

[52] R. Simpson and S. Levine, "Voice control of a powered wheelchair," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 10, no. 2, pp. 122–125, 2002.

[53] F. Alonso-Fernandez, P. Tome-Gonzalez, V. Ruiz-Albacete, and J. Ortega-Garcia, "Iris recognition based on sift features," in *2009 First IEEE International Conference on Biometrics, Identity and Security (BIdS)*. IEEE, 2009, pp. 1–8.

[54] A. Poursaberi and B. N. Araabi, "Iris recognition for partially occluded images: methodology and sensitivity analysis," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, pp. 1–12, 2006.

[55] F. Fuentes-Hurtado, V. Naranjo, J. A. Diego-Mas, and M. Alcañiz, "A hybrid method for accurate iris segmentation on at-a-distance visible-wavelength images," *EURASIP Journal on Image and Video Processing*, vol. 2019, no. 1, pp. 1–14, 2019.

[56] V. Nazmdeh, S. Mortazavi, D. Tajeddin, H. Nazmdeh, and M. M. Asem, "Iris recognition; from classic to modern approaches," in *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2019, pp. 0981–0988.

[57] S. Sukumaran and M. Punithavalli, "Retina recognition based on fractal dimension," *IJCSNS Int J Comput Sci and Netw Secur*, vol. 9, no. 10, pp. 66–7, 2009.

[58] S. A. Tuama and L. E. George, *Retina Recognition Based on Texture Analysis: Building a system for individual recognition based on vascular retina pattern*. LAP LAMBERT Academic Publishing, 2016.

Protecting User Preference in Ranked Queries

Rong Tang
College of Cyber Security
Jinan University, Guangzhou China

Xinyu Yu
College of Cyber Security
Jinan University, Guangzhou China

***Abstract*—Protecting data privacy is of extremely importance for users who outsource their data to a third-party in cloud computing. Although there exist plenty of research work on data privacy protection, the problem of protecting user's preference information has received less attention. In this paper, we consider the problem of how to prevent user preference information leakage from the third-party when processing ranked queries. We propose two algorithms to solve the problem, where the first one is based on distortion of preference vector and the second one is based on homomorphic encryption. We conduct extensive experiments to verify effectiveness of our approaches.**

***Keywords*—Preference privacy; distortion; homomorphic encryption**

I. Introduction

Ranked queries are useful in many real world applications for users to express their tastes in queries, such that the returned query results are the most preferable ones than the other records in database. Currently, there are two kinds of approaches to capturing user's preferences, that is, the quantitative approach [1] and qualitative approach [2]. In the paper, we focus on quantitative approach for modeling user preference.

Given a relational database D = $\{A_1,\ A_2,...,\ A_d\}$ with numerical attribute A_i, i = 1,..., d, i.e., a linear function q = $\omega_1 * A_1 + \omega_2 * A_2 + ... + \omega_d * A_d$, where ω_1, ω_2,..., ω_d are non-negative weights given by the user as preferences toward the corresponding attributes, and $\omega_1 + \omega_2 + ... + \omega_d = 1$. Upon receiving a user ranked query q, for each record $r_i \in D$ the server computes its preference score as $Score(i) = \sum_{j=1}^{d} \omega_j * r_i.A_j$, j = 1,..., d, and returns to the user the top k records with the highest scores.

Although there are extensive research work on processing ranked queries, the privacy issue of user preferences so far has not received attention. There is no doubt that user's preference is a strong and direct link to his/her identity, which the user wants to keep in private, otherwise the information may be utilized by an adversary to against he/she. Thus, one may want to hide his/her preference embedded in the ranked query in order to protect privacy. To illustrate, we give two examples below.

Example 1. Consider a user looking for a second-hand car at a website running by a car dealer. The dealer maintains a used car database, recording for all the cars the features such as Make, Model, Year, MPG (miles per gallon), Mileage, and Price, etc. The user may care more about some of the features, such as MPG, mileage, and Year, and want to find a suitable car for him with the lowest price. By collecting and observing user's queries with preferred features, the dealer knows that the user strongly favors MPG. So the dealer may want to increase a bit the prices of the cars with favorable MPG, such that he is expected to profit more from the user. To have the edge over the car dealer while making a deal, the user wishes to hide his preferences during the querying process.

Example 2. Consider a customer searching for financial information of NYSE-listed companies, through websites such as Yahoo! finance or Google finance. By giving preferences on attributes such as cash flow, P/E (price-to-earnings) ratio, ROA (return on assets) ratio, and debt ratio, the customer intends to search through query interface the favorable companies, based on which he/she may make buy or sale decision for his/her investment portfolio management. On the other hand, a curious adversary may sniff the search results of the customer at the server side and extract the preferable attributes, which may be used to infer the possible investment of the customer. Since customer's portfolio information is critical for him/her to stand up to the fierce business competition, he/she may strongly oppose the exposure of his/her preferences to anyone else.

Problem Statement: Our model includes the user and the service provider (the server). The server maintains a database D and processes users' queries with difference preference on the attributes of D. We assume that the server is semi-honest, that is, the server correctly performs the query processing and returns the results, but he is curious and tries to find out the user's preferences. Each user composes his ranked query as a weight vector f = $\{\omega_1,\ \omega_2,...,\ \omega_d\}$ on attributes of D, and submits it to the server for processing. Our objective is to prevent the server from knowing the exact preferences, i.e., the weight vector of the user, without deteriorating the query processing efficiency and accuracy.

The contributions of this paper are as follows:

- We consider the problem of protecting user's preferences in ranked queries, which is of great importance in many real world applications.
- We propose a simple strategy to distort user's true preferences, and give a test to the strategy.
- To strengthen the privacy level, we devise another algorithm based on homomorphic encryption to protect user's preference.
- We conduct extensive experiments to verify effectiveness of the proposed approaches.

II. Background

A. Preliminaries

Ranked query. Ranked query [3] is very useful in many real-world applications. It is a powerful technology to simulate users' personalized information needs, which has attracted great attention of researchers all over the world. It allows users to express their preferences for queries and specify the specific weight of each query limit, so that the returned query results meet the needs of users better than other data records in the database. At present, there are two methods to describe user preferences, namely qualitative method and quantitative method.

The qualitative method [2] usually uses the preference formula to evaluate which data tuple is more favorable over the others, i.e. determining the partial order relationship between tuples. The quantitative method [1] defines a preference function to express the user's preference for different data attributes, and then finds the data records that best meet the user's needs from the database according to the preference function, that is, measured by specific values. For example, given a hotel dataset containing three dimensions $D = \{d_1, d_2, d_3\}$, where d_1 represents price, d_2 represents score, d_3 represents distance, preference function is defined as $f = \omega_1 * d_1 + \omega_2 * d_2 + \omega_3 * d_3$ according to user preference $f = \{\omega_1, \omega_2, \omega_3\}$. The highest score is the most desirable.

In practical applications, ranked query is often directly related to other problems, such as skyline computing [4] and top-k query [5], [6]. Combining multiple queries can usually get the results you want most quickly.

Top-k query. Top-k query [7], [8] refers to returning the best k data records according to an objective function. It is widely used in many fields, such as e-commerce, recommendation system, search engine and so on. When the Top-k query is associated with the preference problem, it will calculate the score of each data record according to the preference function given by the user, and then return the k objects with the smallest (or largest) scores.

The concept of Top-k query has been around for many years and is widely used in real life. So far, many algorithms for Top-k query have been proposed, which can be roughly divided into three categories. The first is index based algorithm [9], [10]. Its main idea is to divide the whole data set into multiple layers according to the division rules, then index and mark each layer, and finally retrieve layer by layer according to the index order to return the best k results, such as Onion algorithm [11]. The second is the Top-k query algorithm based on view [12]. This kind of algorithm first calculates the scores corresponding to each tuple according to the preference function provided by the user and arranges them in order. The view contains the identifier and score of tuples. After these preparations, it finally returns the Top-k query results. The third is the Top-k query algorithm based on ordered list [13], [14], which is realized by using multiple column files.

B. Related Work

Skyline query. Skyline query problem [4], [15], [16] is a popular technology for processing user preferences and Top-k query. It is used to select a series of objects that meet user preferences and are not dominated by other objects. Given a dataset D with d-dimensions $\{d_1, d_2,..., d_d\}$ and n $objects\{A_1, A_2,..., A_d\}$, where $A_i.d_j$ denotes the j-th dimension value of object A_i. The definition of dominance and skyline are as follows:

Dominance: An object $A_i \in D$ is said to dominate another object $A_j \in D$, denoted as $A_i \prec A_j$, if $A_i.d_r \leq A_j.d_r(1 \leq r \leq d)$ for all d dimensions and $A_i.d_t < A_j.d_t(1 \leq t \leq d)$ for at least one attribute. We call such A_i as *dominant object* and such A_j as *dominated object* between A_i and A_j.

Skyline: An object A_i is said to be a skyline object of D, if and only if there is no such object $A_j \in D(j \neq i)$ that dominates A_i. The skyline of D is the set of skyline objects in D.

At present, there are mainly two kinds of methods for skyline query processing. The first kind of methods do not need to preprocess the data set, but retrieve the query by scanning the whole database at least once, such as block nested loop (BNL), divide and conquer, etc. BNL algorithm [17] adopts the most straightforward method, that is, a point p is compared with each other point to decide whether it is dominated by other points, so as to determine whether the point is part of the skyline. Divide and conquer algorithm [17] divides the universe into several regions, calculates the skyline in each region, and produces the final skyline from the regional skylines. Therefore, the performance of this kind of methods is low because of scanning the whole database.

The second [18] is to reduce the query cost by using the index structure, such as nearest neighbor(NN), branch and bound skyline(BBS), etc. NN and BBS find the skyline by using an R-tree. NN algorithm [19] divides the data space iteratively based on the nearest object in the space, and prunes the dominant object quickly and effectively. However, BBS algorithm [20] uses heap to realize progressive search without redundant query in subspace. Obviously, the difference is that NN issues multiple NN queries [21], whereas BBS performs only a single traversal of the tree. It has been proved [22] that BBS is I/O optimal; that is, it accesses the least number of disk pages among all algorithms based on R-trees (including NN).

Homomorphic encryption(HE). Homomorphic cryptographic system [23], [24] is a public-key cryptosystem that can provide user with the ability to directly perform algebraic operations on ciphertext without decrypting the ciphertext.

Given two messages m_1 and m_2, suppose a homomorphic cryptographic system encrypts them, using public key $\mathcal{PK}$, to ciphertexts $C_1 = E(\mathcal{PK}, m_1)$ and $C_2 = E(\mathcal{PK}, m_2)$. Without knowing the corresponding secret key, one can compute $E(\mathcal{PK}, m_1 + m_2)$, i.e., the ciphertext of the addition of m_1 and m_2, by simply multiplying the two ciphertexts $C_1 \times C_2$. This property is called additive homomorphicsm. Similarly, a crytographic system is multiplicatively homomorphic if one can derive $E(\mathcal{PK}, m_1 \times m_2)$ from C_1 and C_2 directly.

The concept of homomorphic encryption [25], [26] is proposed to directly perform operations in the encryption domain, that is, the results obtained by decrypting the operations performed in the ciphertext domain are consistent with those obtained by performing the same operations in the plaintext domain. However, most existing homomorphic encryption schemes only support accurate computing operations in some discrete spaces, so these schemes can not be applied to tasks requiring floating-point or real number computing. For example, in the bit-wise encryption scheme, the integer is first converted into binary, and then encrypted by bit. The addition and multiplication operations are also based on bits. This scheme cannot be applied to floating-point numbers. For the word-wise encryption scheme, multiple numbers can be encrypted in a single ciphertext, but the rounding operation is difficult to evaluate because it is not expressed as a decimal polynomial.

After Regev [27] introduced the learning with errors (LWE) problem, approximate homomorphic encryption was proposed one after another. Since the key of the method based on LWE problem is realized through matrix, its efficiency will decrease rapidly with the increase of security parameters. Then the ring-LWE problem [28] is proposed. The key of the encryption scheme based on this problem is expressed by several polynomials, which greatly reduces the size of the key and speeds up the encryption and decryption operations. The scheme based on RLWE problem include BFV [26], BGV [29] and HEAAN [30]. BFV and BGV encryption scheme only support accurate computing operations on integers.

However, HEAAN scheme can encrypt floating-point numbers. And the goal of this scheme is efficient approximate calculation on HE. Its main idea is to add a noise to the plaintext number that can reflect important information, so that the addition and multiplication operations of encrypted messages can be approximately calculated. In HEAAN encryption scheme, its decryption structure of the form $\langle \mathbf{c}, sk\rangle = m + e$ (mod q) where e is a small error inserted to guarantee the security. In addition, HEAAN also provides a rescaling operation to remove the error of the least significant bit, which ensures that the length of the error bit increases linearly in proportion to the number of levels consumed, rather than exponentially. The efficiency of HEAAN [30] has been proved in many practical applications, and is still being improved by better bootstrapping algorithms [31], [32]. Therefore, considering that the data used in this paper are floating-point numbers, HEAAN homomorphic encryption scheme will be adopted.

Approximate Algorithm for Comparison Function. Logical operation has always been a difficult point in HE. Bit-wise FHEs encrypt data in a bit-by-bit manner. They support fast logical operations, such as comparison. But they can not support floating point encryption. On the contrary, word-wise FHEs, which store messages as their word-sized numbers, support high-speed arithmetic operations between messages. Therefore, in order to calculate the comparison function, an approximate form of the comparison function is proposed by using polynomials.

Cheon et al. [33] first proposed a new identity $comp(a, b) = \lim_{k\to\infty} a^k/(a^k + b^k)$, and showed that the identity can be computed by an iterative algorithm. Because of the iterative feature, this algorithm is slow in HE implementation. Then, they proposed a new comparison methods SIMD [34], using composite polynomial approximation on the sign function, which is equivalent to the comparison function. That is, repeated compositions of (2n+1)-degree polynomial $f_n(x)$ and $g_n(x)$ output the approximate value of the sign function.

We denote the approximate comparison for two inputs x,y by $(x > y)$ or $(y < x)$. According to the conclusion of [34], given iteration numbers d_f and d_g, $(x > y)$ is computed as follows

$$(x > y) := (f_n^{d_f} \circ g_n^{d_g}(x - y) + 1)/2 \quad (1)$$

TABLE I. An Example Database

Record	A_1	A_2
r_1	1.4	4.4
r_2	2.5	2.9
r_3	3.5	1.0
r_4	3.3	4.1
r_5	4.2	2.8
r_6	1.9	2.3

Here f_d means $f \circ f \circ \cdots \circ f$, i.e., the operation is performed d times.

III. Our Approaches for Preference Protection

Ranked queries are complementary to traditional SQL query semantics. The preferences expressed by the user are considered as *soft constraints* of the queries, whereas the hit-or-miss query conditions, e.g., $\leq$, $>$, and $\neq$, are known as *hard constraints* [1], [2], [35]. To evaluate a ranked query q, the server computes the sum of the linear combination of attributes of the records, based on the user preference vector f = $\{\omega_1$, ω_2,.., $\omega_d\}$(note we use vector and weight interchangeably), then returns to the user the top-k objects with the highest sum.

Table I gives an example of a toy database with 2 attributes A_1 and A_2. Suppose the user expresses in a query his preference as f = {0.4, 0.6}, and wants to pull top 3 favorable records from the server. Starting from the first record r_1, the server computes its preference score *score*(r_1) = 0.4 $*$ 1.4 + 0.6 $*$ 4.4 = 3.2, and the scores of the rest records. Among the computed scores {3.2, 2.7, 2, 3.8, 3.4, 2.1}, r_1, r_4, and r_5 are the top 3 records with highest scores and they are returned as result to the user. As this example shows, there is no protection for user queries, which means that the server can easily obtain the preference information of a user. In this section, we propose two approaches to prevent leakage of user preference, which are described below.

A. The First Approach

Our first approach is called PD, which is based on preference distortion. So far in the model we assume that there is no encryption involved, i.e., the user queries and database content are all in plaintext. We defer discussion of the case in which encryption is employed in the next section.

Our strategy to protect user's true preference is simple, and goes as follows. Instead of directly submitting to the server the true preference vector f, the user distorts f by randomly adding to or subtracting from the components of f a small number, then sends the modified preference vector to the server. Specifically, given a preference vector f = $\{\omega_1$, ω_2,.., $\omega_d\}$, we transform f into a new vector f' = $\{\omega'_1$, ω'_2,.., $\omega'_d\}$, such that $\omega'_i \in [0, 1]$ and $\sum_i^d \omega'_i = 1$. Suppose the increment/decrement added to component ω_i of f is δ_i, based on the relation $\omega'_i = \omega_i + \delta_i$ we can easily verify the following equation

$$\delta_1 + \delta_2 + ... + \delta_d = 0 \quad (2)$$

where $-\omega_i \leq \delta_i \leq 1 - \omega_i$.

We outline the preference distortion in Algorithm 1. We generate d random numbers, and normalize them to maintain the convex property. These random numbers are sorted in descending order, such that the relative preference relation among attributes is maintained. For example, we are given a vector *f* = {0.3, 0.5, 0.2}, and the generated random numbers are *R* = {0.4, 0.2, 0.8}. After normalization and sorting, we get *R* = {0.6, 0.3, 0.1}, from which we get the distorted vector *f'* = {0.3, 0.6, 0.1}, and the increment/decrement vector Δ = {0, 0.1, -0.1}. It is obvious that after distortion, the new vector still preserves the relative preference relation among attributes as in the original vector, but the actual value of preference weights have been changed.

Algorithm 1: Preference Distortion

input : Preference vector *f* = $\{\omega_1, \omega_2,.., \omega_d\}$
output: Distorted preference vector *f'*
R' = Δ = *f'* = $\emptyset$;
Generate a set R of d random numbers;
for $\forall rand_i \in$ R **do**
  $rand'_i = \frac{rand_i}{\sum_j rand_j}$;
  Insert $rand'_i$ into *R'*;
end
Sort *R'* in descending order;
count = len(*f*);
j = 1;
while *count* > 0 **do**
  i = the index of the greatest weight in f;
  $\omega'_i = rand'_j$;
  $\delta_i = \omega'_i - \omega_i$;
  f[index] = 0;
  Insert ω'_i and δ_i into *f'* and Δ,respectively;
  count = count - 1;
  j = j + 1;
end
Return *f'*;

Having obtained the distorted vector *f'*, the user sends it to the server for processing, where at the server side the query processing is just as the same as before and the query results and scores are sent back to the user. Since the scores do not reflect the true values for the original preference vector, the user has to revert the scores returned by the server. Consider the score of a record r_i with respect to *f'*, $score(i) = f' \cdot r_i$, which can be represented as

$$(\omega_1 + \delta_1 \times r_i.a_1 + \omega_2 + \delta_2 \times r_i.a_2 + ... + \omega_d + \delta_d \times r_i.a_d)$$

After re-arrangement of the above formula we have

$$(\omega_1 \times r_i.a_1 + ... + \omega_d \times r_i.a_d) + (\delta_1 \times r_i.a_1 + ... + \delta_d \times r_i.a_d)$$

It is clear that the first parenthesized part is the correct score with respect to *f*, and the second part is the noise artificially added by the distorted vector *f'*. Thus, the user computes the noise for each r_i of the return top *k* records by multiplying Δ with r_i, and then subtracting the noise from the returned score. After the reverting procedure, the user re-orders the resulting records according to the restored scores. We summarize the reverting procedure in Algorithm 2.

The algorithm of preference distortion is simple and efficient for protecting user preferences, however it may introduce false negative and false positive query result, as we will show in the next section.

Algorithm 2: Score Reverting

input : Query result set *S*, and the set *Score*
output: The restored score set *Score'*
Score $= \emptyset$;
for Score$(i) \in$ Score **do**
  Noise $= \delta_1 \times r_i.a_1 + ... + \delta_d \times r_i.a_d$;
  Score'$(i) =$ *Score*$(i) -$ *Noise*;
  Insert *Score'*(i) into *Score'*;
end
Sort the records in *S* according to *Score'*;
Return *Score'*;

Security Discussion. When the user sends the preference vector to the server for query, even if the preference vector is disturbed, the server may predict the user's real vector according to the returned top k result, that is, the higher the precision, the higher the probability of being predicted. It is known that the weight of the user preference vector is a decimal between [0,1], and the server cannot know the specific decimal places of each weight in the real vector. Obviously, the top k result corresponding to the vector in a small range is the same. Therefore, when the server predicts the user preference vector according to the top k query results, there are countless possibilities in the case of uncertain decimal places of each weight in the vector. Especially when the dimension is higher and the number of decimal places is more, there is a lower probability of being predicted by the server.

B. The Second Approach

To further strengthen the privacy level of user's preferences, in this section we propose the second approach called HE, which is based on homomorphic encryption.

Considering that the attributes of a record are floating-point numbers, and HEAAN [30] scheme is selected as the encryption scheme. Assume the secrete/public key pair of HEAAN homomorphic encryption system is $\langle \mathcal{SK}, \mathcal{PK} \rangle$. The user encrypts his plaintext preference weights *f* = $\{\omega_1, \omega_2,..., \omega_d\}$ by using the public key $\mathcal{PK}$, which gives the encrypted weights *E(f)* = {*E*(ω_1), *E*(ω_2),..., *E*(ω_d)}(see Algorithm 3).

Algorithm 3: Preference Encryption

input : Preference vector *f*,the public key of CKKS encryption system $\mathcal{PK}$
output: Encrypted preference vector *E(f)*
E(f) $= \emptyset$;
for $\forall \omega_i \in$ f **do**
  E(ω_i) = $\langle \omega_i, \mathcal{PK} \rangle$;
  Insert *E*(ω_i) into *E(f)*;
end
Return *E(f)*;

After the preference vector encryption operation is completed, the next step is to perform the linear weighted summation operation on the data records in the data set, and select the k results with the top scores through comparison.

The comparison operation of ciphertext is realized by SIMD [34] scheme. In order to reduce the computational consumption of encryption, we consider that the server first performs K-Skyband operation on the dataset *D*, and then only calculates and compares the data records in the K-Skyband dataset. Similar to K nearest-neighbor queries, a K-Skyband [22] query reports the set of points which are dominated by at most $K-1$ points.The definition of K-Skyband is as follows.

K-Skyband: An object $A_i \in D$ is said to be a K-Skyband object of *D*, if A_i is dominated by at most $K-1$ objects in *D*.

As can be seen from the definition, K-Skyband [22] is a variant of skyline, in which K can be regarded as the thickness of skyline; the case $K = 0$ corresponds to a conventional skyline. Thus, we implement K-Skyband by modifying the BBS algorithm. Based on this, the specific process of query is shown in algorithm 4.

Algorithm 4: Query Evaluation

input : Encrypted preference vector *E(f)*
output: Result records and their encrypted scores

```
1 Score = ∅;
2 Compute the set Band_k of k-skyband points of D;
3 for ∀r_i ∈ Band_k do
4 |   E(score(i)) = Σ_{∀a_j∈r_i} E(ω_j) * a_j;
5 |   Insert E(score(i)) into Score,and compare;
6 end
7 Send Score and Band_k to the user;
```

Security Discussion. This scheme encrypts the preference vector. The IND-CPA security nature of FHE scheme ensures that any opponent with the result homomorphic operation between ciphertext and ciphertext cannot extract any information of the message in the ciphertext, our HE method is secure based on the security of FHE, because the server can only access the ciphertext. There is no information leak of user preference.

IV. Experimental Results

In this section we conduct extensive experiments to evaluate performance of the proposed two approaches, and all experiments are conducted on a PC running Ubuntu 18.04.2 LTS. We discuss experiment settings below.

Dataset. We use synthetic datasets of three data distributions, namely, independent (uniform), correlation and anti-correlation, with dimensionality *d* varying in the range [2,5] and cardinality *N* in the range [1k,20k], respectively.

Performance Indicators. For our first approach PD, since it is an approximate method, we focus on precision and recall indicator, i.e., how many real results can be found for PD. On the other hand, for our second approach HE, since it is an exact method, i.e., HE can find all the correct results, we employ running time as the performance indicator for the HE approach. The precision and recall indicator is defined as

$$precision = \frac{TP}{TP + FP} \tag{3}$$

$$recall = \frac{TP}{TP + FN} \tag{4}$$

where TP indicates that the number of positive query results retrieved, FP indicates the number of points which the retrieved positive ones are actually negative, FN indicates the number of points which it has not been retrieved but they are actually positive. Each precision and each recall result is the average of 10 trails.

Suppose $Q(x)$ is defined as the number of data points in area x. Based on the premise of Top-k preference query in this paper, it is easy to obtain

$$Q(TP) + Q(FP) = k$$

and

$$Q(TP) + Q(FN) = k$$

then $Q(FP)$ = $Q(FN)$. Therefore, the results of precision and recall are the same.

A. Effect of Dimensionality d

We vary data dimensionality d from 2 to 5. As shown in Fig. 1, the cardinality N is $10k$ and K is 50. And it can be seen that the average precision and recall results of datasets with independent (uniform), correlation and anti-correlation distribution are basically stable and were not be affected by the dimensionality. It is obvious that the average precision and recall results of correlation data set are the highest, while that of anti-correlation are poor.

B. Effect of K

In order to study the effect of K we carried out experiments when $K = 10, 20, 50, 100$. Fig. 2 shows the precision and recall corresponding to each case. Obviously, the query results are not affected by the K value.

C. Effect of Cardinality N

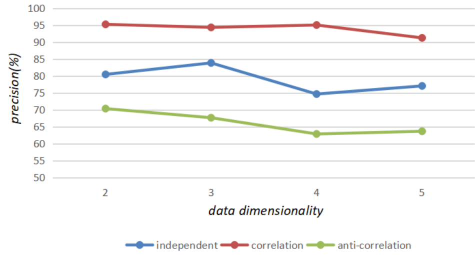
Fig. 3 shows the precision and recall results when the cardinality *N* in the range [1k,20k]. It can be clearly seen that the precision and recall values of the three types of datasets tend to be stable.

In conclusion, our proposed distortion algorithm has little relationship with the size of K, data dimension and data set cardinality. Thus, the disturbance scheme is practical and will not be affected by other factors. In addition, it can be clearly seen that it has a good precision and recall result in correlation datasets but poor in anti-correlation datasets.

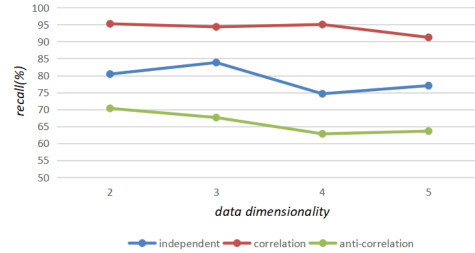
D. Discussion

From the above experimental results, it can be obviously seen that precisions are float up and down. That is due to the existence of false negative and false positive points, and the number of them is uncertain.

For ease of discussion, we use the example database in Table I, and assume the data space range for attribute A_1, A_2 are $x = [0, U_x]$, $y = [0, U_y]$, respectively, and $k = 3$ for the returned top k results. Now consider an original vector *f* = $\{\omega_1, \omega_2\}$ and the corresponding transformed vector *f'* = $\{\omega_1', \omega_2'\}$. After examining each vector against the records in *D*, the qualified records (represented as data points in 2-D space) are highlighted in Fig. 4(A). Specifically, the top 3 records with

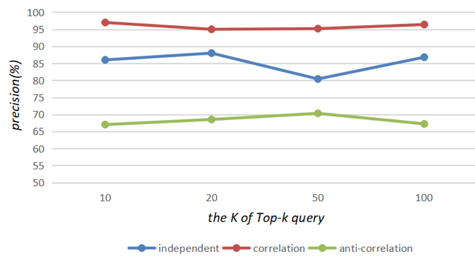


(a) precision

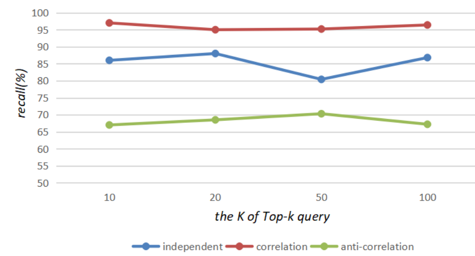


(b) recall

Fig. 1. Precision & Recall on a Dataset with Top-50 Query, $10k$ Records and Varying Dimensionality d.

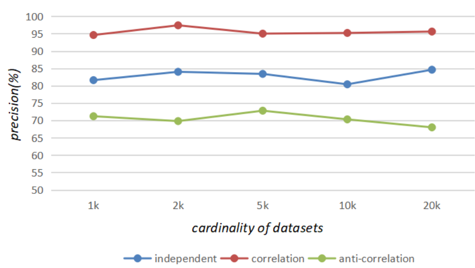


(a) precision

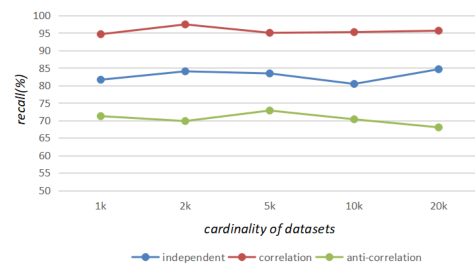


(b) recall

Fig. 2. Precision & Recall on a Dataset of $10k$ Records in $3D$ Versus K.



(a) precision



(b) recall

Fig. 3. Precision & Recall on a Dataset with Top-50 Query and Varying Cardinality N.

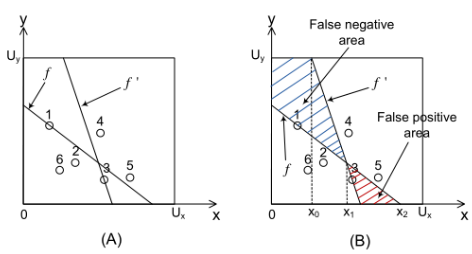


Fig. 4. Original and Distorted Ranked Queries.

respect to the original vector f include r_4, r_5, r_1, whereas the answers for distorted vector f' are r_4, r_5 and r_3. There is a false positive result, r_3, and a false negative r_1, due to the weight distortion.

To evaluate the original and distorted preference vector, one has to quantify the number of false positive and negative results. As depicted in Fig.4(B), it is clear that the larger striped area (denoted by FN) contains false negative points, and the smaller striped area (denoted by FP) false positive points. The area of FN and FP can be easily calculated by computing the following integrals.

$$FN = \int_0^{x_0} (U_y - f) dx + \int_{x_0}^{x_1} (f\text{'} - f) dx \tag{5}$$

$$FP = \int_{x_1}^{x_2} (f - f\text{'}) dx \tag{6}$$

Under the assumption of uniform distribution of the data records in D, we can estimate the number of false negative and false positive results as $N * FN$ and $N * FP$ respectively, where N is the cardinality of D. On the other hand, the computation of the FP and FN becomes complicated in the case of high

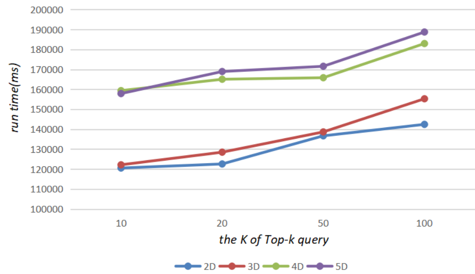


Fig. 5. Running Time on a Dataset by Varying K.

dimensional data. However, one may resort to techniques such as Monte Carlo method to approximate the integral in high-dimensional case. This is a future direction of our work.

E. Performance of HE Approach

For HEAAN, we fix the dimension $N = 2^{17}$, $\Delta = 2^{40}$, and hamming weight $h = 128$ and the secret key is chosen from the ring with ternary secret distribution, i.e., all nonzero coefficients of secret key are ± 1. For SIMD, we employs the approximate comparison method where f_n and g_n are chosen by $n = 3$, where

$$f_3(x) = (35x - 35x^3 + 21x^5 - 5x^7)/2^4 \quad (7)$$

$$g_3(x) = (4589x - 16577x^3 + 25614x^5 - 12860x^7)/2^{10} \quad (8)$$

We conduct experiments with different k and dimensionality d varying in the range [2,5], and report the corresponding running time. Fig. 5 shows the running time when k = 10, 20, 50, 100 with $10k$ objects.

When K in K-Skyband is large, the data to be calculated decreases relatively, but the running time still increases steadily with the increase of dimensionality.

V. Conclusion

In order to protect user preference privacy information, we propose two approaches to solve the problem of user preference privacy protection in ranked queries. The first method, called PD, hides the user's real preference information by introducing perturbation to the user ranked query vector. This scheme will lead to a slight degradation in precision in the query results. Moreover, the experiment results also show that PD achieves the best performance on dataset with correlation distribution, whereas it performs relatively poor on dataset with anti-correlation distribution. Therefore, PD is suitable for real-time ranked query processing scenarios with less accurate requirement for the query result. In order to get exact query result, a homomorphic encryption-based method called HE is proposed to encrypt the preference vector, which enables the third-party server to process ranked query by calculating the ciphertext, so as to fully protect the privacy of user preference.

Acknowledgment

We would like to thank the anonymous reviewers for reviewing this paper. This work is supported in part by National Natural Science Foundation of China (Grant No. 61972177), Science Plan of Yunfu city (Grant No. 2021020302), Guangdong China. Rong Tang and Xinyu Yu contribute equally to this paper.

References

[1] R. Agrawal and E. L. Wimmers, "A framework for expressing and combining preferences," in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 2000, pp. 297–306.

[2] J. Chomicki, "Preference formulas in relational queries," *ACM Transactions on Database Systems (TODS)*, vol. 28, no. 4, pp. 427–466, 2003.

[3] Y. Gao and Q. Liu, "Introduction to preference query analysis and optimization," in *Preference Query Analysis and Optimization*. Springer, 2017, pp. 1–7.

[4] Y. Wang, Z. Shi, J. Wang, L. Sun, and B. Song, "Skyline preference query based on massive and incomplete dataset," *IEEE Access*, vol. 5, pp. 3183–3192, 2017.

[5] P. K. Agarwal, S. Sintos, and A. Steiger, "Efficient indexes for diverse top-k range queries," in *Proceedings of the 39th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*, 2020, pp. 213–227.

[6] T. M. N. Le, M. L. T. Be, and D. H. T. Ngoc, "A processing of top-k aggregate queries on distributed data," in *IOT with Smart Systems*. Springer, 2022, pp. 815–826.

[7] A. Shanbhag, H. Pirk, and S. Madden, "Efficient top-k query processing on massively parallel hardware," in *Proceedings of the 2018 International Conference on Management of Data*, 2018, pp. 1557–1570.

[8] W. Ren, X. Lian, and K. Ghazinour, "Effective and efficient top-k query processing over incomplete data streams," *Information Sciences*, vol. 544, pp. 343–371, 2021.

[9] S.-Y. Ihm, K.-E. Lee, A. Nasridinov, J.-S. Heo, and Y.-H. Park, "Approximate convex skyline: a partitioned layer-based index for efficient processing top-k queries," *Knowledge-Based Systems*, vol. 61, pp. 13–28, 2014.

[10] D. Xin, C. Chen, and J. Han, "Towards robust indexing for ranked queries," in *VLDB*. Citeseer, 2006, pp. 235–246.

[11] Y.-C. Chang, L. Bergman, V. Castelli, C.-S. Li, M.-L. Lo, and J. R. Smith, "The onion technique: Indexing for linear optimization queries," in *Proceedings of the 2000 ACM SIGMOD International Conference on Management of data*, 2000, pp. 391–402.

[12] G. Das, D. Gunopulos, N. Koudas, and D. Tsirogiannis, "Answering top-k queries using views," in *Proceedings of the 32nd international conference on Very large data bases*, 2006, pp. 451–462.

[13] R. Fagin, A. Lotem, and M. Naor, "Optimal aggregation algorithms for middleware," *Journal of computer and system sciences*, vol. 66, no. 4, pp. 614–656, 2003.

[14] Z. Gong, G.-Z. Sun, J. Yuan, and Y. Zhong, "Efficient top-k query algorithms using k-skyband partition," in *International Conference on Scalable Information Systems*. Springer, 2009, pp. 288–305.

[15] K. Mouratidis, K. Li, and B. Tang, "Marrying top-k with skyline queries: Relaxing the preference input while producing output of controllable size," in *Proceedings of the 2021 International Conference on Management of Data*, 2021, pp. 1317–1330.

[16] Z. Zheng, M. Zhang, M. Yu, D. Li, and X. Zhang, "User preference-based data partitioning top-k skyline query processing algorithm," in *2021 IEEE International Conference on Industrial Application of Artificial Intelligence (IAAI)*. IEEE, 2021, p. 436—444.

[17] S. Borzsony, D. Kossmann, and K. Stocker, "The skyline operator," in *Proceedings 17th international conference on data engineering*. IEEE, 2001, pp. 421–430.

[18] N. Sukhwani, V. R. Kagita, V. Kumar, and S. K. Panda, "Efficient computation of top-k skyline objects in data set with uncertain preferences," *International Journal of Data Warehousing and Mining (IJDWM)*, vol. 17, no. 3, pp. 68–80, 2021.

[19] D. Kossmann, F. Ramsak, and S. Rost, "Shooting stars in the sky: An online algorithm for skyline queries," in *VLDB'02: Proceedings of the 28th International Conference on Very Large Databases*. Elsevier, 2002, pp. 275–286.

[20] D. Papadias, Y. Tao, G. Fu, and B. Seeger, "An optimal and progressive algorithm for skyline queries," in *Proceedings of the 2003 ACM SIGMOD international conference on Management of data*, 2003, pp. 467–478.

[21] G. R. Hjaltason and H. Samet, "Distance browsing in spatial databases," *ACM Transactions on Database Systems (TODS)*, vol. 24, no. 2, pp. 265–318, 1999.

[22] D. Papadias, Y. Tao, G. Fu, and B. Seeger, "Progressive skyline computation in database systems," *ACM Transactions on Database Systems (TODS)*, vol. 30, no. 1, pp. 41–82, 2005.

[23] J. H. Cheon and D. Stehlé, "Fully homomophic encryption over the integers revisited," in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 2015, pp. 513–536.

[24] Y. Doröz, Y. Hu, and B. Sunar, "Homomorphic aes evaluation using the modified ltv scheme," *Designs, Codes and Cryptography*, vol. 80, no. 2, pp. 333–358, 2016.

[25] J. W. Bos, K. Lauter, J. Loftus, and M. Naehrig, "Improved security for a ring-based fully homomorphic encryption scheme," in *IMA International Conference on Cryptography and Coding*. Springer, 2013, pp. 45–64.

[26] Z. Brakerski, "Fully homomorphic encryption without modulus switching from classical gapsvp," in *Annual Cryptology Conference*. Springer, 2012, pp. 868–886.

[27] O. Regev, "On lattices, learning with errors, random linear codes, and cryptography," *Journal of the ACM (JACM)*, vol. 56, no. 6, pp. 1–40, 2009.

[28] V. Lyubashevsky, C. Peikert, and O. Regev, "On ideal lattices and learning with errors over rings," in *Annual international conference on the theory and applications of cryptographic techniques*. Springer, 2010, pp. 1–23.

[29] Z. Brakerski, C. Gentry, and V. Vaikuntanathan, "(leveled) fully homomorphic encryption without bootstrapping," *ACM Transactions on Computation Theory (TOCT)*, vol. 6, no. 3, pp. 1–36, 2014.

[30] J. H. Cheon, A. Kim, M. Kim, and Y. Song, "Homomorphic encryption for arithmetic of approximate numbers," in *International Conference on the Theory and Application of Cryptology and Information Security*. Springer, 2017, pp. 409–437.

[31] J. H. Cheon, K. Han, A. Kim, M. Kim, and Y. Song, "Bootstrapping for approximate homomorphic encryption," in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 2018, pp. 360–384.

[32] J.-P. Bossuat, J. R. Troncoso-Pastoriza, and J.-P. Hubaux, "Bootstrapping for approximate homomorphic encryption with negligible failure-probability by using sparse-secret encapsulation," *Cryptology ePrint Archive*, 2022.

[33] J. H. Cheon, D. Kim, D. Kim, H. H. Lee, and K. Lee, "Numerical method for comparison on homomorphically encrypted numbers," in *International Conference on the Theory and Application of Cryptology and Information Security*. Springer, 2019, pp. 415–445.

[34] J. H. Cheon, D. Kim, and D. Kim, "Efficient homomorphic comparison methods with optimal complexity," in *International Conference on the Theory and Application of Cryptology and Information Security*. Springer, 2020, pp. 221–256.

[35] W. Kießling, "Foundations of preferences in database systems," in *VLDB'02: Proceedings of the 28th International Conference on Very Large Databases*. Elsevier, 2002, pp. 311–322.

(IJACSA) International Journal of Advanced Computer Science and Applications,
Vol. 13, No. 4, 2022

Towards Security Awareness of Mobile Applications using Semantic-based Sentiment Analysis

Ahmed Alzhrani, Abdulmjeed Alatawi, Bandar Alsharari, Umar Albalawi, Mohammed Mustafa
College of Computing and Information Technology, University of Tabuk, Tabuk, 71491, Saudi Arabia
Industrial Innovation and Robotics Center, University of Tabuk, KSA, Tabuk, 71491, Saudi Arabia

***Abstract*—With the rapid increase of smartphones and the growing interest in their applications, e.g., Google Play Apps, it becomes necessary to analyze users' reviews whether they are expressed as ratings or comments. This is because recent studies reported that users' reviews could provide us with useful clues and valuable features that can help in understanding the broad opinion about some applications in term of security awareness. Several techniques have been developed for this crucial task and significant progress have been achieved such as Semantic and Sentiment Analysis, Topic Modelling, and Clustering. The majority of the existing methods are mainly based on representing reviews' words in a Bag-Of-Words vector space with String-matched approaches without considering the common polysemy and synonymy problems of words. This is true due to the fact that users who make use of these applications are often from a diverse background and thus, different vocabulary. This paper proposes a new approach to classifying security opinions about applications from users' reviews while considering special features of synonymous and polysemous words. To achieve this task, the proposed model makes use of word embedding, topic modelling, Bi-LSTM, and n-grams approach. For the proposed model, a new dataset is built that contains reviews about 18 popular applications. The application's selection was primarily governed by making the dataset diverse in its domain. The experiment results showed that the proposed ensemble model which combines the prediction of the extracted features, which in turn captures synonymy, polysemy, and dependency of words-is significantly useful, and it achieves better results with an accuracy approaching 90% compared to the use of each technique separately. The model could contribute in preventing mobile users from unsafe applications.**

Keywords*—*Security awareness; semantic analysis; sentiment analysis; mobile applications; topic modelling; clustering

I. Introduction

The history of smartphones dates to the 1980s; however, the revolution in the industry came when the first iPhone was introduced in the market. This revolution has completely changed the life of humanity and how people interact; and what seemed like a distinct dream has become a tangible reality. After 2007, users felt the power within their hands, with smartphones becoming an integral part of every life. Today, there are close to six billion smartphone users, which amounts to over 70% of the world population. It is estimated that access to smartphones would reach over 7.5 billion in the next five years [1], as shown in Fig. 1. Smartphones are also expected to affect all aspects of life on earth in the coming years. Smartphones have evolved over the last decade and have come a long way from being a prized possession to an ultimate necessity. When IBM introduced the first smartphone,

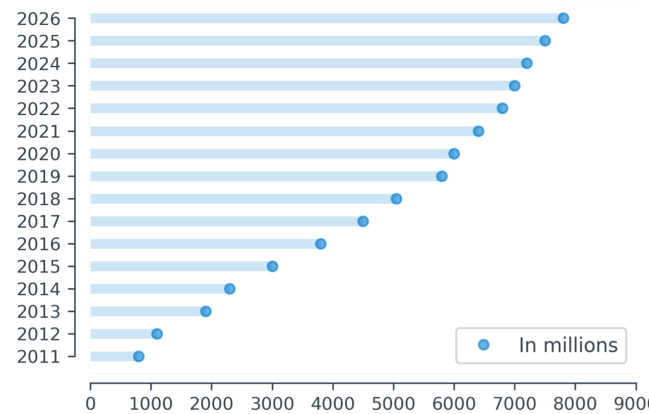


Fig. 1. Number of Smartphone Users.

it was more about the touchscreen, email, fax, and notes than anything fancy or smart power in hand. The first smartphone was never that powerful and looked like it would take ages to reach the common people. It was in the year 2007, however, when the first iPhone was introduced, and it captured the market by storm. It possessed a number of features, including the capacity to play music and capture images, not to mention that it was priced at something that people could think to own [2].

Today, the 'New Generation' of smartphones have increasingly become more innovative and eye-catching. During the current pandemic (COVID-19), smartphone users have increased [3]. While in lockdown, such users have befittingly found solace in these gadgets, allowing them to stay connected with their families, gain access to information, and use them to join work platforms, such as Teams and Zoom, in addition to other users. Many companies started producing smartphones and application distribution platforms. For example, Samsung entered the market strongly and distinctively in 2009 through the Android operating system and has since been continuously developing its famously known application store 'Google Play' (GP). There are millions of different smartphone apps that are emerging as the number of users increases and technology develops. In 2017, the Google Play store had approximately 3.5 million apps [4], as shown in Fig. 2.

One of the advantages of the applications is that they allow users to review and evaluate. Therefore, any application downloaded by any of the operating systems can get

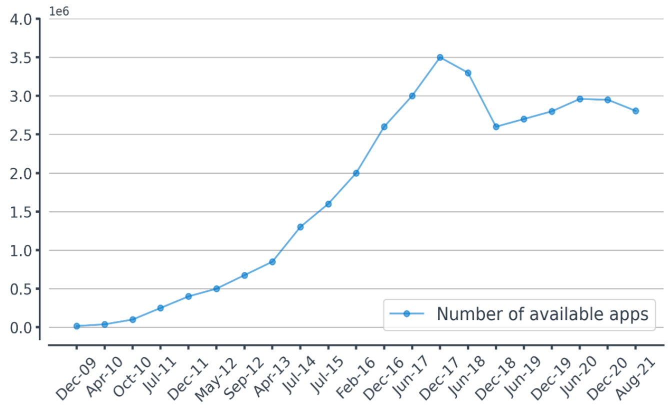


Fig. 2. Number of Available Applications in Google Play.

reviewed by its users, which lets others know about the app's benefits, drawbacks, or any security issues. User reviews and ratings are some of the most critical factors in determining whether or not an app will survive and prosper in the market. Furthermore, data scraping can be used with these reviews to benefit from them or to provide information. According to a survey published in 2018 [5], 94% of buyers avoided a business after reading unfavorable reviews about it online. Interestingly, 80% of clients would avoid doing business with a company if it had less than four stars, which is also relevant to the app industry. Hence, it is crucial to take app reviews into consideration. Through the advancements in new apps, software, and technologies, IT security is forced into continuously being upgraded to prevent it from breaching privacy. With so many apps, as well as social media platforms available to communicate and express ideas freely, technology has aided an explosion of data in recent years. This has created a loophole through which smartphone users' personal information can be easily accessed, raising larger security concerns and difficulties.

Smartphones invading human life has produced multiple apps that are related to almost everything a user does in their routine life, indicating that the user of a smartphone is never alone and always being watched by someone or the other. Today, there are apps developed for everything: job-hunting, reading books, watching movies, or anything else throughout the day [6]. These apps give their users the freedom and ease to complete the requirements ranging from simple data collection or information gathering to entertainment or any further assistance. According to [7], the number of individuals using mobile banking applications has surpassed two billion, accounting for around 40% of the global adult population. A few fact checks of the security issues in these apps [8] are as follows:

- In the second quarter of 2018, mobile applications and mobile browsers accounted for 71% of all fraud transactions.
- In 2018, tools and leisure applications accounted for 54% of malicious mobile apps.
- The United States accounted for 25% of all mobile malware incidence, India for 23%, and China for only 3 %.

Modern mobile operating systems include a variety of security features. Installed apps can only access files in their own sandbox folders by default, with user privileges preventing them from changing system files. However, mistakes made by developers while designing and coding for mobile applications leave security vulnerabilities that adversaries may exploit. A lack of multi-factor authentication is one of the most frequent risks to mobile app security. Without this protection, a hacker only needs a tiny amount of personal information to access the data.

The rest of this paper is organized a follows: The Background and related works are described in Section II. The database collection of this paper is summarized in Section III. Section IV discusses the datasets analysis. In Section V, the proposed model is discussed. Section VI illustrates the experiment results, while Section VII concludes the paper and briefly discusses areas of future work.

II. Background and Related Work

Several methods and techniques have been proposed in the literature for the task of how to extract sentiments in user's review texts. Some of these methods were built on top of supervised learning, in which a large set of data is being labelled, whereas others were based on unsupervised learning. However, the common theme among the provided studies is how to extract useful features that can help in identifying sentiments or opinions. Besides the use of conventional classification techniques that are based on the traditional frequency of both positive and negative words, the utilized approaches for feature extraction that are related to this study can be informally categorized into three main categories, these are Sentiment-based, Semantic-based and Topic-based techniques. Next section discusses each of which in more detail.

A. Sentiment-Based Analysis

Prior studies have proposed various sentiment analysis approaches. The authors [9] used text mining to summarize user evaluations and extract key features from Android app reviews. They then utilized a natural language processing (NLP) technique to write rules. They used SAS® Enterprise MinerTM 7.1 to summarize reviews and pull-out features, and SAS® Sentiment Analysis Studio 12.1 is utilized to perform sentiment analysis. Reviews from two Apps from two different categories were used in the experiment; namely: "Where is my Perry" from "Brain & Puzzle" category and "Beautiful widgets" from "Personalization" category. Data was gathered from the Google Play Android App Store. Textual reviews with rich content were collected from the App Store website. Six hundred reviews were extracted, A corpus of 500 reviews was used for text mining, constructing sentiment models, and developing sentiment rules, while a testing dataset of 100 reviews was used. Each textual review is categorized into a positive or negative directories based on overall ratings depending on the 5-star rating scale used by Google Play. Rule based models were applied to the testing datasets. The NLP rule-based model outperformed the SAS® Sentiment Analysis Studio 12.1 default statistical model in predicting sentiment in test data. In analyzing customer sentiment, NLP rule-based models give deeper insights than statistical models.

In the research [10], A series of comparison studies were carried out using classification algorithms to classify emotions. Review length and feature representation. The experiments were carried out on over 1400000 mobile app reviews based on four characteristics: 1) short average length; 2) large span of length; 3) power-law distribution; 4) Significant difference in polarity. This research makes a comparison between two classification techniques; namely: Support Vector Machine (SVM) and Naïve Bayes. The experimental results stated that: 1) Bayesian method is proved to perform better than the SVM on the classification link; 2) N-Gram is applied (N=2) for the best result on feature representation after the Chinese word segmentation; 3) Feature extraction process can improve the sentiment classification accuracy when reviews have more than 150 words; 4) Short reviews are more easily classifiable then long ones.

Zhang et al. [11] built a novel system for named MoSa (Mobile Sentiment Analysis) for data analysis. The proposed system employs algorithms for detecting and analyzing different types of application big data, such as news comments. They proposed a sentiment analysis approach relying on a mixed sentiment dictionary and new algorithms for calculating comment sentiment scores and categorizing comment sentiment tendencies. The mixed sentiment dictionary is more effective at classifying short texts according to experimental data. The average word length should be less than 2.3 when utilizing a sentiment dictionary to analyze brief texts. They also suggest certain unique statistical models that describe public behavior. In doing so, they discovered that standard deviation is a better indicator of public sentiment than other statistics.

The study [12] proposed a security-related and evidence-based ranking scheme, called SERS. This approach is utilizing the principles of Subjective Logic operations, theory of evidence, Sentiment Analysis techniques, Static taint analysis, and NLP. SERS achieves a holistic rank ordering of related apps and generates insights linked to apps available on the Google Play Store, in both structured and unstructured contexts. They tested their method on publicly available apps from the Google Play Store and compared their results to common ranking methods like average star ratings.

The review paper [13] focuses on the sentiment analysis (SA) for users' review of mobile apps to extract user requirements for developing new apps or improving existing ones, i.e., requirements evolution. They have investigated the approaches employed in the literature during the years 2009 to 2015 to answer the following research questions: 1) "Are there any publications about using SA of mobile app users' reviews for requirements evolution?", 2) "What are the methods used?", and 3) "What are the tools used?". Results demonstrate the value of using SA to analyze user reviews and reports, automated methods, and tools for evaluating reviews with features and sentiments.

The research published by Rizk et al. [14] utilized the naïve Bayes method and LDA algorithm for determining the customer sentiment on mobile banking apps and what aspects need to be maintained or improved in the app. The data used for the experiment were collected from the user reviews on Google Play store. Data were manually labelled and two classes (positive and negative) were generated. Naïve Bayes is utilized for the sentiment analysis process, LDA is used for the process of topic modelling. The experiment's findings revealed that the Nave Bayes approach had a high level of accuracy, recall, and precision. The value of k=5, which equals 86 % accuracy, 93 % recall, and 92 % precision, has the highest accuracy, recall, and precision. The most common topics in negative classes, according to the LDA method, are OTP code delivery constraints, application login problems, and network connection issues. The most often common topics in positives classes, on the other hand, were ease, simplicity, and helpfulness.

Lavanya et al. [15] proposed an opinion mining algorithm utilizing the word alignment model for extracting opinion targets and opinion words from online reviews extracted from Twitter. The proposed project aimed to design opinion words and opinion target predicting algorithms that mine user reviews posted on Twitter to analyze the market status of a particular product. The results of the experiments show that the pro-posed method achieves higher accuracy in a more efficient manner. Furthermore, the project's ultimate goal is to create possible consumer-oriented items such as mobile phones, laptops, and so on.

The study [16] has merged three techniques; namely, 1) Sentiment analysis (SA), 2) NLP and 3) Text Analysis (TA), to classify the reviews of mobile apps into categories related to software maintenance and evolution. This study provides a high-level taxonomy of sentences' categories included in user reviews that are related to the maintenance and evolution of mobile applications. Moreover, it provides a novel approach using NLP techniques to extract the user Intentions expressed in user reviews. The study's findings revealed that combining NLP, TA, and SA methods allows app developers to discover meaningful phrases with higher precision (75.2%) and recall (74.2%) than utilizing each methodology separately. The authors also demonstrated that increasing the size of the training set improves accuracy and recall in specific setups. They also discovered that a classifier that is trained using both structure (NLP) and sentiment (SA) characteristics performs considerably better than the one that is just trained with text (TA) features.

B. Semantic-based Analysis

The authors [17] proposed a framework for automatically analyzing the differences and similarities between app reviews from the Google Play Store and tweets based on the semantics of the words. They demonstrated that the framework can automatically identify similarities and differences using statistical tests and human expert review. This system may be used instead of the costly and unreliable crowd sources (due to the evaluation of non-experts). By filtering similar and different issues, it can decrease redundant information and group the main points. The results of multiple experiments, which were compared to expert evaluation, showed that it may be used to find similarities and differences between extracted topics, n-grams, and user comments.

Yadav et al. [18] proposed a novel framework that employs Word2Vec word embeddings to incorporate semantics into app user feedback analysis. They use Google Play, Store and Twitter as a case study to see if their method can detect similar/different comments in the two well-studied types of bug reports and feature requests in the literature. Statistical analysis and human expert evaluation both validated the result. The result demonstrated that this approach could measure the semantic differences between users' comments in both groups automatically. The framework may be used to create intelligent tools that combine user input from various platforms while also allowing for automated analysis of reviews.

C. Topic Modelling and Clustering

The authors [19] conducted a study employing machine learning techniques to compare national cybersecurity strategies (NCSs) for different nations. Topic modeling and clustering methods were utilized to investigate the similarity and differences between NCSs and to identify the underlying topics that are appearing in them. A total of 60 NCSs developed between 2003 and 2016 were gathered and examined. They discovered that membership in international intuitions could be a determinant factor for NCS harmonization and integration using institutional theories. The study indicated that quantitative analytical methods such as LDA and clustering can be used to acquire a wider picture and insights during the formation of NCS while analyzing qualitative data such as textual policies strategies, and legislations. The clustering method's results assisted us in gaining a better understanding of the overall similarities between NCSs. The findings suggested that members of a group such as NATO or other like-minded allies have established more integrated and coordinated NCS.

Moubayed et al. [20] introduced a tool that summaries, categorizes, and models such data sets, as well as a search engine that allows users to query the model created from the data. The tool is based on a technique known as probabilistic topic modeling, which goes beyond document lexical analysis to model the intricate relationships between words, documents, and abstract concepts. It will help academics query the papers' underlying models and access the library of documents, allowing them to be sorted thematically. LDAVis, a third-party tool, is also included to provide a more in-depth understanding of the LDA model's inner structure. The tool has simple and intuitive navigation features that display the words inside a subject as well as the relationships between topic and words. This is a very useful tool for security experts to assess the quality of the created model and fine-tune the search queries they provide to the search engine.

D. Analyzing the Current State of the Research in Users' Reviews (Security)

To build a well-designed architecture, firstly existing studies are analyzed briefly, in terms of whether the used model can capture the unique characteristics of English language that often present in reviews.The main four features are whether they are covered by the described studies or not. These features are synonymy, polysemy, contextual information inside text, hidden and latent variables in texts, and word dependency. Table I briefly discusses the outcomes and the main shortcomings in each of the listed studies.

III. Database Collection

Existing corpora are either distributed under some license policy (not accessible) or they do not suite this study, which is the classification of English reviews for some of the most popular applications from a security perspective. For this reason, we decided to collect our own corpus with an eye to make it public for researchers in the near future. English reviews were crawled by making use of a quantitative approach from the mobile app metadata of Google Play Store during the period of January 2021 to September 2021. In particular, several tools including the most popular Google Play Scrapper are used for this purpose. Accordingly, a Python application is built based on the API of Google Play Scrapper. One of the elegant features of Google Play Store is that it does not require any dependencies. In the process, however, a list of some popular keywords in security and cryptography domains was prepared by an expert in the field. Examples for the employed tokens include words like cryptography, confidentiality, breach, crack, etc.

Recently, the research [21] showed that racing game applications are often downloaded by users and their reviews are frequently reviewed. The same study also showed that education apps are the most dominant in Google Store. Motivated by the arguments and with an eye on analyzing security in several domains, six categories have been chosen in the most popular domains and from which the majority of the applications are often downloaded. These are games, finance, education, shopping, entertainment, and social media. The domains were specified from mobile app metadata.

This study focus is concentrated around 18 common and well-known applications, each of which is under a specific category. In particular, in each domain, three applications are chosen among the top popular ones. The intended period was from 2013 to 2021. Given these constrains, about 250435 reviews are collected. Due to the messy theme of the raw reviews (and before the features were extracted), the data was firstly gone through various preprocessing. The main of this processing is to populate text into a standard and canonical form and thus, it can be analyzed. Those preprocessing steps are as follows:

1) Awkward Reviews: the data was filtered to remove awkward reviews and those were in mixed languages (i.e., English and French) and emojis.
2) Punctuation Marks and Symbols: to have clean text, all punctuations, marks, and symbols were eliminated.
3) Normalization: texts were normalized, trimmed, and cases folded, started with removing stopwords. Stopwords are often eliminated to compress corpus size. Additionally, they are non useful and of little importance because they are present approximately in the majority of the reviews. However, since the data had many misspelled words, a list of additional awkward stopwords is identified that were removed during this cleaning phase. For example, the phrase doesn't (does not) has been removed.
4) Porter Stemmer: After that, the texts in the dataset were stemmed using Porter stemmer. Porter stemmer is a light stemmer that can chop off the suffixes to render different forms from a single stem.

TABLE I. THE COMPARISON STUDY OF SENTIMENT AND SEMANTIC-BASED ANALYSIS

References	Contribution	Approach	Potential shortcoming
[9]	A sentiment feature-based model that combines text mining techniques with NLP. In particular, the learned features were automatically used to extract keywords.	Text clustering techniques from text mining plus Part-Of- Speech (POS) in NLP.	- weighting method used for clustering was based on simple Frequency. - Contextual information was not exploited. - Dataset size was relatively small. - POS needs large data to correctly identify sentiment in reviews. - The impact of stopwords removal was not investigated.
[13]	Investigating the importance of employing sentiment analysis techniques in providing a picture about the mobile applications.	A general review paper.	- No empirical analysis for the employed approaches and techniques. - The tools that were covered by the study were not assessed.
[14]	Identifying the weaknesses and strengths in mobile banking apps through the analysis of topics in users' reviews comparing results to those obtained through Naïve Bayes classifier.	Latent Dirichlet Allocation (LDA) topic model.	- Neither Hidden Topic Markov Model (HTMM) nor N-gram language models were used in the study. - Additionally, the use of LDA is not a good option when dealing with short texts, especially those with messy nature like reviews. In such cases, the use of NMF is the most adequate approach.
[15]	An opinion mining algorithm was proposed to extract opinion words and target opinion from Some online reviews extracted.	The IBM-1 alignment model, which is originally used for translation task in NLP.	IBM model needs a considerable large amount of parallel data to extract lexical weights. However, a small size of sentences were used in the study.
[16]	Classify the reviews of mobile apps into groups related to software maintenance and evolution.	Sentiment analysis, NLP, Text Analysis.	The study needs to be extended to a larger number and variety of apps.
[18]	Integrating users' reviews and feedback about specific application(s) from different platforms.	Google's word2vec Pre-trained word embedding model.	- Topicality has not been modeled in the developed approach. - Dependency of words in text was not investigated and thus, words were handled as if they were independent inside the texts.
[20]	Building a topic-based (semantic-based) search engine to tackle the change in security threats.	Latent Dirichlet Allocation (LDA) topic model.	Bag-Of-Words topic modelling does not consider Markovian relations between topics. Neither HTMM nor N-gram language models were used in the study. Thus, topics may be incoherent.

TABLE II. DISTRIBUTION OF COLLECTED REVIEWS IN DIFFERENT APPLICATIONS AND DOMAINS NORMALIZATION PROCESSES

Category	Application	Number of Reviews
Games	Pubg Mobile	58139
	Call of Duty	
	Among us	
Education	Google Classroom	9160
	Duolingo	
	Google Play Books	
Shopping	Amazon Shopping	24228
	AliExpress	
	Wish	
Social Media	Snapchat	28534
	Tik Tok	
	Twitter	
Entertainment	YouTube	38168
	Prime Video	
	Netflix	
Finance	PayPal	18907
	AlRajhi	
	Investing	
Total		177136

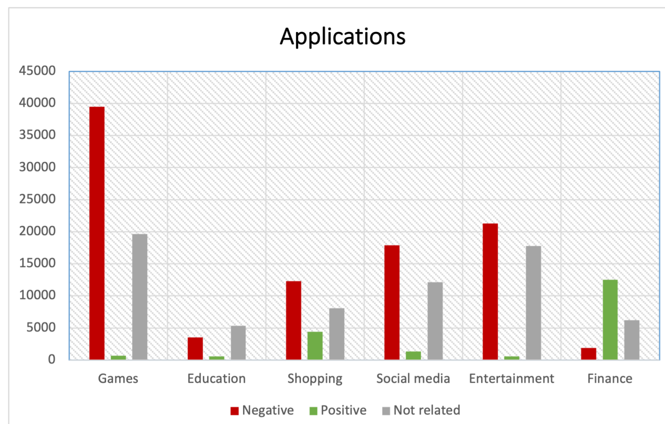


Fig. 3. The Distribution of the Three Classes According to Category.

Following this, the extracted reviews were stored into a CSV format file containing their corresponding metadata. After the above normalization processes, Table II illustrates the distributions and the total number of reviews of each application with its category. For Training phase, the dataset needs to be labeled. Accordingly, we decided to classify the sentiments of the review texts into three main categories (Positive, negative and Neutral). The word neutral means the review is actually not related, but it was collected because of one of its constituent words is on the keywords list. One might ask why to eliminate such reviews. After a deep discussion, we decided to keep them so as to mimic what actually happening in the real world. From that perspective, a human judgment process was conducted to label the reviews in one of the three aforementioned classes. Fig. 3 shows the distributions of the three classes according to category.

IV. DATASETS ANALYSIS

At first, the distribution of labelled classes from two prospective: users' ratings and the class labelling is investigated. In all reviews that were collected from different domains, the user's rating is of 1 to 5 stars. Fig. 4 plots the distribution of different ratings using a pie chart. In the figure, it is obvious that the higher rating for all applications together is found to be of user rating level 1.

This indicates that the overall satisfaction in terms of the selected is high. However, users' ratings may not be perfect. Therefore, the correlation between human labelled classes and these users' rating levels is studied. In particular, exploring whether the sentiment polarity is correlated with users' ratings. Therefore, the ratings levels is categorized as shown in Table III. In the same table, the total number of different review categories plus labelling on the same categories, are also provided.

From Table III, it is evident that the distribution between the two types of labelling (users' rating vs. our labelling)

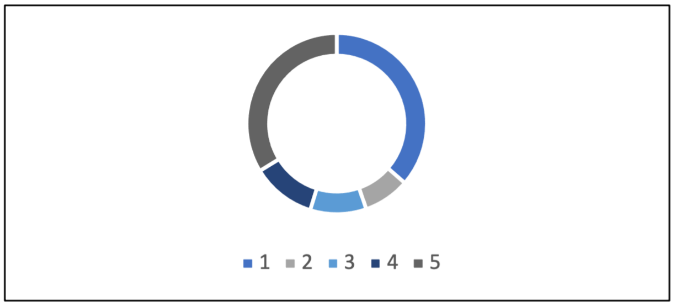


Fig. 4. The Destruction of the users' Ratings for All Categories Together.

TABLE III. THE CATEGORIZED LABELS FOR USERS' RATING AND OUR LABELLING

Rating	Categorized Label	Users' Rating Class	Our Labelling
≥ 4	Positive	73774	17216
$\doteq 3$	Natural	86357	66406
≤ 2	negative	17005	93514

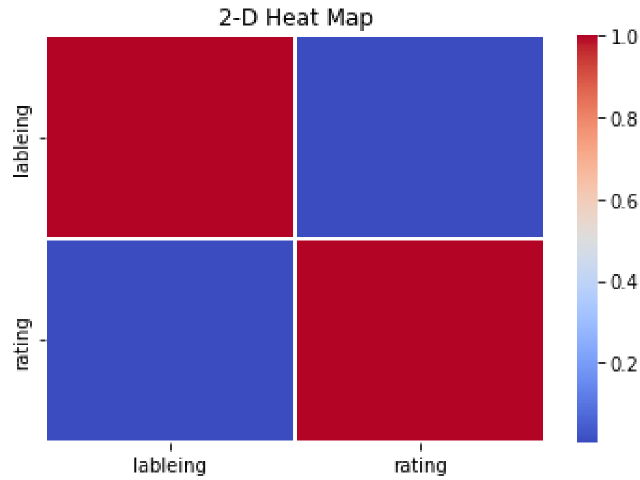


Fig. 5. The Destruction of the users' Ratings and Human Judgment Labeling.

are significantly inconsistent and extremely varied. We cannot claim that the labelling is perfect, but it has been judged by the same team, unlike the users' ratings, which are judged by a large number of users, and thus, we can claim that the labelling (human labelling for distinguishing purposes) is better than the identified classes of human labels. This is an important conclusion because often researchers attempt to avoid the boring and time-consuming process of labelling. Therefore, unless the ratings are checked statistically, we cannot depend on users' ratings, which are very subjective among a large number of users.

Fig. 5 shows the correlation between the two identified types of labels (Human labelling vs user-rating labels) using HeatMap. It is clear that the two types of ratings are highly uncorrelated. This is an indication for the fact that users' ratings are not aligned with human judgment labelled about reviews. The conclusion is aligned with Table III. Next step is to analyze why there are many unrelated reviews. We went through several reviews, not related to the study but they were crawled. From the proposed analysis, it is showing that the main problem for this phenomenon is the polysemy of words. Polysemy means a single word that have different meanings. For example, the word apple in the information technology domain has a very different meaning from its meaning in the agriculture domain. In the former field, it is likely related to apple phone, whereas in the latter, it refers to fruits. This polysemy nature of words in English texts cannot be disambiguated unless contextual information (topicality) and homonymes inside the texts are captured. There are many topical model and approaches that can be utilized to mitigate this problem, e.g., Latent Dirichlet Analysis (LDA), Words' Embeddings, and BERTopic (Bidirectional Encoder Representations of Transformers).

Nevertheless, the use of such models may not resolve this problematic phenomenon and the availability of contextual information is doubtable. This is because texts in reviews are often too short with relatively weak co-occurrence associations among words. As a consequent result, the texts may not be interpreted (even the hidden topics are captured) due to lack of contextual semantics and topicality. This is one of the main targets for this research, to extract some useful features based on hidden and contextual information in reviews' texts. In Fig. 3, it is also evident that the probability distribution of the collected reviews is imbalanced. This bias can easily make the proposed model, as it will be described later, biased towards a specific category and/or classes. For this reason, in the experiments, a random up-sampling is implemented for each category to be aligned with the main category in the dataset.

V. PROPOSED MODEL

A. Motivation

Traditional approaches for representing texts in documents/reviews/tweets attempt to create some degree of association between words and documents (reviews in this case) to build some co-occurrence word-document matrix to standard weighting schemes for normalized and stemmed words. In such models, the order of words does not have any impact on the entire process, known as Bag-Of-Words (BOW) model. The main drawback that may occur here is that the model does not reveal the structural semantics among words and the hidden meanings of the entire context inside text. This means that the linguistic contexts of words are only eliminated due to their frequency in each document and among documents within an unordered set of words.

The problem here is that words in human language are polysemous by their nature. The term polysemy means a single word which can have several meanings and senses. On the other hand, synonymy is very prevalent in natural human languages. Synonymy refers to the opposite in which different words may have the same meaning. It is evident that traditional techniques for representing words are limited towards problematic synonymy and polysemy; and susceptible to mismatching problems. Synonyms and polysemy cannot be adequately captured unless three main features in the texts are considered:

1) How to represent words? it is a essential question. In traditional approaches, words are represented using

standard TF*IDF weighting. In such a case, it cannot be predicted the best word from its surrounding text.

2) Capturing hidden topics and latent semantics and variables. Discovering such hidden structures is extremely valuable because more functions (more than classification or clustering technique) can be utilized for different purposes.
3) Capturing the dependence between texts' words. This means that the order is a matter here. This is not an easy task, because the dependency can be captured bidirectionally (in a forward or backward direction). This is be-cause human usually does not begin their thoughts from scratch, there are always some dependencies that are represented inside text. For example, individual words can be understood according to their previous and next words.

B. The Proposed Architecture

Inspired by the arguments above, the proposed model is a carefully designed architecture that can predict reviews while considering their contextual information, ordering and hidden topics. In a brief description as shown in Fig. 6, the proposed model uses:

1) Traditional n-grams TF*IDF, which is the n-gram occurrence of a ”n-word” phrase. In this case, it will capture which words are more likely to occur in reviews and which words are more likely occur together.
2) Distributional Representation: the proposed model will use word embeddings. The latter allows us the next representation of words that can understand their meanings.
3) Bi-Directional Long-Short Term Memory (Bi-LSTM), which is a Recurrent Neural Network (RNN) model used to capture dependencies inside texts.
4) For topic modelling, the proposed model will use the Non-Negative Matrix Factorization (NMF). NMF is a model used to find hidden and latent topics, which generate text according to some hidden topic distribution. One main feature of the NMF is its ability to handle short texts as those present in reviews.

VI. Experimental Results

Down-sample technique is used to balance the distributions across different classes in the dataset. Accordingly, the labelling classes with the highest frequencies were down sampled to the lowest one. The outputs were then merged and then randomly sampled together (shuffled) resulting into 17829 for each class instead of 91283, 17829, and 68024 for negative, positive, and neutral classes, respectively. Following this, the dataset was split into training, testing, and validation partitions with 80%, 10%, and 10% of the entire dataset, respectively, as illustrated in Table IV.

A. Traditional N-Gram (TF-IDF) Model

The model starts to create various features with word n-grams model with n [1, 3]. Hence, the entire reviews were converted into a Bag-Of-Words model. Since the number of n-grams was found to be extremely high, the model tries to use top k-grams with k to be determined empirically. Following this, the features were fed to four different classifiers one by one, these are Logistic Regression, Naive Bayes, Decision Tree and Random Forest. The best result was achieved by employing the bigrams model on the top 296633 features with the logistic regression model. In particular, the overall accuracy performance was found to be 88%, 76%, and 77% for training, validation, and testing datasets. Fig. 7 plots the accuracy for different n-grams.

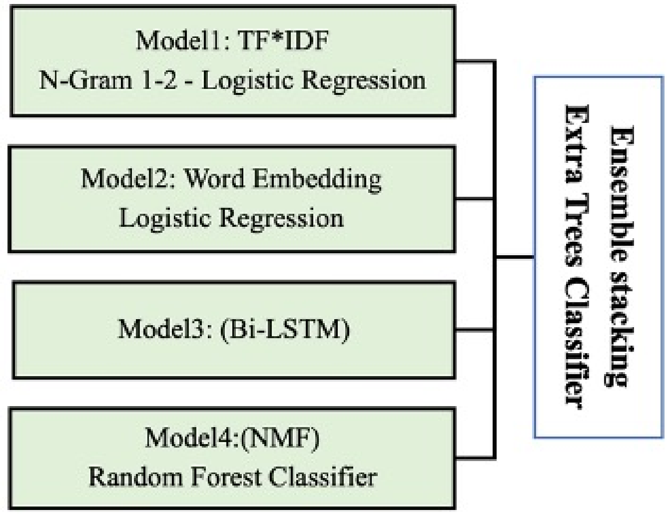


Fig. 6. The Proposed Ensemble Stacking Model.

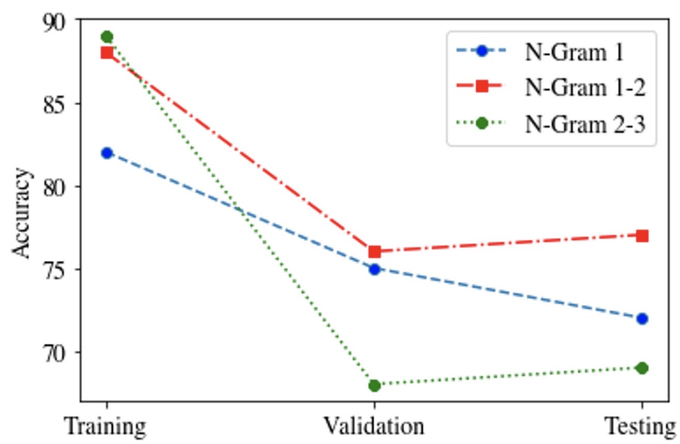


Fig. 7. The Accuracy of Logistic Regression Classifier with n-grams-based Features.

B. Word Embedding Model

In the second run of the experiments, the model extracts features using the word2vec skip gram pretrained model of distributional representation of reviews' words. Hence, the embedding of each word in each review is extracted from the pretrained model and the average of all embeddings of the voter words is computed for each review. In the experiments, 300 features were used for each word. The computed vectors

TABLE IV. Distribution of Collected Reviews in Different Applications of their Domains after Being Balanced

Category	Application	# of Reviews	After balancing process			
			All	Negative	Positive	Natural
Games	Pubg Mobile	58139	13171	7375	642	5154
	Call of Duty					
	Among us					
Education	Google Classroom	9160	2542	625	520	1397
	Duolingo					
	Google Play Books					
Shopping	Amazon Shopping	24228	8574	2423	4094	2057
	AliExpress					
	Wish					
Social Media	Snapchat	28534	7259	3080	1142	3037
	Tik Tok					
	Twitter					
Entertainment	YouTube	38168	9086	3981	536	4569
	Prime Video					
	Netflix					
Finance	PayPal	18907	12855	345	10895	1615
	AlRajhi					
	Investing					
Total		177136	53487	17829	17829	17829

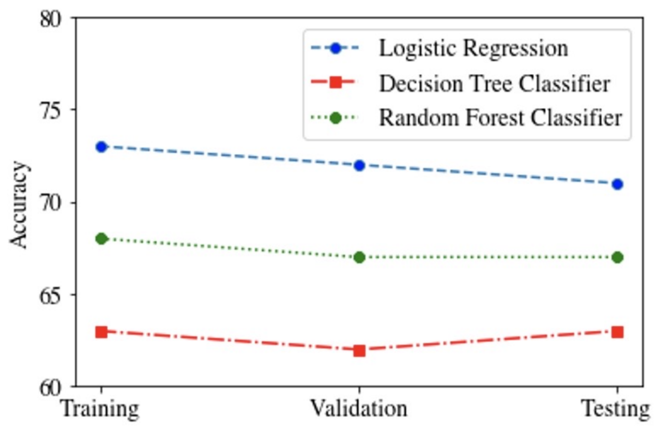


Fig. 8. The Accuracy of the Three used Classifier Word Embedding Model.

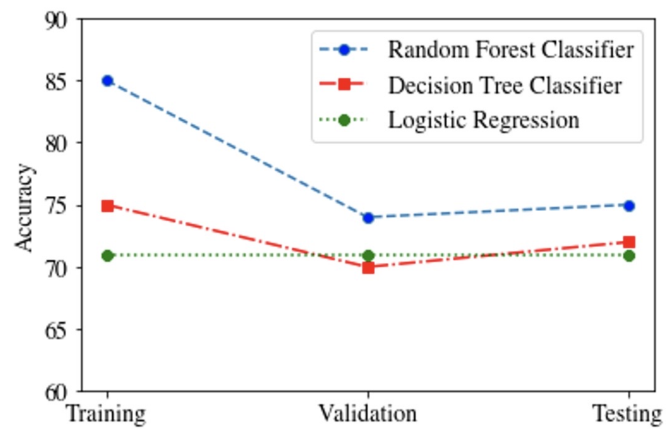


Fig. 9. The Accuracy of the used Classifiers with NMF Topic Modelling Features.

are then fed to logistic regression, decision tree and random forest classifier. Logistic regression classifier also obtains the best accuracy 73%, compared to decision tree and random forest, which both result in almost similar results - 72% and 71%, respectively. The Fig. 8 shows the results of words' embedding using the three classifiers.

C. Bi-Directional Long-Short Term Memory (LSTM) Model

The fundamental concept of LSTM is the feedback connection. At first, the model tokenizes the corpus with a maximum unique word = 25000 words out of +26000 words in the whole corpus, then fitted the tokenizer to the train dataset, and transform all datasets. After that, it pads the tokenized dataset to keep all records with the same length and only needed and effective maximum sequence length which equals 120 features. After that, the model builds the Bi-LSTM architecture, train the Recurrent Neural Network (RNN) model on the padded training dataset, and evaluate the RNN through prediction accuracy. The best model is a simple Bi-LSTM NN with 1 embedding, 1 bidirectional layer, 1 dense layer with dropout 30% with 120 features the maximum sequence length used in the padding step with accuracy 85%, 77% and 78% for training, validation and testing datasets.

D. Non-Negative Matrix Factorization (NMF) Model

In the forth experiment, the model extracts hidden topics that cause the texts in review to occur. For this purpose, it uses NMF. NMF is an unsupervised model that has the ability to discover such latent variables from a large amount of data and the end product is a probability distribution for every topic over words and how likely a word belongs to each topic is also extracted as a distribution over topics. One accredited feature of NMF is that the model has shown to be more effective - compared to other topic models - when it is used for short texts, as those present in reviews. From that prospective, the frequency of the words were extracted using the countvectorizer. the model attempted a different number of hidden topics and empirically concluded that the best number of topics to be extracted is 150. Results, which are plotted in Fig. 9, show that the best result by this method is obtained when a random forest classifier was used with an accuracy approaching 86% in the training dataset and 75% in the testing dataset.

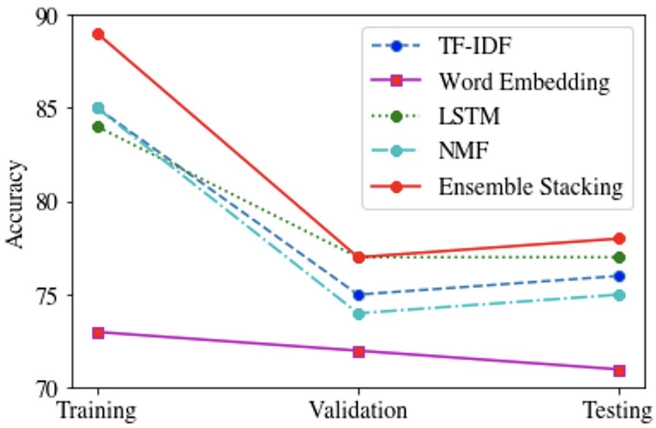


Fig. 10. The Accuracy of the Models Compared to the Proposed Ensemble Model.

E. Ensemble Stacking Model

In the final set of the experiments there is an ensemble model that employs predictions of the base model that created to build a new model. Therefore, instead of averaging all extracted features from different models, we used the prediction output to feed the ensemble stacking model. Accordingly, the parameters of the model were estimated, and the results showed that the accuracy was increased significantly (it almost approaches 89%) compared to the base models. Fig. 10 plots the accuracy of all models when they were compared to the proposed ensemble model.

VII. Conclusion

Several methods and techniques have been developed for the vital problem of extracting and classifying sentiment inside users' reviews of mobile applications. The problem is challenging because in human language, users can express their thoughts in different vocabularies and words, which are synonymous and polysemous by their nature. In particular, polysemy, in which a single word can have several senses, is significantly difficult. Additionally, texts convey contextual information that cannot be captured using conventional techniques like word-document matrices. In other words, words often tend to be generated according to some hidden topic distributions and they are also dependent inside text.

Inspired by these arguments, this paper attempts to contribute to security threat control by classifying automatically mobile applications based on their accompanied reviews. Accordingly, the proposed model combined the words embedding' features that are extracted from reviews of approximately 18 applications with hidden contextual information inside their texts, which were extracted using NMF topic model. However, to capture the dependency among words, the proposed model employed the use of Bi-LSTM in which the text will be analyzed in both forward and backward directions. We also add another type of features, which are those extracted from word embeddings as the belief is that similar words tend to occur in similar texts.

For the experiment, we built the dataset and the reported experiments showed that the proposed model outperforms traditional models for classifying user reviews and it is able to capture the announced features in review text. In particular, the proposed model accuracy is above 89% and the improvement compared to the proposed base model is significantly high. The experiments showed that NMF model topic model is able to capture hidden topics inside short texts of users' reviews. On the other hand, both Bi-LSTM and word embeddings are extremely useful for handling synonymous and polysemous words.

In the future, we are going to investigate the impact of stemming in improving model performance. In addition, it is really interesting to look into the effect of the dataset word embedding, therefore we plan to train the model using the dataset instead of employing pretrained word embedding models. We are also going to expand the model to Arabic reviews. the latter language is rich in its vocabulary and the use of synonymy and polysemy is widespread.

Acknowledgment

The authors would like to thank The University of Tabuk for providing research support and facilities.

References

[1] Ericsson, "Ericsson mobility report november 2021," november, 2021. [Online]. Available: https://www.ericsson.com/en/reports-and-papers/mobility-report/reports/november-2021

[2] SimpleTexting, "History and evolution of smartphones," July, 2021. [Online]. Available: https://simpletexting.com/where-have-we-come-since-the-first-smartphone

[3] H. S. Maghded, K. Z. Ghafoor, A. S. Sadiq, K. Curran, D. B. Rawat, and K. Rabie, "A novel ai-enabled framework to diagnose coronavirus covid-19 using smartphone embedded sensors: design study," in *2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI)*. IEEE, 2020, pp. 180–187.

[4] Statista, "Number of available applications in the google play store from december 2009 to september 2021," October, 2021. [Online]. Available: https://www.statista.com/statistics/266210/number-of-available-applications-in-the-google-play-store

[5] T. Manifest, "4 reasons why you need mobile app reviews," October, 2018. [Online]. Available: https://themanifest.com/app-development/blog/benefits-mobile-app-reviews

[6] M. Hatamian, "Engineering privacy in smartphone apps: A technical guideline catalog for app developers," *IEEE Access*, vol. 8, pp. 35 429–35 445, 2020.

[7] ptsecurity, "Mobile application vulnerabilities and threads 2019," July, 2019. [Online]. Available: https://www.ptsecurity.com/upload/corporate/ww-en/analytics/Mobile-Application-Vulnerabilities-and-Threats-2019-eng.pdf

[8] CPO, "How hacker attack mobile apps," April, 2019. [Online]. Available: https://www.cpomagazine.com/cyber-security/how-hackers-attack-mobile-apps

[9] J. Liu, M. K. Sarkar, G. Chakraborty *et al.*, "Feature-based sentiment analysis on android app reviews using sas® text miner and sas® sentiment analysis studio," in *Proceedings of the SAS Global Forum 2013 Conference*, vol. 250. Citeseer, 2013.

[10] L. Zhang, K. Hua, H. Wang, G. Qian, and L. Zhang, "Sentiment analysis on reviews of mobile users," *Procedia Computer Science*, vol. 34, pp. 458–465, 2014.

[11] Y. Zhang, W. Ren, T. Zhu, and E. Faith, "Mosa: A modeling and sentiment analysis system for mobile application big data," *Symmetry*, vol. 11, no. 1, p. 115, 2019.

[12] N. S. Chowdhury and R. R. Raje, "Sers: A security-related and evidence-based ranking scheme for mobile apps," in *2019 First IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA)*. IEEE, 2019, pp. 130–139.

[13] N. M. Rizk, A. Ebada, and E. S. Nasr, "Investigating mobile applications' requirements evolution through sentiment analysis of users' reviews," in *2015 11th International Computer Engineering Conference (ICENCO)*. IEEE, 2015, pp. 123–130.

[14] M. E. Permana, H. Ramadhan, I. Budi, A. B. Santoso, and P. K. Putra, "Sentiment analysis and topic detection of mobile banking application review," in *2020 Fifth International Conference on Informatics and Computing (ICIC)*. IEEE, 2020, pp. 1–6.

[15] T. Lavanya, M. J. P. JC, and K. Veningston, "Online review analytics using word alignment model on twitter data," in *2016 3rd International Conference on Advanced Computing and Communication Systems (ICACCS)*, vol. 1. IEEE, 2016, pp. 1–6.

[16] S. Panichella, A. Di Sorbo, E. Guzman, C. A. Visaggio, G. Canfora, and H. C. Gall, "How can i improve my app? classifying user reviews for software maintenance and evolution," in *2015 IEEE international conference on software maintenance and evolution (ICSME)*. IEEE, 2015, pp. 281–290.

[17] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[18] A. Yadav, R. Sharma, and F. H. Fard, "A semantic-based framework for analyzing app users' feedback," in *2020 IEEE 27th International Conference on Software Analysis, Evolution and Reengineering (SANER)*. IEEE, 2020, pp. 572–576.

[19] F. Kolini and L. Janczewski, "Clustering and topic modelling: A new approach for analysis of national cyber security strategies," 2017.

[20] N. Al Moubayed, D. Wall, and A. S. McGough, "Identifying changes in the cybersecurity threat landscape using the lda-web topic modelling data search engine," in *International Conference on Human Aspects of Information Security, Privacy, and Trust*. Springer, 2017, pp. 287–295.

[21] P. B. P. Reddy and R. Nallabolu, "Machine learning based descriptive statistical analysis on google play store mobile applications," in *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*. IEEE, 2020, pp. 647–655.

Deep Multi View Spatio Temporal Spectral Feature Embedding on Skeletal Sign Language Videos for Recognition

SK. Ashraf Ali[1], M. V. D. Prasad[2], P.Praveen Kumar[3], P. V. V. Kishore[4]
Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, India.[1,2,4]
Department of Information Technology, Vignan's Institute of Information Technology, Duvvada, Visakhapatnam 530049, India.[3]

***Abstract*—To build a competitive global view from multiple views which will represent all the views within a class label is the primary objective of this work. The first phase involves the extraction of spatio temporal features from videos of skeletal sign language using a 3D convolutional neural network. In phase two, the extracted spatio temporal features are ensembled into a latent low dimensional subspace for embedding in the global view. This is achieved by learning the weights of the linear combination of Laplacian eigenmaps of multiple views. Subsequently, the constructed global view is applied as training data for sign language recognition.**

***Keywords*—Laplacian eigenmaps; 3D convolutional networks; sign language recognition; multi view; skeletal data**

I. Introduction

Sign Language Recognition (SLR) is extremely coordinated movements of hands captured through sensors as 1/2/3D data and translated into text or voice by a machine learning interface [1]. Sign language is a communication medium for hearing impaired people which consists of hand movements and finger shapes that operate independently or collaboratively with respect to upper body parts. SLR is considered an extension of human action recognition (HAR) [2]. Automated HAR or SLR is accomplished through machine learning approaches on multi modal datasets such as RGB, Depth and skeletal information in image, video and data formats. The RGB and depth formats provide appearance information whereas the skeletal joint data exclusively models pose details. Although SL knowledge representation is largely modelled in RGB video formats, it is bottlenecked by motion blurring and spatial resolution of fingers with respect to the frame size. Therefore, the skeletal data has obtained wide acceptance for human action or sign language recognition problems. The 3D skeletal data has been used as vectorized, image and RGB video formats for recognition.

However, the pattern identification process on skeletal 3D video data for building a real time application is a supremely challenging task. Traditional models employed vectorized 3D data for recognition with deep neural networks(DNN) [3]. Above all the DNN models on 3D skeletal action data, long short-term memory (LSTM) [4] networks have shown greater reliability and robustness for HAR tasks. Similarly, 3D skeletal SLR on vectorized data was successfully designed and experimented with color coded Spatio-Temporal features [5]. Singularly, most of these methods presented results related to cross view testing with poor performance as these models received only single view training. As a result, the above methods failed to generalize on building a real time engine for HAR or SLR.

Meanwhile, the above problem is finding solutions in the form of multi view training on Deep Learning Models. Though multi view processing of video data is having 2 decades of research history, it has gained extensive attention in the last few years due to the progress in deep learning approaches. Earlier DNN proposed were constructed with multiple streams feeding into individual views independently whose Softmax scores are fused for getting a final recognition score. Later, learning approaches have trained multiple CNNs for each view and then learned the concatenated features in the dense layers. This approach has allowed for multiple views to share features across classes. Specifically, this process does not restrict the features that were not significant in the decision making. Additionally, the view specific features that play a major role in articulating the desired outcome are ignored.

To overcome the above challenges, we propose to learn a global synthesized target view by linearly combining the independent multiple views as suggested in [6]. However, these intra class independent views have shown to exhibit unequal similarities with other views which biases the result towards the false positives. Hence, to overcome this uniformity across views that influence the target class, we propose higher order Laplacian eigenmaps from [7]. This enables the target feature reconstruction to have a complete non uniform distribution across the multiple independent views. Consequently, we learn a nonuniform linear combination of weights on independent views which can be generalized for any target view. Finally, the synthesized target view features of all classes are classified using standard deep learning architectures. The proposed methodology called multi view spatio temporal feature embedding (MVSTFE) is illustrated in the following Fig. 1.

The proposed MVSTFE is investigated on our 3D skeletal video datasets of sign language (KLEF3DSL_2Dskeletal) [8] and four other multiview action datasets NTU RGB-D [9], SBU Kinect Interaction [10], KLYoga3D [11] and KL3D_MVaction [12]. The performance of the proposed deep networks was tested for the proposed method against the state-of-the-art on datasets. The remaining paper is clustered into four sections. The second section highlights the key historical aspects associated with multi view learning, sign language

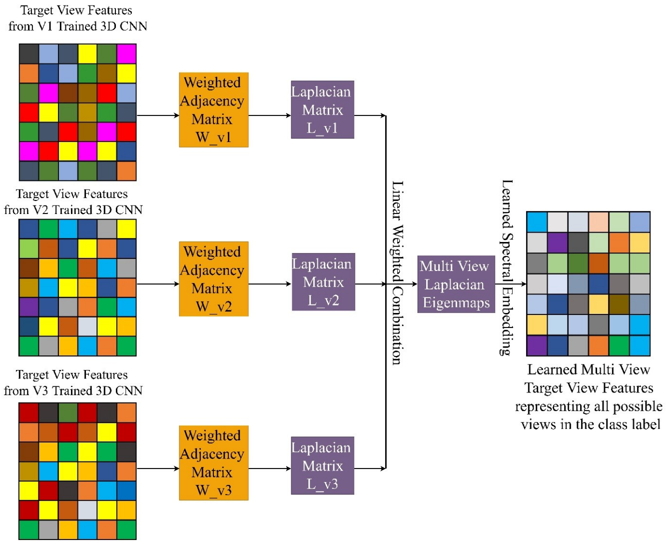


Fig. 1. Illustration of the Proposed Multi View Spatio Temporal Feature Embedding on Skeletal Sign Language Video Data.

recognition and deep networks. The methodology is packaged in the third section and the obtained results for experimentation with analysis are presented in section four. Finally, conclusions were drawn from the analytical insights gained on the overall performance of the proposed models.

II. Literature Review

This section of the paper dwells on the advantages and disadvantages of the previous methods of sign language and action recognition in multiple views. Additionally, it also discusses the current models in deep metric learning.

With the advent of deep learning frameworks, the 2D video based SLR has become powerful with the option of feature learning rather than feature extraction. A large contingent of them is available for perusal [13]. The accuracies reported by these methods are not reproducible or they simply fail to generalize on the video quality or the signer. This has motivated researchers towards higher dimensional data such as RGB D or 3D skeletal representations. Multi modal video sequences that are fed into multiple streams of a CNN are predominantly researched which have shown evidence of exceptional performances in real time for sign (action) recognition applications [14]. The recognition accuracies were better than the single modal datasets. However, the training requires higher computing powers, and the datasets are captured with special devices making it an unfeasible deployable solution.

Eventually, to develop a real time SLR or HAR system, it is intuitive to learn multi views across datasets. This has initiated action recognition research to move in the direction of developing view-based learning algorithms [15], [16]. Multi view HAR has evolved through research using dictionary learning [17], neural networks with adaptable views [18], convolutional neural networks [19] and deep attention models [20], to name a few. However, the most widely researched and acknowledged models are from deep learning networks. Moreover, visual attention models with deep CNNs have established themselves as a formidable solution to multi view learning [21]. Despite their success, attention models are specific to a particular view and the view specific features are to be fused accordingly for classification by the dense layers. The fusion mechanisms ensemble the view specific features into a multi view feature vector that has failed to capture the variations in multi view data [22].

Primarily multiview approaches were classified as multi-view learning and view invariant models. In multiview learning, the video input is considered as a time series of data frames in different views which are learned independently by the classifier [23], [24]. Most of the methods used low level observable features for generating discriminative features [17]. Subsequently, multiple training methods were employed for each of the views to find a set of consistent features between a pair of views [25], [26]. The algorithms are used for finding relationships between views canonical correlation analysis(CCA) [27] and projection matrices [28]. Extending to the above methods are matrix factorization [29] and low rank constrained matrix factorization [30] for capturing view similarities. All these models have shown good performance on instances where the number of views were limited and require extensive computational power for deployment.

Alternatively, view invariant models developed linear descriptors to transfer information between views. Accordingly, these models consider target views as a linear combination of views within a class label [6], [7]. Subsequently, the weight vectors are computed by applying optimization in Laplacian space. Moreover, these works assume that all views contribute equally to the target view features. However, in sign language recognition with video data from multiple source views it is difficult to impose the above assumption in real time. To overcome the disadvantage of equal contribution by all views to the target view, we propose to learn these contributions in the Laplacian space using deep learning.

The following points make the proposed method unique from the existing ones:

1) To design an unequal linear view combiner to extract target view features.
2) To construct highly discriminative Spatio-Temporal features in the Laplacian space.
3) To reconstruct learned target vectors into a Spatio-Temporal feature representation with 3D CNNs.

In order to find an appropriate solution for multiview problems, the following objectives are being formulated:

1) To design an unequally contributing linear view combiner to identify the linear combinations.
2) To learn the mapping function for generating a singularly trainable view invariant Spatio-Temporal feature.
3) To initiate anyone view testing model. We call our proposed model multi view spatio temporal feature embedding (MVSTFE).

III. Multi View Spatio Temporal Feature Embedding (MVSTFE)

This section describes the proposed multi view spatio temporal feature embedding model for multi view sign language recognition on skeletal video datasets. First, a cluster of 3D CNNs is trained independently on individual views for all classes in the dataset. Secondly, a target view is selected randomly which is referenced on the pre trained 3D CNNs for feature extraction. The extracted features from independent view streams are learned by compiling Laplacian eigenmaps to construct a combined target view. This combined target view features will represent a linear combination of Laplacian eigen maps from multiple views generating a highly discriminative feature for all views of the target view class. Finally, these learned target view features will be used for training any deep classifier for sign or action recognition.

A. Independent View 3D CNN Model

The primary step in the process of multi view sign language recognition is to design and train a 3D convolutional neural network (3D CNN). The 3D CNN takes input as the skeletal video sequences as input for supervised training. The number of 3D CNN streams are equal to the number of source views available for training. The 3D CNN architecture used in this work is shown in Fig. 2. The model has 4 pairs of 3D convolutional layers with one set of batch normalization and maximum pooling layers after each pair respectively. The input of the network is a 2D skeletal video sequence of size $256 \times 256 \times 3$ with 100 frames. The features at the end of the convolutional layers are flattened and inputted to two fully connected layers with the last layers being Softmax.

Let $X^{vc} = (x_v = \{S_v\} \forall v = 1 \ to \ V, c = 1 \ to \ C)$ be the RGB skeletal video sequences in V views with $V \in R^3$. The 3D CNN model will extract the features f^v from x_v with view specific labels y_v using the trainable parameters θ_{3D} by optimizing a loss function L on the overall multi view dataset as

$$\theta_{3D} = \arg\min_{\theta_{3D}} L(\theta_{3D}; x_v, y_v) \quad (1)$$

For classification tasks, we need a global loss function to

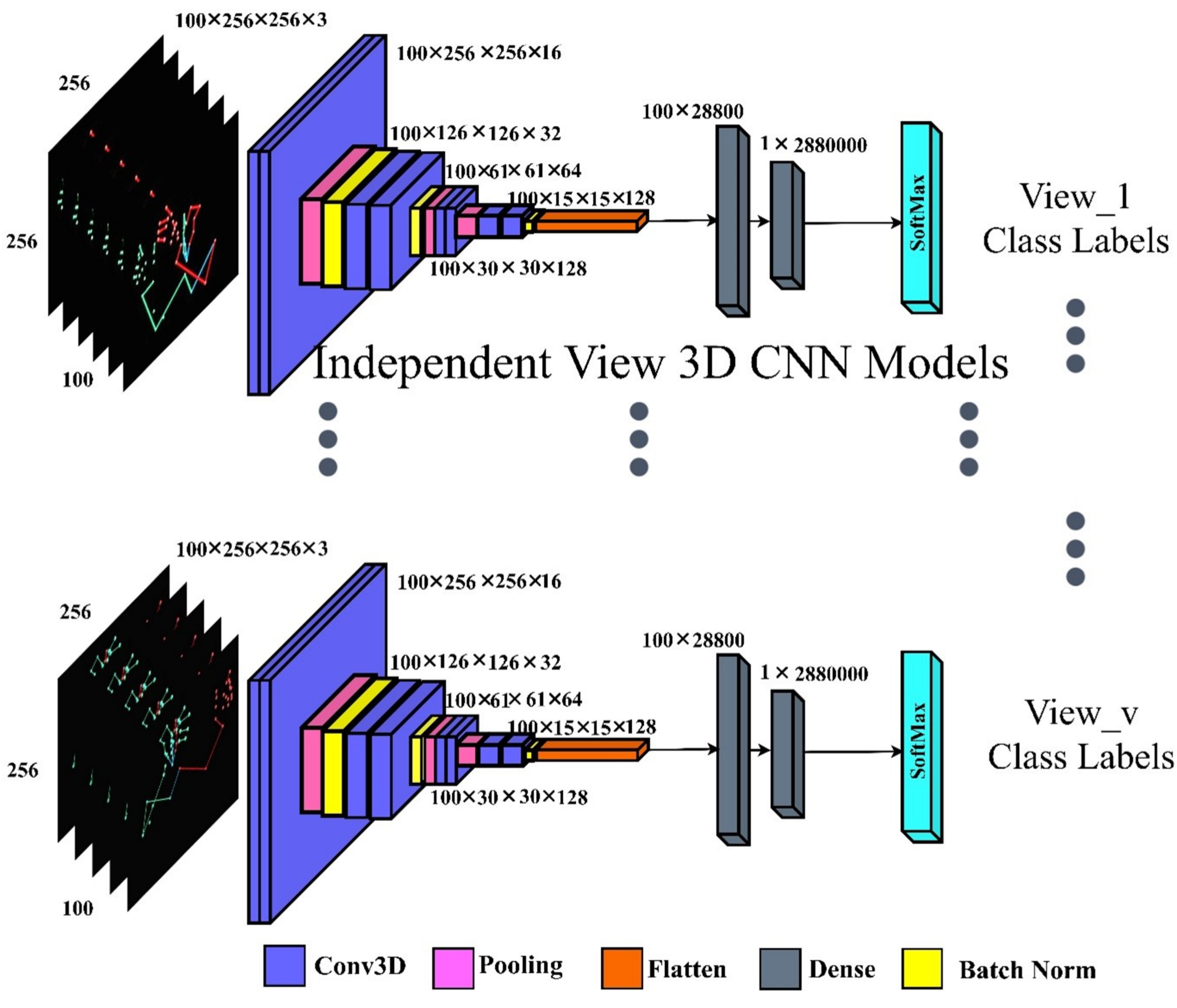


Fig. 2. 3D CNN Architecture for Training Multiple Views Independently across All Classes.

discriminate the classes with the help of SoftMax layers. The class label prediction is computed on the embedding space using the cross-entropy loss functional defined as

$$l_{CrossEnt} = -\sum_{i=1}^{C} (y_i \log(\widehat{y_i}) + (1 - y_i) \log(1 - \widehat{y_i})) \quad (2)$$

The $l_{CrossEnt}$ is the loss function for training the network. The $(\widehat{y_i})$ is the predicted label and y_i is the actual. The C defines the total number of classes in the dataset. Each stream in the network is view independent with the specifications as shown in Fig. 2. Consequently, weight and biases are initialized using unit variance zero mean Gaussian random variable. The filter sizes in all 3D CNN layers is fixed at $3 \times 3 \times 3$. Moreover, the learning rate is dynamically controlled with 10% decrease rate from the previous valued whenever the loss became constant across 10 epochs. The initial learning rate was selected as 0.0001. Stochastic gradient descent optimizer is applied to update the wights and biases in the network. This trained network will be used to extract spatio temporal features from a target view which are further used to construct a combined view features. These constructed view features have the ability to represent all the views within a class label.

B. Combined View Feature Generation

Given a sign class in a specific target view x_{vt} as input the trained model θ_{3D}, the output features f_v at the end of dense layers are represented as

$$\{f_v\}_{v=\{1,V\}} = \sum_{i=1}^{I} \sum_{j=1}^{J} x_v(i,j) * K(k-i, k-j) \, \forall k \in \text{kernel size} \quad (3)$$

The features extraction network is shown in Fig. 3. The network consists of four pairs of convolutional layers with rectified linear activations followed by a 2×2 window maximum pooling layer. The strides of the kernels in convolution layers is one and that of maximum pooling is two. After maximum pooling a batch normalization layer is added to standardize the inputs to the deeper layers. Finally, two fully connected layers are added to learn on the feature extracted in the convolutional layers. Subsequently, the spatial features at the output of dense layers are concatenated along the frames to generate a complete spatio temporal feature matrix representing the 2D skeletal video sequence. Altogether, V streams operate independently in the network generating view specific class features $F^{cv} = \{f_{ic}\} \; \forall i = 1$ to $V \in R^{g \times N}$, Where g is the dimensionality of the features and N is the number of frames. The model is trained with categorical cross entropy loss with stochastic gradient descent optimizer on the entire dataset. The trained model $\widetilde{\theta_{3D}}$ is applied on all the input video frames to extract the feature samples as

$$F^{cv} = \widetilde{\theta_{3D}}(w,b) \times x^c_{vt} \forall V \;\; \& \;\; C \in R^{g \times N} \quad (4)$$

The spatio temporal feature matrix F^{cv} consists of the target view features inferenced from independently trained views across all classes. The objective is to generate a feature matrix that will represent all views in a class as a linear combination of the extracted features. Traditionally, this is achieved by considering the all the mixing coefficients are equally distributed across all views. However, equally distribution of information across all views has produced ambiguous recognition accuracies. To overcome this, non-uniform distribution is proposed [7] with Laplacian eigenmaps. In this work, we incorporate the process of spectral embedding using Laplacian eigenmaps to calculate the mixing coefficients of the linear combination.

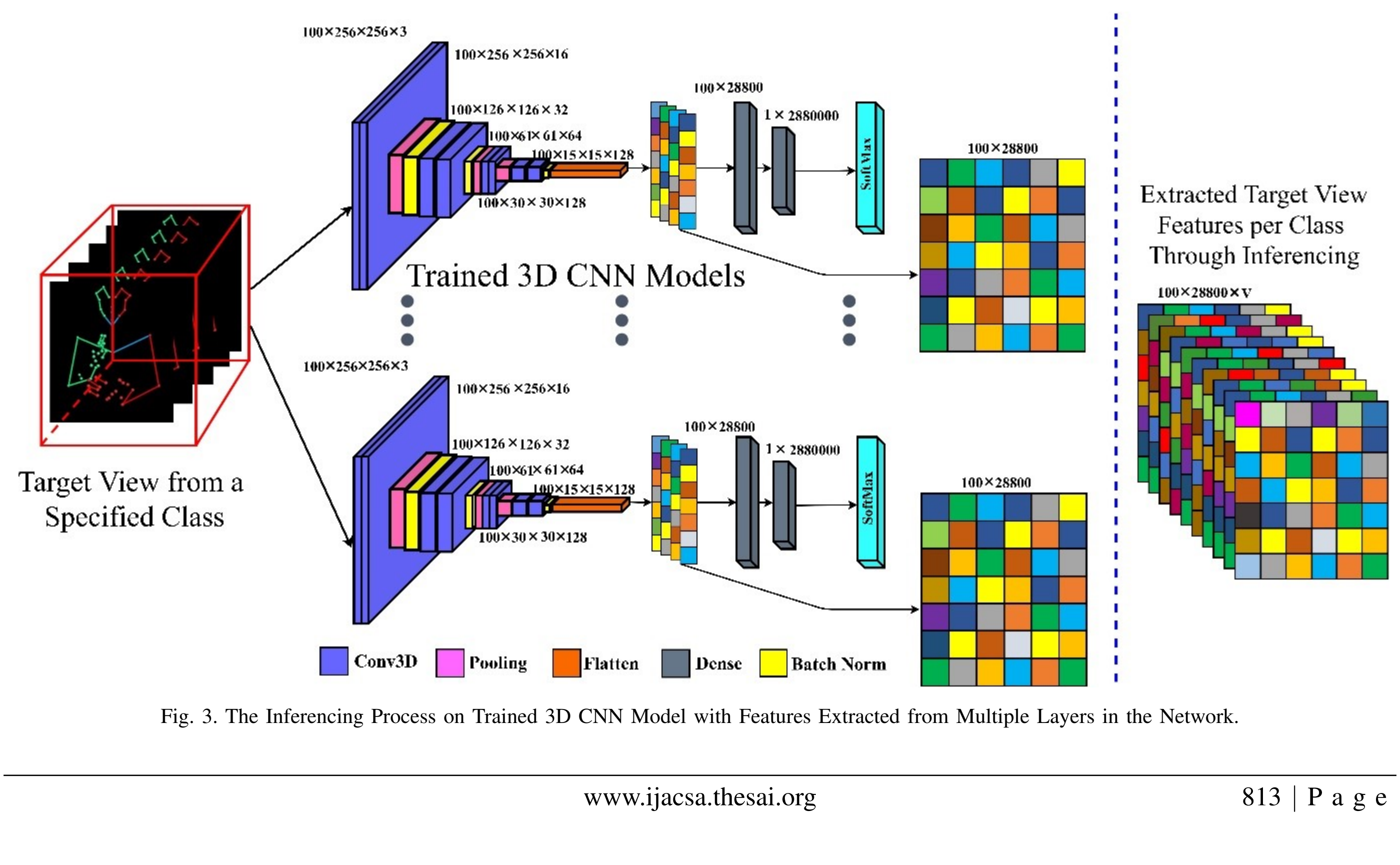


Fig. 3. The Inferencing Process on Trained 3D CNN Model with Features Extracted from Multiple Layers in the Network.

C. Constructing the Non-uniform View Linear Combiner

Given a set of spatio temporal target view features $F^{cv} \in R^{d\times V}$ from a particular class label $y_{i\in C}$ with V views, these views can be linearly combined with coefficients as

$$F^{cv}_{Comb} = \sum_{i=1}^{V} \lambda^i F^{cv(i)} \tag{5}$$

Where, V is the total number of source views and d is the feature matrix dimensionality. F^{cv}_{Comb} is the combined feature representation of the target feature. The mixing coefficients $\lambda^i\ \forall\ i = 1\ to\ V$ is the weighted combination. The constraint on the mixing coefficient is

$$\sum_{i=1}^{V} \lambda^i = 1,\ \lambda^i > 0 \tag{6}$$

The intent in the above representation is to generate a global view that is compatible with all the views in the class. Mostly, the coefficient λ^i is considered as the average $1/V$ across all the views. However, in reality, the views that are in close proximity with the target view contribute more than ${}^{1}/{}_{V}$. Consequently, the obtained linearly combined global view features are least compatible for representing all the views in a class. This problem is solved by evaluating the mixing coefficients of individual views with the help of cost function derived using Laplacian eigenmaps [7].

First, the target features are arranged a V data matrices $F = \left\{F^{cv} \in R^d\right\}_{v=1}^{V}\ \forall d = R^{g\times N\times V}$ as shown in the output of Fig. 3. The objective is to calculate a set of mixing coefficients $\lambda = \left\{\lambda^i\right\}_{i=1}^{V}$. We start by initializing $\lambda = \left[\frac{1}{V}, ..., \frac{1}{V}\right]$. Subsequently, set the $g \times N$ feature points obtained from trained network in the t^{th} target view.

To compute the combined target view embedding features, we subsequently compute the weighted adjacency matrix A^t on the target features and the Laplacian matrix L^i of the individual views with $i \in (1, 2, ..., V)$. Consequently, the global Laplacian L^G of the entire target view class is computed as a linear combination of initial weights. The spectral encoding, Y^G can be computed from eigen value decomposition of L^G as a Laplacian eigen map. Accordingly, select the smallest eigen values other than the zeroth one, reconstruct the spectral encoding Y^{G*}. Using the reconstructed spectral encoding Y^{G*} and Y^G, update the mixing coefficients of the linear combination λ^i. Optimize till the distance between the reconstructed and the original spectral encoding are less than a set experimental threshold.

D. Construction of Laplacian Eigenmaps and Spectral Embedding

Given the feature data points in the t^{th} target view $\left\{F^{cv} \in R^d\right\}_{v=1}^{V}$ with $g \times N$ data points, we first compute the adjacency matrix A^t as

$$\left[A^t\right]_{i,j} = e^{\left(-\frac{\left\|F^{ci} - F^{cj}\right\|_2^2}{\sigma^2}\right)}\ \forall\ F^{ci}\ \ \&\ \ F^{cj}\ are\ associated \tag{7}$$

Where, A^t is a symmetric matrix of size $gN \times gN$. The value of σ is selected as 2. The adjacency matrix establishes a link between the target features extracted from trained CNN in Fig. 2 in all views. If the distance between the features is small, the value in the $(i,j)^{th}$ position tends towards 1 and vice versa. Consequently, A^testablishes a relationship between the features points formed by a set of d data points in multiple views.

Subsequently, to compute a single view feature combination from multiple target view features Laplacian eigenmaps were used from [30]. Laplacian eigenmaps reduces the data by projecting data on a different spectral view without compromising on the relationships between the feature points. Accordingly, the spectral encoding Y^{G*} can be computed by minimizing the cost function defined as

$$f\left(Y^G\right) = \sum_{i,j\in\{\forall gN\}} \left\|y_i^G - y_j^G\right\|^2 \left[A^t\right]_{i,j} \tag{8}$$

The above representation gives the difference between two embedding features in multiple views modulated by their association values in adjacency matrix. If the feature points in multiple views are in close proximity, the adjacency matrix value is large, thus contributing more to cost function. As a result of this, similar data points are preserved in the spectral embedding from different views. Eventually, the solution to the optimization is transformed into a minimization problem as described in [30] as

$$Y^{G*} = \arg \min_{Y^{G^T} D Y^G = 1, Y^{G^T} D1 = 0} tr\left(\left(Y^G\right)^T L^G Y^G\right) \tag{9}$$

The global laplacian matrix L^G is computed as $L^G = D - A^t$, where D gives the degree of connectivity in the data as $[D]_{i,i} = \sum_{j=1}^{gN} \left[A^t\right]_{i,j}$. Computing Y^{G*} in (9) is equivalent to fining eigen vectors of Y^{G*} as $L^G Y^{G*} = \alpha D Y^{G*}$. The spectral embedding Y^{G*} can also be calculated by simply computing the eigen values of L^G. Finally, the laplacian eigen maps L^G and spectral embedding Y^{G*} are used to compute the cost function to find the mixing coefficients as

$$\lambda^i = \frac{tr\left(\left(Y^G\right)^T L^i Y^{G*}\right)}{\sum_{i=1}^{V} tr\left(\left(Y^G\right)^T L^i Y^{G*}\right)} \tag{10}$$

Overall, the convergence of (10) can be decided based on the $l2$ norm between iterations as

$$\sqrt{\sum_{i=1}^{V} \left(\lambda_k^i - \lambda_{k-1}^i\right)^2} < \delta \tag{11}$$

Here, λ_k^i is the value of mixing coefficients at k^{th} iteration and λ_{k-1}^i is the value at $(k-1)^{th}$ iteration. The constant δ is a user defined parameter less than 1. Eventually, the value of λ^i will be different from $\frac{1}{V}$where multiple views are contributing differently to the target view. Finally, by multiplying the obtained mixing coefficients with target features from different views, we obtain a global view feature that closely relates to the target view features. Furthermore, the resulting single view target view feature is highly discriminative across classes and has found have close proximity with all the views from within a class label. The following section describes the datasets and experiments conducted to ascertain the performance of the proposed method.

IV. Experimentation

The proposed view invariant method, Deep Metric Encoder Decoder (DMED) was trained and tested on multi view skeletal sign (action) video datasets in multiple ratios. We present a one – to – one, one – to – many, many – to – one and many – to – many cross view training and testing approaches on DMED. Further, we compare the results of our approach with other state – of – the – art multi view methods. Finally, multiple CNNs architecture's for classification were tested to check the robustness of the proposed feature extraction process in generating view invariant features.

A. Skeletal Video Datasets and Evaluation Metrics

The multi view sign language dataset KLEF3DSL_2Dskeletal with $V = 15$ views, 200 classes is generated at KL Biomechanics and Vision Computing Research Centre using 3D motion capture technology [8]. Further, the proposed model is evaluated on multi view benchmark skeletal action datasets such as NTU RGB-D [9], SBU Kinect Interaction [10], KLYoga3D [11] and KL3D_MVaction [12]. A small subset of data sample from KLEF3DSL_2Dskeletal is presented in Fig. 4 for a sign basketball. In this work we are limiting our views to 15 due to computational constraints. The training testing ratios are kept constant across all networks and datasets. The selected train test ratios are one – to – one and one – to – many. The remaining views were also evaluated but are not presented here as they have not produced any noticeable performance changes when compared to the selected ones. Since there are no multi view sign language datasets, we evaluated our model on multi view benchmark action datasets. Despite the availability of huge classes in action datasets, we selected only 40 action classes for training with 15 views from each class for maintaining uniformity during comparison. In some cases, unavailability of views has prompted us to generate random views by altering the viewing angles of joints. Here, the evaluation is performed independent of the type of view in which the action is recorded. Fig. 5(a), (b) and (c) shows samples from NTU RGB-D, KL3D_MVaction and KLYoga3D dataset respectively. We used mean recognition accuracy (mRA) for performance evaluations.

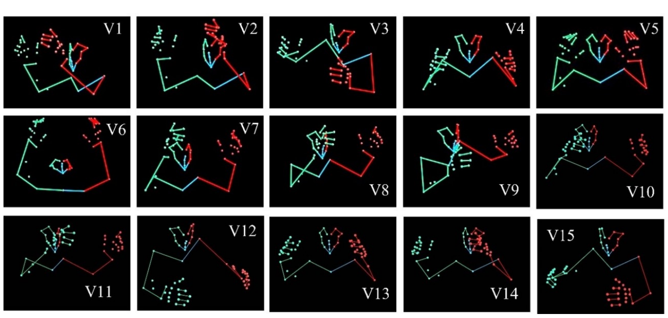


Fig. 4. KLEF3DSL_2Dskeletal Sign Language Video Dataset. A Sample Frame in 15 Different Views from the Skeletal Video Sign "Basketball".

The first 3D CNN network in Fig. 2 extracts the features from skeletal sign (action) video datasets. The network in Fig. 2 is trained on all the available views with similar hyper parameters except for the learning rate and number of epochs. The learning rate for KLEF3DSL_2Dskeletal sign language video dataset is 0.001 and it was 0.005 for all other action datasets. However, the KLYoga3D was trained on a learning rate of 0.0001 for 200 epochs due to large number of skeletal joints. The remaining datasets were trained for 150 epochs. The maximum recognition accuracy achieved during training was around 0.973 for KLEF3DSL_2Dskeletal sign language, 0.942 for NTU RGB-D, 0.845 for SBU Kinect Interaction, 0.902 for KLYoga3D and 0.985 for KL3D_MVaction datasets respectively. Consequently, these individual view trained 3D CNN streams will be inferenced for all dataset samples to generate global view features which represent all views within a class label.

To accomplish the proposed objectives of MVSTFE, we select a target view from each class for inferencing on the trained 3D CNN in Fig. 2 as shown in Fig. 3. The output of Fig. 3 are the features extracted from each of the individual views for the inputted target view. These target view features are combined using the non – uniform linear combiner by computing the value of linear combination value λ^i using spectral embedding of Laplacian eigenmaps. The hyperparameter $(\delta)$for MVSTFE on KLEF3DSL_2Dskeletal$(\delta = 0.54)$, NTU RGB-D$(\delta = 0.71)$, SBU Kinect Interaction $(\delta = 0.94)$, KLYoga3D $(\delta = 0.83)$ and KL3D_MVaction $(\delta = 0.57)$ is selected iteratively. Finally, the generated combined view target features are used for classification. Specifically, to test the robustness of the features in the classification process, we standardized it by training and inferencing on benchmark CNN architectures. However, these architectures are miniaturized in layers and depth to source the feature inputs of size 100×100. Moreover, the regular 2D Convolutional layers in these models were replaced with 3D layers. This has been done to directly extract spatio temporal features from the network. To demonstrate the actual usefulness of these view invariant features, which resulted in the formulation of multiple performance evaluation procedures on the classifier as presented in the following sections.

B. One – to – One Classifier Performance Evaluation

The one – to – one cross view recognition experiment is conducted by training the classifier in Fig. 6. with one view global target feature representing all views and inferencing on a different views. Specifically, the key aspect of this experiment is to test the robustness of the generated view invariant features in estimating a class label based on its constituent views on which it is formulated. To demonstrate this, we designed a CNN network inspired from VGG-16 with 6 convolutional layers, 3 maximum pooling, one flatten and 2 dense layers. The network is trained with the generated view invariant features in each class and tested with view specific features. Consequently, we selected the learning rate of 0.01 for this network with categorical cross entropy loss and Adam optimizer. Subsequently, the above procedure is repeated for all datasets with the same hyper parameters. Furthermore, three benchmark architectures such as Inception – V4, GoogleNet and ResNet – 50 were trained and tested. However, vanishing gradients and overfitting problem were eliminated by re-designing the

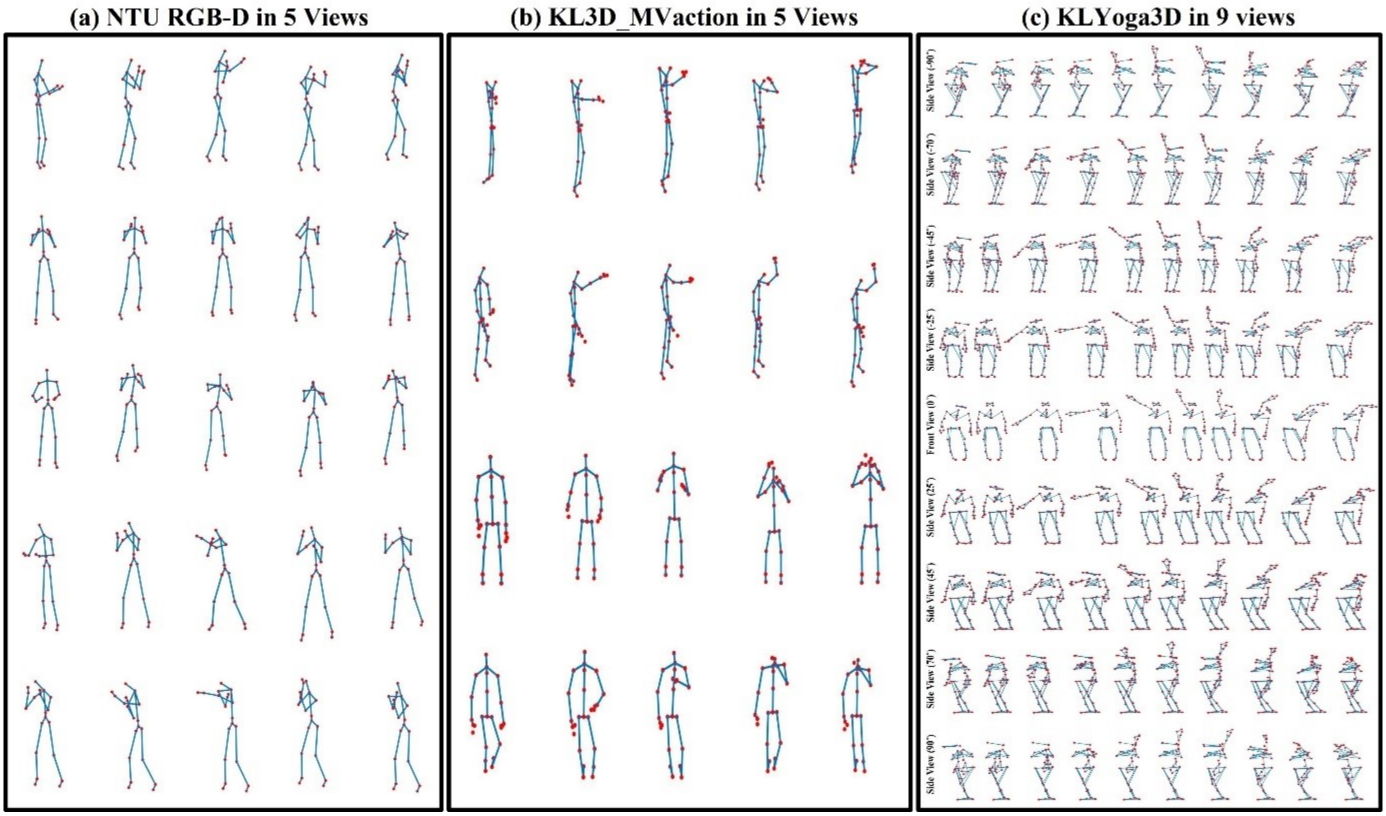


Fig. 5. Multi View Benchmark Action Datasets for Model Evaluation.

architectures with only half the layers than the original models. On the other hand, the structure of the original models were preserved to achieve highest performance. Eventually, mRA is computed during inferencing and the 10-fold maximum value is presented in Table I for all the datasets.

After examining the mRA in Table I, it is evident that all the models perform well on test views that have more visual information when compared to views with overlapping joints. The outcomes from Table I also suggests that the view target global features have shown to reduce false positives in all

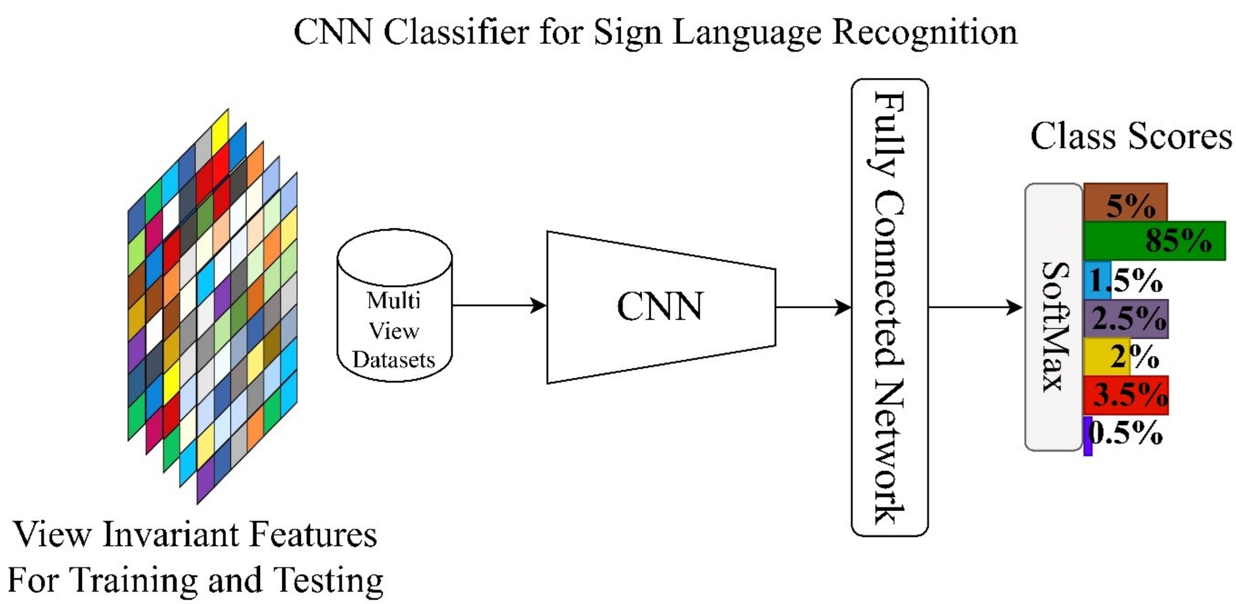


Fig. 6. CNN Architecture for Classification.

TABLE I. ONE – TO – ONE PERFORMANCE EVALUATION OF THE SELECTED CLASSIFIERS ON SKELETAL VIDEO DATASETS TRAINED WITH THE ONE VIEW OF TARGET VIEW FEATURE AND TESTED WITH ALL SPECIFIC VIEW FEATURES. THE PERFORMANCE OF THE CLASSIFIER IS MEASURED USING MRA

Classifiers	Views Datasets	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15
Tiny VGG – 16	KLEF3DSL_2Dskeletal	0.599	0.619	0.634	0.615	0.664	0.604	0.614	0.559	0.544	0.552	0.554	0.635	0.666	0.596	0.622
	NTU RGB-D	0.578	0.631	0.64	0.62	0.676	0.622	0.6	0.601	0.599	0.624	0.594	0.611	0.662	0.602	0.661
	SBU Kinect Interaction	0.58	0.593	0.589	0.601	0.584	0.591	0.595	0.531	0.529	0.5	0.52	0.566	0.619	0.602	0.59
	KLYoga3D	0.622	0.629	0.663	0.652	0.698	0.652	0.641	0.595	0.601	0.629	0.619	0.689	0.701	0.601	0.622
	KL3D_MVaction	0.619	0.604	0.601	0.629	0.609	0.59	0.616	0.576	0.563	0.565	0.559	0.626	0.649	0.601	0.62
Inception - V4	KLEF3DSL_2Dskeletal	0.671	0.691	0.706	0.687	0.736	0.676	0.686	0.631	0.616	0.624	0.626	0.707	0.738	0.668	0.694
	NTU RGB-D	0.65	0.703	0.712	0.692	0.748	0.694	0.672	0.673	0.671	0.696	0.666	0.683	0.734	0.674	0.733
	SBU Kinect Interaction	0.652	0.665	0.661	0.673	0.656	0.663	0.667	0.603	0.601	0.572	0.592	0.638	0.691	0.674	0.662
	KLYoga3D	0.694	0.701	0.735	0.724	0.77	0.724	0.713	0.667	0.673	0.701	0.691	0.761	0.773	0.673	0.694
	KL3D_MVaction	0.691	0.676	0.673	0.701	0.681	0.662	0.688	0.648	0.635	0.637	0.631	0.698	0.721	0.673	0.692
GoogleNet	KLEF3DSL_2Dskeletal	0.708	0.728	0.743	0.724	0.773	0.713	0.723	0.668	0.653	0.661	0.663	0.744	0.775	0.705	0.731
	NTU RGB-D	0.687	0.74	0.749	0.729	0.785	0.731	0.709	0.71	0.708	0.733	0.703	0.72	0.771	0.711	0.77
	SBU Kinect Interaction	0.689	0.702	0.698	0.71	0.693	0.7	0.704	0.64	0.638	0.609	0.629	0.675	0.728	0.711	0.699
	KLYoga3D	0.731	0.738	0.772	0.761	0.807	0.761	0.75	0.704	0.71	0.738	0.728	0.798	0.81	0.71	0.731
	KL3D_MVaction	0.728	0.713	0.71	0.738	0.718	0.699	0.725	0.685	0.672	0.674	0.668	0.735	0.758	0.71	0.729
ResNet - 50	KLEF3DSL_2Dskeletal	0.665	0.718	0.727	0.707	0.763	0.709	0.687	0.688	0.686	0.711	0.681	0.698	0.749	0.689	0.748
	NTU RGB-D	0.667	0.68	0.676	0.688	0.671	0.678	0.682	0.618	0.616	0.587	0.607	0.653	0.706	0.689	0.677
	SBU Kinect Interaction	0.709	0.716	0.75	0.739	0.785	0.739	0.728	0.682	0.688	0.716	0.706	0.776	0.788	0.688	0.709
	KLYoga3D	0.706	0.691	0.688	0.716	0.696	0.677	0.703	0.663	0.65	0.652	0.646	0.713	0.736	0.688	0.707
	KL3D_MVaction	0.735	0.755	0.77	0.751	0.8	0.74	0.75	0.695	0.68	0.688	0.69	0.771	0.802	0.732	0.758

TABLE II. MANY – TO – ONE PERFORMANCE EVALUATION OF THE CLASSIFIERS TRAINED WITH MULTIPLE SETS OF TRAINING VIEWS AND TESTED WITH ONLY ONE TARGET VIEW FEATURE GENERATED USING MVSTFE

Classifiers	Training Views Datasets	1	1	3	4	5	6	7	8	9	10	11	12	13	14	15
Tiny VGG – 16	KLEF3DSL_2Dskeletal	0.597	0.602	0.612	0.613	0.662	0.67	0.674	0.698	0.71	0.731	0.757	0.767	0.817	0.838	0.871
	NTU RGB-D	0.587	0.607	0.623	0.65	0.67	0.696	0.707	0.727	0.748	0.77	0.794	0.812	0.84	0.852	0.893
	SBU Kinect Interaction	0.58	0.596	0.608	0.605	0.617	0.628	0.66	0.687	0.703	0.724	0.756	0.785	0.816	0.839	0.866
	KLYoga3D	0.585	0.599	0.613	0.63	0.65	0.664	0.674	0.699	0.716	0.738	0.77	0.788	0.797	0.827	0.872
	KL3D_MVaction	0.586	0.608	0.612	0.617	0.65	0.657	0.667	0.7	0.727	0.737	0.76	0.77	0.8	0.832	0.861
Inception - V4	KLEF3DSL_2Dskeletal	0.652	0.657	0.667	0.668	0.717	0.725	0.729	0.732	0.744	0.765	0.791	0.801	0.851	0.872	0.905
	NTU RGB-D	0.642	0.662	0.678	0.705	0.725	0.751	0.762	0.761	0.782	0.804	0.828	0.846	0.874	0.886	0.927
	SBU Kinect Interaction	0.635	0.651	0.663	0.66	0.672	0.683	0.715	0.721	0.737	0.758	0.79	0.819	0.85	0.873	0.9
	KLYoga3D	0.64	0.654	0.668	0.685	0.705	0.719	0.729	0.733	0.75	0.772	0.804	0.822	0.831	0.861	0.906
	KL3D_MVaction	0.641	0.663	0.667	0.672	0.705	0.712	0.722	0.734	0.761	0.771	0.794	0.804	0.834	0.866	0.895
GoogleNet	KLEF3DSL_2Dskeletal	0.62	0.625	0.635	0.636	0.685	0.701	0.705	0.708	0.72	0.741	0.767	0.786	0.836	0.857	0.89
	NTU RGB-D	0.61	0.63	0.646	0.673	0.693	0.727	0.738	0.737	0.758	0.78	0.804	0.831	0.859	0.871	0.912
	SBU Kinect Interaction	0.603	0.619	0.631	0.628	0.64	0.659	0.691	0.697	0.713	0.734	0.766	0.804	0.835	0.858	0.885
	KLYoga3D	0.608	0.622	0.636	0.653	0.673	0.695	0.705	0.709	0.726	0.748	0.78	0.807	0.816	0.846	0.891
	KL3D_MVaction	0.609	0.631	0.635	0.64	0.673	0.688	0.698	0.71	0.737	0.747	0.77	0.789	0.819	0.851	0.88
ResNet - 50	KLEF3DSL_2Dskeletal	0.601	0.606	0.616	0.617	0.666	0.674	0.678	0.702	0.714	0.735	0.761	0.771	0.821	0.842	0.875
	NTU RGB-D	0.591	0.611	0.627	0.654	0.674	0.7	0.711	0.731	0.752	0.774	0.798	0.816	0.844	0.856	0.897
	SBU Kinect Interaction	0.584	0.6	0.612	0.609	0.621	0.632	0.664	0.691	0.707	0.728	0.76	0.789	0.82	0.843	0.87
	KLYoga3D	0.589	0.603	0.617	0.634	0.654	0.668	0.678	0.703	0.72	0.742	0.774	0.792	0.801	0.831	0.876
	KL3D_MVaction	0.59	0.612	0.616	0.621	0.654	0.661	0.671	0.704	0.731	0.741	0.764	0.774	0.804	0.836	0.865

classes. Moreover, the proposed work also highlights the used of any single view for testing as against the previous models, where all views are required as input. Consequently, it will be interesting to test the many – to – one cross view performance, where the models are trained with view specific features and tested with only one target view invariant feature.

C. Many – to – One Classifier Performance Evaluation

Here, we train the classifiers with all the views and test it only one target view feature. Table II shows mRA values for multiple sets of training views. The results in Table II show that the performance of the MVSTFE model has increased when trained with multiple view features. On the other hand, Inception – V4 has shown to outperform all other classifiers used for experimentation due to the fact that it contains multiple attention layers for selecting maximally contributing vectors.

D. Comparisons against other View Invariant Generation Techniques

The previous models applied spectral clustering with matrix factorization [28], auto – weighted spectral clustering [7] and multi view temporal ensemble [6] are designed to generate complimentary views and correspondingly reconstructing a global view. Additionally, the number of views used in these models is comparatively lower than our proposed work. Increasing the number of views in the above models will increase the computational complexity, which was reduced in MVSTFE. Table III presents the comparisons of the above multi view recognition methods with MVSTFE.

E. Validation of MVSTFE with State – of – the – Art Multi View Methods

Historical validation of the proposed MVSTFE is performed by comparing it with state – of – the – art multi view methods in Table IV. The methods selected for comparison have applied some kind of deep learning algorithms

TABLE III. PRESENTS THE RESULTS OF [28], [7] AND [6] ALONG WITH OUR PROPOSED MVSTFE MODEL ON BENCHMARK DATASETS

Multi View Algorithms	Classifiers / Train Test Methods / Datasets	Tiny VGG – 16		Inception - V4		GoogleNet		ResNet – 50	
		One – to – one	Many – to – one	One – to – one	Many – to – one	One – to – one	Many – to – one	One – to – one	Many – to – one
Spectral clustering via structured low-rank matrix factorization [28]	KLEF3DSL_2Dskeletal	0.526	0.621	0.598	0.716	0.524	0.675	0.568	0.666
	NTU RGB-D	0.558	0.719	0.627	0.755	0.602	0.736	0.594	0.714
	SBU Kinect Interaction	0.504	0.657	0.575	0.67	0.549	0.648	0.551	0.643
	KLYoga3D	0.559	0.696	0.638	0.777	0.617	0.749	0.611	0.728
	KL3D_MVaction	0.539	0.652	0.611	0.711	0.586	0.7	0.587	0.696
Auto-weighted multi-view clustering via spectral embedding [7]	KLEF3DSL_2Dskeletal	0.623	0.718	0.67	0.813	0.621	0.752	0.665	0.763
	NTU RGB-D	0.655	0.816	0.724	0.852	0.699	0.833	0.691	0.811
	SBU Kinect Interaction	0.601	0.754	0.672	0.767	0.646	0.745	0.648	0.74
	KLYoga3D	0.656	0.793	0.735	0.874	0.714	0.846	0.708	0.825
	KL3D_MVaction	0.636	0.749	0.708	0.808	0.683	0.797	0.684	0.793
Multi-view temporal ensemble [6]	KLEF3DSL_2Dskeletal	0.549	0.674	0.591	0.749	0.554	0.675	0.568	0.671
	NTU RGB-D	0.581	0.702	0.62	0.768	0.599	0.738	0.595	0.715
	SBU Kinect Interaction	0.527	0.652	0.575	0.679	0.552	0.657	0.544	0.648
	KLYoga3D	0.592	0.693	0.637	0.778	0.612	0.751	0.607	0.728
	KL3D_MVaction	0.572	0.65	0.607	0.721	0.589	0.7	0.585	0.692
MVSTFE Proposed	KLEF3DSL_2Dskeletal	0.668	0.793	0.71	0.868	0.673	0.794	0.687	0.79
	NTU RGB-D	0.7	0.821	0.739	0.887	0.718	0.857	0.714	0.834
	SBU Kinect Interaction	0.646	0.771	0.694	0.798	0.671	0.776	0.663	0.767
	KLYoga3D	0.711	0.812	0.756	0.897	0.731	0.87	0.726	0.847
	KL3D_MVaction	0.691	0.769	0.726	0.84	0.708	0.819	0.704	0.811

TABLE IV. COMPARISON AMONG DIFFERENT VIEW-BASED RECOGNITION TECHNIQUES

		V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15
NTU RGB+D	[16]	0.608	0.59	0.574	0.61	0.645	0.659	0.574	0.55	0.601	0.572	0.579	0.568	0.581	0.61	0.568
	[17]	0.617	0.599	0.583	0.619	0.654	0.668	0.583	0.559	0.61	0.581	0.588	0.577	0.567	0.604	0.638
	[18]	0.587	0.569	0.553	0.589	0.624	0.638	0.553	0.529	0.58	0.551	0.558	0.547	0.541	0.578	0.612
	[19]	0.616	0.598	0.581	0.618	0.652	0.667	0.581	0.557	0.608	0.579	0.586	0.575	0.576	0.612	0.647
	[21]	0.59	0.572	0.555	0.592	0.626	0.641	0.555	0.532	0.582	0.553	0.56	0.549	0.581	0.618	0.652
	[25]	0.625	0.607	0.59	0.626	0.661	0.675	0.59	0.566	0.617	0.588	0.595	0.584	0.57	0.606	0.641
	MVSTFE	0.723	0.776	0.785	0.765	0.821	0.767	0.745	0.746	0.744	0.769	0.739	0.756	0.807	0.747	0.806
SBU Kinect Interaction	[16]	0.63	0.612	0.595	0.632	0.666	0.681	0.595	0.572	0.623	0.593	0.6	0.589	0.579	0.615	0.65
	[17]	0.628	0.61	0.594	0.63	0.665	0.679	0.594	0.57	0.621	0.592	0.599	0.588	0.549	0.585	0.62
	[18]	0.637	0.619	0.603	0.639	0.674	0.688	0.603	0.579	0.63	0.601	0.608	0.597	0.577	0.614	0.648
	[19]	0.607	0.589	0.573	0.609	0.644	0.658	0.573	0.549	0.6	0.571	0.578	0.567	0.551	0.588	0.622
	[21]	0.636	0.618	0.601	0.638	0.672	0.687	0.601	0.577	0.628	0.599	0.606	0.595	0.586	0.622	0.657
	[25]	0.61	0.592	0.575	0.612	0.646	0.661	0.575	0.552	0.602	0.573	0.58	0.569	0.591	0.628	0.662
	MVSTFE	0.725	0.738	0.734	0.746	0.729	0.736	0.74	0.676	0.674	0.645	0.665	0.711	0.764	0.747	0.735
KLYoga3D	[16]	0.645	0.627	0.61	0.646	0.681	0.695	0.61	0.586	0.637	0.608	0.615	0.604	0.53	0.566	0.601
	[17]	0.65	0.632	0.615	0.652	0.686	0.701	0.615	0.592	0.643	0.613	0.62	0.609	0.539	0.575	0.61
	[18]	0.638	0.62	0.604	0.64	0.675	0.689	0.604	0.58	0.631	0.602	0.609	0.598	0.509	0.545	0.58
	[19]	0.647	0.629	0.613	0.649	0.684	0.698	0.613	0.589	0.64	0.611	0.618	0.607	0.537	0.574	0.608
	[21]	0.617	0.599	0.583	0.619	0.654	0.668	0.583	0.559	0.61	0.581	0.588	0.577	0.511	0.548	0.582
	[25]	0.646	0.628	0.611	0.648	0.682	0.697	0.611	0.587	0.638	0.609	0.616	0.605	0.546	0.582	0.617
	MVSTFE	0.767	0.774	0.808	0.797	0.843	0.797	0.786	0.74	0.746	0.774	0.764	0.834	0.846	0.746	0.767
KL3D_MVaction	[16]	0.62	0.602	0.585	0.622	0.656	0.671	0.585	0.562	0.612	0.583	0.59	0.579	0.551	0.588	0.622
	[17]	0.655	0.637	0.62	0.656	0.691	0.705	0.62	0.596	0.647	0.618	0.625	0.614	0.586	0.622	0.657
	[18]	0.66	0.642	0.625	0.662	0.696	0.711	0.625	0.602	0.653	0.623	0.63	0.619	0.591	0.628	0.662
	[19]	0.598	0.58	0.564	0.6	0.635	0.649	0.564	0.54	0.591	0.562	0.569	0.558	0.53	0.566	0.601
	[21]	0.607	0.589	0.573	0.609	0.644	0.658	0.573	0.549	0.6	0.571	0.578	0.567	0.539	0.575	0.61
	[25]	0.577	0.559	0.543	0.579	0.614	0.628	0.543	0.519	0.57	0.541	0.548	0.537	0.509	0.545	0.58
	MVSTFE	0.764	0.749	0.746	0.774	0.754	0.735	0.761	0.721	0.708	0.71	0.704	0.771	0.794	0.746	0.765
KLEF3DSL_2Dskeletal	[16]	0.606	0.588	0.571	0.608	0.642	0.657	0.571	0.547	0.598	0.569	0.576	0.565	0.537	0.574	0.608
	[17]	0.58	0.562	0.545	0.582	0.616	0.631	0.545	0.522	0.572	0.543	0.55	0.539	0.511	0.548	0.582
	[18]	0.615	0.597	0.58	0.616	0.651	0.665	0.58	0.556	0.607	0.578	0.585	0.574	0.546	0.582	0.617
	[19]	0.62	0.602	0.585	0.622	0.656	0.671	0.585	0.562	0.613	0.583	0.59	0.579	0.551	0.588	0.622
	[21]	0.583	0.565	0.548	0.585	0.619	0.634	0.548	0.524	0.575	0.546	0.553	0.542	0.514	0.551	0.585
	[25]	0.557	0.539	0.522	0.559	0.593	0.608	0.522	0.499	0.549	0.52	0.527	0.516	0.488	0.525	0.559
	MVSTFE	0.744	0.764	0.779	0.76	0.809	0.749	0.759	0.704	0.689	0.697	0.699	0.78	0.811	0.741	0.767

for generation and classification of view video data. Since the data used in these methods were different, we recreated these models from scratch as given in their respective manuscripts. All the experiments were conducted on the benchmark skeletal datasets used in this work with one – to – one train – test pattern. We presented our best result obtained from inception V4 classifier in this comparison. However, the hyper parameters for the comparison networks was adopted from our Inception V4. The proposed MVSTFE has outperformed the existing models as can be seen in Table IV.

V. CONCLUSION

This work proposed a deep learning based spectral embedding method for generating a single global view from a set of multi view features. We trained a 3D CNN on each of the available views and inferencing on a target view video data to extract features. Eventually, these target features are combined linearly by calculating the mixing coefficients for making a global feature representation for all possible views. Consequently, the mixing coefficients are computed using spectral embedding in Laplacian eigen space which preserves proximity between views within the class label. Experimentation has shown that the proposed MVSTEF on 2D video based skeletal sign language dataset and the benchmark action datasets has outperformed the previous multiview baseline models.

REFERENCES

[1] Razieh Rastgoo, Kourosh Kiani, and Sergio Escalera. Sign language recognition: A deep survey. *Expert Systems with Applications*, 164:113794, 2021.

[2] Imen Jegham, Anouar Ben Khalifa, Ihsen Alouani, and Mohamed Ali Mahjoub. Vision-based human action recognition: An overview and real world challenges. *Forensic Science International: Digital Investigation*, 32:200901, 2020.

[3] Eepuri Kiran Kumar, PVV Kishore, Maddala Teja Kiran Kumar, Dande Anil Kumar, and ASCS Sastry. Three-dimensional sign language recognition with angular velocity maps and connived feature resnet. *IEEE Signal Processing Letters*, 25(12):1860–1864, 2018.

[4] Jun Liu, Gang Wang, Ling-Yu Duan, Kamila Abdiyeva, and Alex C Kot. Skeleton-based human action recognition with global context-aware attention lstm networks. *IEEE Transactions on Image Processing*, 27(4):1586–1599, 2017.

[5] E Kiran Kumar, PVV Kishore, M Teja Kiran Kumar, and D Anil Kumar. 3d sign language recognition with joint distance and angular coded color topographical descriptor on a 2–stream cnn. *Neurocomputing*, 372:40–54, 2020.

[6] Bee Hock David Koh and Wai Lok Woo. Multi-view temporal ensemble for classification of non-stationary signals. *IEEE Access*, 7:32482–32491, 2019.

[7] Shaojun Shi, Feiping Nie, Rong Wang, and Xuelong Li. Auto-weighted multi-view clustering via spectral embedding. *Neurocomputing*, 399:369–379, 2020.

[8] PVV Kishore, D Anil Kumar, AS Chandra Sekhara Sastry, and E Kiran Kumar. Motionlets matching with adaptive kernels for 3-d indian sign language recognition. *IEEE Sensors Journal*, 18(8):3327–3337, 2018.

[9] Amir Shahroudy, Jun Liu, Tian-Tsong Ng, and Gang Wang. Ntu rgb+ d: A large scale dataset for 3d human activity analysis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1010–1019, 2016.

[10] Meng Li and Howard Leung. Multiview skeletal interaction recognition using active joint interaction graph. *IEEE Transactions on Multimedia*, 18(11):2293–2302, 2016.

[11] Teja Kiran Kumar Maddala, PVV Kishore, Kiran Kumar Eepuri, and Anil Kumar Dande. Yoganet: 3-d yoga asana recognition using joint angular displacement maps with convnets. *IEEE Transactions on Multimedia*, 21(10):2492–2503, 2019.

[12] D Srihari, PVV Kishore, E Kiran Kumar, D Anil Kumar, M Kumar, MVD Prasad, Ch Prasad, et al. A four-stream convnet based on spatial and depth flow for human action classification using rgb-d data. *Multimedia Tools and Applications*, 79(17):11723–11746, 2020.

[13] Neena Aloysius and M Geetha. Understanding vision-based continuous sign language recognition. *Multimedia Tools and Applications*, 79(31):22177–22209, 2020.

[14] Yuling Xing and Jia Zhu. Deep learning-based action recognition with 3d skeleton: A survey, 2021.

[15] Tanveer Hussain, Khan Muhammad, Weiping Ding, Jaime Lloret, Sung Wook Baik, and Victor Hugo C de Albuquerque. A comprehensive survey of multi-view video summarization. *Pattern Recognition*, 109:107567, 2021.

[16] Zan Gao, Hua Zhang, GP Xu, YB Xue, and Alexander G Hauptmann. Multi-view discriminative and structured dictionary learning with group sparsity for human action recognition. *Signal Processing*, 112:83–97, 2015.

[17] Jingjing Zheng, Zhuolin Jiang, and Rama Chellappa. Cross-view action recognition via transferable dictionary learning. *IEEE Transactions on Image Processing*, 25(6):2542–2556, 2016.

[18] Pengfei Zhang, Cuiling Lan, Junliang Xing, Wenjun Zeng, Jianru Xue, and Nanning Zheng. View adaptive recurrent neural networks for high performance human action recognition from skeleton data. In *Proceedings of the IEEE international conference on computer vision*, pages 2117–2126, 2017.

[19] Amin Ullah, Khan Muhammad, Tanveer Hussain, and Sung Wook Baik. Conflux lstms network: A novel approach for multi-view action recognition. *Neurocomputing*, 435:321–329, 2021.

[20] Chenyang Si, Wentao Chen, Wei Wang, Liang Wang, and Tieniu Tan. An attention enhanced graph convolutional lstm network for skeleton-based action recognition. In *proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1227–1236, 2019.

[21] Dongang Wang, Wanli Ouyang, Wen Li, and Dong Xu. Dividing and aggregating network for multi-view action recognition. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 451–467, 2018.

[22] PVV Kishore, Manoj VD Prasad, Ch Raghava Prasad, and R Rahul. 4-camera model for sign language recognition using elliptical fourier descriptors and ann. In *2015 International Conference on Signal Processing and Communication Engineering Systems*, pages 34–38. IEEE, 2015.

[23] Imran N Junejo, Emilie Dexter, Ivan Laptev, and Patrick Púrez. Cross-view action recognition from temporal self-similarities. In *European Conference on Computer Vision*, pages 293–306. Springer, 2008.

[24] Hossein Rahmani and Ajmal Mian. Learning a non-linear knowledge transfer model for cross-view action recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2458–2466, 2015.

[25] Hossein Rahmani, Arif Mahmood, Du Huynh, and Ajmal Mian. Histogram of oriented principal components for cross-view action recognition. *IEEE transactions on pattern analysis and machine intelligence*, 38(12):2430–2443, 2016.

[26] Jiang Wang, Xiaohan Nie, Yin Xia, Ying Wu, and Song-Chun Zhu. Cross-view action modeling, learning and recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2649–2656, 2014.

[27] Lichen Wang, Zhengming Ding, Zhiqiang Tao, Yunyu Liu, and Yun Fu. Generative multi-view human action recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6212–6221, 2019.

[28] Yang Wang, Lin Wu, Xuemin Lin, and Junbin Gao. Multiview spectral clustering via structured low-rank matrix factorization. *IEEE transactions on neural networks and learning systems*, 29(10):4833–4843, 2018.

[29] Chuan Sun, Imran Nazir Junejo, Marshall Tappen, and Hassan Foroosh. Exploring sparseness and self-similarity for action recognition. *IEEE Transactions on Image Processing*, 24(8):2488–2501, 2015.

[30] Yu Kong, Zhengming Ding, Jun Li, and Yun Fu. Deeply learned view-invariant features for cross-view action recognition. *IEEE Transactions on Image Processing*, 26(6):3028–3037, 2017.

Human Activity Recognition in Car Workshop

Omar Magdy
Faculty of Computer Science,
October University for Modern
Sciences and Arts (MSA),
Giza, Egypt

Ayman Atia
HCI-LAB, Faculty of Computers and Artificial Intelligence,
Helwan University.
Faculty of Computer Science,
October University for Modern
Sciences and Arts (MSA), Giza, Egypt

***Abstract*—Human activity recognition has become so widespread in recent times. Due to the modern advancements of technology, it has become an important solution to many problems in various fields such as medicine, industry, and sports. And this subject got the attention of a lot of researchers. Along with problems like wasted time in maintenance centers, we proposed a system that extracts worker poses from videos by using pose classification. In this paper, we have tested two algorithms to detect worker activity. This system aims to detect and classify positive and negative worker's activities in car maintenance centers such as (changing the tire, changing oil, using the phone, standing without work). We have conducted two experiments, the first experiment was for comparison between algorithms to determine the most accurate algorithm in recognizing the activities performed. The experiment was done using two different algorithms (1 dollar recognizer and Fast Dynamic time warping) on 3 participants in a controlled area. The one-dollar recognizer has achieved a 97% accuracy compared to the fastDTW with 86%. The second experiment was conducted to measure the performance of a one-dollar algorithm with different participants. The results show that a 1 dollar recognizer achieved an accuracy of 94.2% when tested on 420 different videos.**

***Keywords*—Machine learning; human activity recognition; pose identification; industry analysis**

I. Introduction

Human activity recognition has gained importance in recent years because of its applications in various fields like health, security and surveillance, entertainment, and intelligent environments. Human activity recognition accounted a lot of researches in different approaches, like wearable devices [1][2][3][4][5],object-tagged[6][7][8], and device-free [9][10][11], to acknowledge human activities. Wasted time has become a prominent problem in various fields of work, and this problem affects the percentage of products and services required in our daily lives. The survey provided by Salary.com [12] found that 89% of workers admitted to wasting time at work every day. The survey showed that 61% claim to waste between 30 minutes to an hour a day. While this may not seem like much, it can add up to 5 hours a week or 260 hours a year/per employee. Maintenance places and factories are the most vulnerable to this issue. The survey depend heavily on human activity, which harms production rates and disrupts many required services. The field of car maintenance has become an important field in our daily life, and car workshops have spread greatly in the recent period, but difficulties have begun to appear inside the maintenance centers. The problem is that workers waste a lot of time while working by doing some negative activities. Fig. 1 shows a worker who uses the phone at work. Moreover, there are a lot of negative activities such as (eating, drinking, talking to others, standing without work...). Besides, it is difficult to have an employee responsible for monitoring the worker for no less than 12 hours.

Fig. 1. A Negative Wasting Time Activity (using Mobile while being in Work).

The maintenance workshop could have dozens of activities some of them are consider positive and other consider negative. However,some activities look very similar such as shown in Fig. 2(a) and Fig. 2(b). The posture of changing oil activity looks very similar in motion trajectory to engine rebuild activity.

(a) Changing oil (b) Engine rebuild

Fig. 2. A Sample for Positive Activity in Workshop.

The proposed system works by taking an input of recording videos of the worker's activities inside the maintenance center. The video is classified to determine the number of activities the worker performed such as (Changing the oil, changing the tire, or using the phone... Etc.). The classification was done by taking the path (x, y) of the important points in the skeleton as shown in Fig.3, Fig. 4 and comparing them with the points that were taken from the collector of the data set. The system can determine the type of activities that the worker performed in the video.

Fig. 3. Worker Changing Oil.

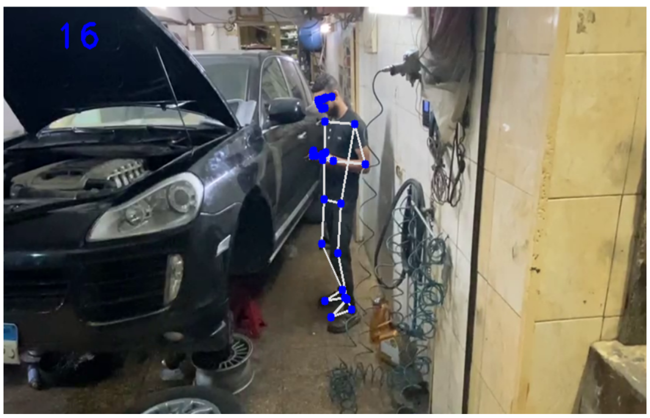


Fig. 4. Worker use Mobile.

The main contribution of this paper is to create a human activity recognition system that extract worker poses from the videos by using pose classification. The system used a dollar recognizer (1$) to detect worker positive and negative activities in maintenance workshop. As a result of the presence of many maintenance workshops, a problem occurred which is many workers neglect their work resulting in the occurrence of negative activities. Therefore, this paper is needed to detect the positive and negative activities and differentiate between them without the need for a person to monitor these activities and depend on the computer as an alternative. The purpose of this paper as well as to facilitate the reduction of the waste time done by each worker so they can be more productive. Moreover, positive activity is detected to reward the committed workers and to evaluate the performance of the workers in general.

II. Related Work

Due to the popularity of human activity recognition systems and with the rapid advancement of computer technology, lots of research efforts were dedicated to this subject. Congcong Liu et al. [6], has proposed a activity recognition method that can identify abnormal human activity recognition in surveillance video using a combination of Bayes Classifier and CNN (Convolution Neural Network) to detect the activities and also use KTH dataset as the input of Bayes Classifier and CNN. Another system that is able to recognize human activity is proposed by Bagate et al. [1], which identifies the activity using RGB-D sensors that is developed with deep learning model CNN (Convolution Neural Network) and using knight depth camera for capturing 3-D skeleton data. For Human detection and Motion tracking, Sandar et al. [8], used frame wise displacement and recognition is based on the skeletal model with the deep learning framework to understand human behavior in the indoor and outdoor environment. And in the field of Human activity recognition by sensors, Murat et al. [13], automatically identifies human activity using joint coordinates skeletons and uses two types of deep learning to make classification and use data set of multiple people in the images. Song et al. [5], propose (1D) Convolution Neural Network (CNN) -based method for recognizing the activities using collected accelerometer data from smartphones and this method gave high accuracy of 92.71. And using wearable devices, Tahera et al. [2], used the eSense accelerometer sensor to detect the matching of activity between the head and the mouth, from this collected data some activities of the head and mouth were identified and using the machine learning and deep learning for data classification. Nitin et al. [14], propose to use the TCN (temporal Convolutional Network) to recognize the activities because it better than other deep learning methods , it has strong ability to capture long-term dependencies. Godwin et al. [9], combined gyroscope sensors with accelerometers to detect human activity and perform analysis and recognition using ANN (artificial neural networks). Tsokov et al. [15], use of the 1D synaptic neural network (CNN) with accelerometer data to make recognition of human activity more accurate.

Isah et al. [3] collected hip motion from the different waist mounted sensors, and convert each signal into spectrum image and use them as input to the CNN (Convolution Neural Network). Nacer et al. [16], use entropy point estimate for 1D heat map to separate between human maps and animal maps to give high accuracy in human activity recognition. Selçuk et al. [10] used a novel design to reduce the number of sensors used for human activity recognition and detection by using (EMD) empirical mode decomposition.And Jiewen et al. [4], identified and interacted by focusing on two wearable cameras and the interactive activities that involve only two people. Peter Washington et al. [11], addressed the topic of identifying human activity in the treatment of autism and grouped movements with a handheld camera and used the classifier CNN (Convolution Neural Network) for detecting headbanging in home videos. Hristov et al. [17]proposed a method that classifies human activity by using 3D skeleton data and normalizing it beforehand and it was represented in 2 forms. They applied this method to the UTDMHAD dataset, the system has achieved a 92.4% accuracy rate. Heilym et al. [18] was opposed to the idea of wearing devices and sensors to determine human activities and pointed out that these devices could cause inconvenience to the bearer and could give false results if used in crowded places, so he relied on the camera and determining the activity through the human skeleton features. Salahuddin Saddar et al. [19]have an objective which is to compare some machine learning algorithms

TABLE I. COMPARISON WITH SIMILAR SYSTEMS

Paper	Activities	Algorithms	Methods	Area	Accuracy
[7]	stopping the furnace operating checking the solid fuel tank checking the gear motor and auger tightening the mounting screws of the gear motor	CNN CNN+SVM Yolov3	Image classification	controlled area	95.7%
[22]	grab tools hammers nail wrench use screwdriver	CNN	Image classification	controlled area	87%
This paper	changing tire changing oil use mobile stand without work	one-dollar Fastdtw	Pose classification	uncontrolled area	95%

that were used in human activity recognition such as (SVM, Decision Trees, Random Forests, XGBoost). They tested these algorithms with measurement sensor data that was recently released from the LARA dataset. The XGBoost has achieved the best accuracy with a rate of 78.6%. Ismael et al. [20] took the topic of identifying human activity in terms of reducing aggressive actions inside prisons and on the streets to reduce aggression and used "handcrafted/learned" as a hybrid feature framework that gave it very high accuracy rates. Yusuf Erkan et al. [21]used depth sensor to classify 27 different activities and by using long-short term memory, they analyzed skeleton data. It has achieved an accuracy rate 93%. Halikowski et al. [7], presented a system for monitoring activities inside the factory using (CNN, CNN+SVM, Yolov3) algorithms. They used some activities such as (stopping the furnace operating, checking the solid fuel tank, checking the gear motor and auger, tightening the mounting screws of the gear motor) and achieved an accuracy 94%. The work presented utilize deep learning for extracting features without considering human post estimation. Zhaozheng et al. [22] presented a system for detecting activities in smart manufacturing. They used some activities (grab tools, hammer nail, wrench use, rest lever, screwdriver). They captured these activities using IMU and sEMG signals obtained from a MYO armband. They extract feature using a convolutional neural network (CNN) model. The CNN model is evaluated on this data set and achieves 98% and 87% recognition accuracy in the half-half and leave-one-out experiments. All of the previous explained the importance of identifying human activity in solving some problems in various fields. Alghyaline et al. [23] has proposed system that detects different actions in the street such as (walking, running, stopping). They measure the movement type by using three different techniques which are (Yolo, Kalman filter, Homography). The method was tested by CCTV camera and BEHAVE dataset, it has achieved an accuracy of 96.9% for the Behave dataset and achieved 88.4% for the dataset that was collected by CCTV camera.Arzani et al. [24]proposed a structural prediction strategy proposed by this system to recognize the simple and complex actions by using probabilistic graphical models (PGMs). These activities require various model parametrization to be spanned, category-switching scheme is used to deal with this parametrization. Three datasets were used to cover the two action types which are (CAD-60, UT-Kinect, and Florence 3-D). This system could recognize simple and complex activities while the previous systems focused on only one type of these two. The system proposed by Archana et al.[25] recognized human activity with Resnet and 3D CNN without using the LSTM- attention model as the 3D CNN is achieved by modifying the 2D Resnet in order to achieve better accuracy, so that the development of detecting, and recognizing real-time human motion has been achieved.The system proposed by Zheng Dong et al. [26]resolves the issue of incomplete feature extraction by a new framework called CapsGaNet which proposed multi-feature extraction, and gated recurrent units (GRU) with attention mechanisms. The constructed dataset was a daily and aggressive activity dataset (DAAD). Moreover, the paper approved that CapsGaNet has efficiently improved the accuracy of recognition. Radhika V. et al. [27]proposed a system that used Random Forest Algorithm(RFA) to recognize human activity using Smartphones. RFA algorithm has different decision trees that is used in classification of the dataset. There were four various evaluation parameters used to measure the performance of the system such as F1 score, accuracy, precision, and sensitivity. The accuracy of the system achieved 98.34%. The system proposed by Navita et al. [28] detects the activity of aged people using the Internet of Things (IoT) monitoring model to monitor the activity of their health state. The SVM has attained 98.03%. The proposed system by Yin Tang et al. [29], a new CNN model that used hierarchical-split (HS) for a huge number of varieties in human activity recognition. Each one feature layer uses multi-scale feature representation by capturing a wide range of receptive fields of human activities. The proposed HS model can achieve high recognition performance compared to similar models complexities. The system achieved 94.10% SOTA accuracy on human activity recognition dataset.The proposed system by Maciej A. Noras et al. [30] discussed the topic of far-field electric field sensors, which accompany different physical events. The determination of activities in the proximity of the sensor is done by field signature signals. Moreover, the paper provided enhancements for electric field sensor usage and signal processing in human and animal motion recognition, perimeter monitoring, moving objects recognition, and electric power faults detection.

Table I shows a comparison between our system and different systems. The first system, Halikowski et al. [7] proposed this system to measure the performance of the worker in the factory, they recognized four different activities which are (stopping the furnace operating, checking the solid fuel tank, checking the gear motor and auger, tightening the mounting screws of the gear motor) in a controlled area using image classification method. They used more than one algorithm such as (CNN, CNN+SVM, Yolov3), their system has achieved a 95.7% accuracy rate. The second system which is proposed by Zhaozheng et al. [22] was used for qualification and evaluation of the workers. They also used image classification to detect four activities which are (grab tools, hammer nail, wrench use, rest lever, screwdriver) in a controlled area, and they achieved an accuracy rate of 87%. Our system was proposed to detect the activities of the worker inside the car maintenance center and differentiate between the negative and the positive activities. To help the workshop owner in measuring the worker performance, the system has used the pose classification method to detect four different activities which are (changing oil, changing tire, use mobile, stand without work). The system was used in an uncontrolled area and on different body characteristics such as(height and weight).

III. Proposed System

We presented a method that recognized and classified human activity in car workshops performed and captured from videos. It differentiated between the positive activities and the negative activities based on a comparison between input video and dataset stored in the templates. The proposed system used mediapipe for collecting key points of the skeletal joints and used the one-dollar and Fastdtw algorithms to classify the poses.

Fig. 5 shows the system overview, the system has two different way to input which is videos or live camera. The processing part starts with face recognition to differentiate between the workers because there is a large number of workers in the maintenance center. Then the mediapipe starts to extract the poses by calculating the path of each point in the skeleton. The mediapipe can extract 32 points, but this system focuses on extracting five important points which are (shoulder, elbow, wrist, hip, knee). The path of points was saved in a file to be ready for classification by the algorithms. The algorithms start to match these points with the points stored in the data set and send the results to the database to create a report that the managers and workshop owners can see.

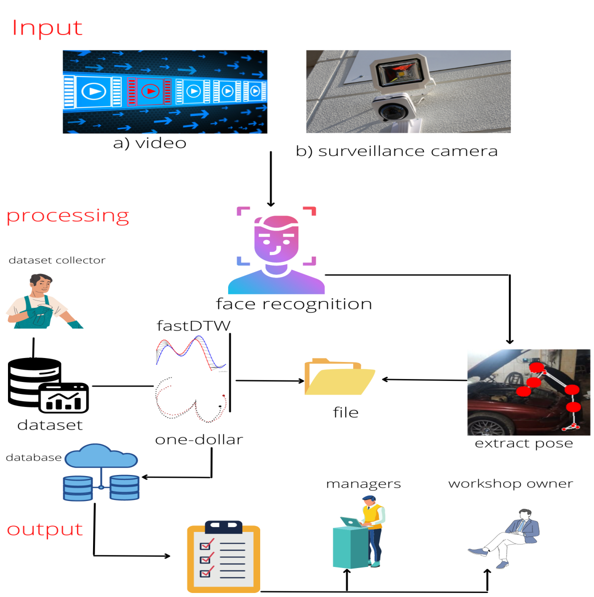


Fig. 5. System Overview.

The dataset was collected by recording videos of a specialist in the activities of the worker inside the maintenance workshop. These videos were entered into the mediapipe and OpenCV to extract the human pose from the videos showing the path of each point during the video. Each list of points has been saved in a file contains the points and the name of the activity. With the aid of a partner in the industry, we managed to collect 560 videos including the four activities which are (changing tires, changing oil, using a mobile, and standing without work), to start testing them and in order to have a huge number of data to help in testing.

A. FastDTW Algorithm

Fast Dynamic Temporal Warping (FastDTW) is a time series alignment algorithm that was initially designed for speech recognition. Its goal is to align two sequences of feature vectors by warping the time axis iteratively until an optimal match is discovered. The two sequences can be placed on the sides of a grid, one on top of the other, with the Y time-series axis on the top and the X time-series axis on the left. In each cell, a distance metric can be set, comparing the relevant elements of the two sequences. Path is detected through the grid by determining the best match alignment between the two sequences which are ”I”, and ”j”, and results in the minimum overall distance between each cell. Finding all different routes possibilities through the grid and computing the total distance regarding each one is the process for computing this total distance (D) as mentioned in Equation 1. The total distance is calculated by dividing the sum of the distances between individual items on the path by the weighting function’s sum. This is achieved by making an equal number of points in each of the two series, then calculating the Euclidean distance between the first and each subsequent point in the first series.

$$D(i,j) = |t(i) - r(j)| + min \begin{cases} D(i-1,j) \\ D(i-1,j-1) \\ D(i,j-1) \end{cases} \tag{1}$$

B. 1 Dollar Algorithm

The one-dollar is a geometric template matcher, the previously stored templates (T) are compared to the candidate strokes (C) resulting in the match that is the closest in 2-D Euclidean space as mentioned in Equation 2. Thus, we have exactly N points that will allow us to calculate the distance between C[k] to T[k] in which k= 1 to N. The most used pairwise point comparisons in the one-dollar algorithm are scale, rotation, and position invariant. One-dollar cannot differentiate between gestures whose identities rely on special orientations, ratios, and locations. One-dollar algorithm does not contain usage of time, as a result, gestures cannot be separated with respect to the speed.

$$d_i = \frac{\sum_{K=1}^{N} \sqrt{\left(C\left[K_x\right] - T_i\left[K_x\right]\right)^2 + \left(C\left[K_y\right] - T_i\left[K_y\right]\right)^2}}{N} \tag{2}$$

IV. Experiment

A. Experiment 1

The objective from conducted this experiment was to find out the most accurate algorithm in determining the activity performed by the worker inside the maintenance center. Besides, how it is estimated to differentiate between positive and negative activities.

This experiment was conducted on 3 participants by recording 3 video streams with a duration of 7 minutes for each video. The system has 84 videos of four different activities, including the positive and the negative activities in the different sequences. Each activity was repeated 7 times in each video 4 from the right side and 3 from the left side. The input to our system is split videos for each sequence of activities, the duration is 25 seconds for each video. consequently, After that, the videos were inserted into the system and extracted the important points in the skeleton to determine the path of each point (the X-coordinate and the Y-coordinate) using mediapipe, and save them in a file to be ready to be tested by the algorithms. The system calculated the accuracy of each algorithm in identifying and distinguishing between activities.

In this experiment, we were able to choose the best algorithm by testing each one of them separately, the one-dollar recognizer was able to give the best accuracy rate of 97% regarding the following activities: (changing tires, changing oil, using mobile, and standing without work). The rate of the fastDtw algorithm was 86% in the same previously mentioned activities as shown in Table II.

TABLE II. The Results of Experiment 1

Algorithms	Activites	Accuracy
One-dollar	changing tire changing oil use mobile stand without work	97%
Fastdtw	changing tire changing oil use mobile stand without work	86%

B. Experiment 2

This experiment was conducted to measure the performance of one-dollar algorithm with an increase in the number of participants.

We asked 15 participants to do a sequence of activities in different stream videos. The scenario of the activities was changing tire, use mobile, changing oil, stand without work and it was changeable from one to another. The average duration of the activity in the video was 30 seconds, Each activity was repeated 7 times 4 from the right side and 3 from the left side. The average age of the participants ranged from 19 to 23 years, and the characteristics of the body were also different. Afterwards, the videos were entered into the system, and the system was able to extract 420 videos for a range of different activities in uncontrolled environments. Through this experiment, we were able to measure the accuracy of the system in identifying activities in different conditions.

In this experiment, with large number of participants, experts in the field of mechanics, as well as workers from maintenance centers, the system was able to extract 420 videos, and the one-dollar algorithm started the stage of identifying activities. It achieved an accuracy rate of 94.2%.

C. Discussion

After the two experiments have been done, there's a discussion to explain why these results appear in both. The first experiment was to determine the most accurate algorithm, and the result was that one dollar was more accurate than the fastdtw, which is because the One dollar recognizer has been built over fast dynamic time warping (fastDTW), and both can determine path differences between two trajectories. But the difference between them is that one dollar recognizer reduces the noise in orientation by taking the angles into its calculations, which gives preference to one dollar in the accuracy ratios.The second experiment was to test the performance of the one-dollar algorithm in determining the activity of a large number of participants. The results showed that there was a significant difference in accuracy ratios between the activities. The activity of changing the tire was more subtle because this activity differed from the other activities in the motion trajectory, and the activity of using mobile and standing without work had high accuracy, but there were few similarities between them. The oil change activity has a similar motion trajectory with the two activities (use mobile, standing without work), which reduced its accuracy as shown in Fig. 6.

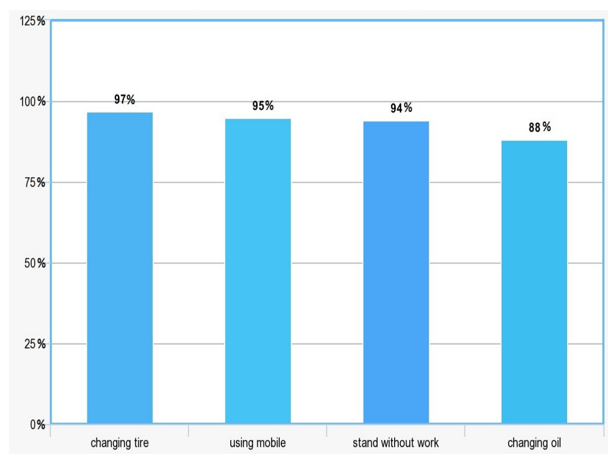


Fig. 6. The Results of the Activities.

The aim of this paper was achieved by using the method of pose classification to extract the pose points from the input video, after extracting the points, a one-dollar algorithm is used to match the extracted points with the collected dataset. As a result, we detect the activities and classify them whether they are positive activities or negative activities. We tested the previous method on 18 participants to prove the accuracy of the mentioned method which reached 94.2%.

V. Conclusion and Future Work

In this paper, we proposed a human activity recognition system in a car workshop that would help maintenance center owners to identify positive or negative worker activities correctly and efficiently. Our system used pose classification to extract workers poses and used two different algorithms to detect the activities. The system achieved an accuracy rate of 94.2% using this method. However, we are confident that the accuracy of this system will improve with more test videos and that we will increase the activities on which the experiment was conducted in the future. Our future work focus on calculating the wasted time of each worker. Besides, solving problems such as obstacles (engine hood) that appear while identifying activities and affecting the accuracy of the system.

References

[1] A. Bagate and M. Shah, "Human activity recognition using rgb-d sensors," in *2019 international conference on intelligent computing and control systems (ICCS)*. IEEE, 2019, pp. 902–905.

[2] T. Hossain, M. S. Islam, M. A. R. Ahad, and S. Inoue, "Human activity recognition using earable device," in *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, 2019, pp. 81–84.

[3] I. A. Lawal and S. Bano, "Deep human activity recognition using wearable sensors," in *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, 2019, pp. 45–48.

[4] J. Zhao, R. Han, Y. Gan, L. Wan, W. Feng, and S. Wang, "Human identification and interaction detection in cross-view multi-person videos with wearable cameras," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 2608–2616.

[5] S.-M. Lee, S. M. Yoon, and H. Cho, "Human activity recognition from accelerometer data using convolutional neural network," in *2017 ieee international conference on big data and smart computing (bigcomp)*. IEEE, 2017, pp. 131–134.

[6] C. Liu, J. Ying, F. Han, and M. Ruan, "Abnormal human activity recognition using bayes classifier and convolutional neural network," in *2018 IEEE 3rd international conference on signal and image processing (ICSIP)*. IEEE, 2018, pp. 33–37.

[7] J. Patalas-Maliszewska, D. Halikowski, and R. Damaševičius, "An automated recognition of work activity in industrial manufacturing using convolutional neural networks," *Electronics*, vol. 10, no. 23, p. 2946, 2021.

[8] S. Win and T. L. L. Thein, "Real-time human motion detection, tracking and activity recognition with skeletal model," in *2020 IEEE Conference on Computer Applications (ICCA)*. IEEE, 2020, pp. 1–5.

[9] G. Ogbuabor and R. La, "Human activity recognition for healthcare using smartphones," in *Proceedings of the 2018 10th international conference on machine learning and computing*, 2018, pp. 41–46.

[10] S. Sezer and E. Surer, "Information augmentation for human activity recognition and fall detection using empirical mode decomposition on smartphone data," in *Proceedings of the 6th International Conference on Movement and Computing*, 2019, pp. 1–8.

[11] P. Washington, A. Kline, O. C. Mutlu, E. Leblanc, C. Hou, N. Stockham, K. Paskov, B. Chrisman, and D. Wall, "Activity recognition with moving cameras and few training examples: Applications for detection of autism-related headbanging," in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–7.

[12] Z. Hussain, Q. Sheng, and W. E. Zhang, "Different approaches for human activity recognition: A survey," 06 2019.

[13] M. Yalçın, N. Tüfek, and H. Yalcin, "Activity recognition of interacting people," in *2018 26th Signal Processing and Communications Applications Conference (SIU)*. IEEE, 2018, pp. 1–4.

[14] N. Nair, C. Thomas, and D. B. Jayagopi, "Human activity recognition using temporal convolutional network," in *Proceedings of the 5th international Workshop on Sensor-based Activity Recognition and Interaction*, 2018, pp. 1–8.

[15] S. Tsokov, M. Lazarova, and A. Aleksieva-Petrova, "Evolving 1d convolutional neural networks for human activity recognition," in *International Conference on Computer Systems and Technologies' 21*, 2021, pp. 49–54.

[16] A. Naser, A. Lotfi, J. Zhong, and J. He, "Human activity of daily living recognition in presence of an animal pet using thermal sensor array," in *Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, 2020, pp. 1–6.

[17] P. Hristov, A. Manolova, and O. Boumbarov, "Deep learning and svm-based method for human activity recognition with skeleton data," in *2020 28th National Conference with International Participation (TELECOM)*. IEEE, 2020, pp. 49–52.

[18] H. Ramirez, S. A. Velastin, I. Meza, E. Fabregas, D. Makris, and G. Farias, "Fall detection and activity recognition using human skeleton features," *IEEE Access*, vol. 9, pp. 33 532–33 542, 2021.

[19] A. S. Syed, Z. S. Syed, M. S. Shah, and S. Saddar, "Using wearable sensors for human activity recognition in logistics: A comparison of different feature sets and machine learning algorithms," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 9, 2020. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2020.0110976

[20] I. Serrano, O. Deniz, J. L. Espinosa-Aranda, and G. Bueno, "Fight recognition in video using hough forests and 2d convolutional neural network," *IEEE Transactions on Image Processing*, vol. 27, no. 10, pp. 4787–4797, 2018.

[21] Y. E. Görgülü and K. Taşdelen, "Human activity recognition and temporal action localization based on depth sensor skeletal data," in *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*. IEEE, 2020, pp. 1–5.

[22] W. Tao, Z.-H. Lai, M. C. Leu, and Z. Yin, "Worker activity recognition in smart manufacturing using imu and semg signals with convolutional neural networks," *Procedia Manufacturing*, vol. 26, pp. 1159–1166, 2018.

[23] S. Alghyaline, "A real-time street actions detection," *International Journal of Advanced Computer Science and Applications*, vol. 10, 01 2019.

[24] M. Arzani, M. Fathy, A. Azirani, and E. Adeli, "Switching structured prediction for simple and complex human activity recognition," *IEEE Transactions on Cybernetics*, vol. PP, pp. 1–12, 01 2020.

[25] N. Archana and K. Hareesh, "Real-time human activity recognition using resnet and 3d convolutional neural networks," in *2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS)*. IEEE, 2021, pp. 173–177.

[26] X. Sun, H. Xu, Z. Dong, L. Shi, Q. Liu, J. Li, T. Li, S. Fan, and Y. Wang, "Capsganet: Deep neural network based on capsule and gru for human activity recognition," *IEEE Systems Journal*, 2022.

[27] V. Radhika, C. R. Prasad, and A. Chakradhar, "Smartphone-based human activities recognition system using random forest algorithm," in *2022 International Conference for Advancement in Technology (ICONAT)*. IEEE, 2022, pp. 1–4.

[28] P. Mittal *et al.*, "Machine learning (ml) based human activity recognition model using smart sensors in iot environment," in *2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*. IEEE, 2022, pp. 330–334.

[29] Y. Tang, L. Zhang, F. Min, and J. He, "Multi-scale deep feature learning for human activity recognition using wearable sensors," *IEEE Transactions on Industrial Electronics*, 2022.

[30] M. Noras, "Activity detection and recognition with passive electric field sensors," *IEEE Transactions on Industry Applications*, 2021.

BMP: Toward a Broker-less and Microservice Platform for Internet of Thing

Lam Nguyen Tran Thanh[1], Khoi Le Quoc[2], The Anh Nguyen[3], Huong Hoang Luong[4], Hong Khanh Vo[5], Tuan Dao Anh[6], Hy Nguyen Vuong Khang[7], Khoi Nguyen Huynh Tuan[8], Hieu Le Van[9], Nghia Huynh Huu[10], Khoa Tran Dang[11], Khiem Huynh Gia[12]
VNPT Information Technology Company, Ho Chi Minh city, Vietnam[1]
FPT University, Can Tho City, Viet Nam[2,3,4,5,6,7,8,9,10,11,12,13]

***Abstract*— The Internet of Things (IoT), currently, is one of the most interesting technology trends. IoT is the foundation and driving force for the development of other scientific fields based on its ability to connect things and the huge amount of data it collects. The IoT Platform is considered the backbone of every IoT architecture that not only allows the transfer of data between user and device but also the feed of high-level applications such as big data or deep learning. As a result, the optimal design of the IoT Platform is a very important issue, which should be carefully considered in many aspects. Although the IoT is applied in multiple domains, there are three indispensable features including (a) data collections, (b) devices and users management, and (c) remote device control. These functions usually come with some requirements, for example, security, high-speed transmission, low energy consumption, reliable data exchange, and scalable systems. In this paper, we propose the IoT Platform,called BMP (Broker-less and Microservice Platform) designed according to microservice and broker-less architecture combined with gRPC protocol to meet the requirements of the three features mentioned above. Our IoT Platform addresses five issues: (1) address the limited processing capacity of devices, (2) reduce energy consumption, (3) speed up transmission rate and enhance the accuracy of the data exchange, (4) improve security mechanisms, and (5) improve the scalability of the system. Moreover, we describe the evaluation to prove the effectiveness of the BMP (i.e., proof-of-concept) in three scenarios. Finally, a source code of the BMP is publicized on the GitHub repository to engage further reproducibility and improvement.**

Keywords—Internet of Things (IoT); gRPC; Single Sign-On; Broker-Less; Kafka; Microservice; Role-based Access Control (RBAC)

I. Introduction

The application fields of the Internet of Thing (IoT), currently, are increasingly diverse including smart cities, healthcare, supply chains, industry, agriculture. According to [1], there will be approximately 75.44 billion IoT connected devices in 2025. Iot Platform is an intermediary system that acts as an "adhesive layer" to connect devices with users. There are many architectures outlined to optimize the IoT system, including 5-layer architecture in order from low to high: Things, Connect, Collect, Learn and Do introduced in book [2]. This architecture facilitates it easy to separate each specific group of roles required of an IoT Platform. The fields of application of the current IoT are varied, it is challenging to use a specific architecture suitable for every practical application. However, we generally draw features that are indispensable for an IoT system, namely, i) data collection; ii) device and user management; and iii) remote device control. These three features are the Things, Connect and Collect layers, respectively [2].

The **Things layer** are the group of physical devices that directly collect data or perform an action based on a control command from the user. The device is usually limited in power, processing and bandwidth [3]. So, the most important requirement of the **Things layer** is balanced between processing capabilities and power consumption of devices (1).

The **Connect layer** is responsible for transmitting data. So, this layer is determined by transmission protocols which are suitable for the hardware and network processing capabilities of the devices in the **Things layer**. So, the most important requirement of the **Connect layer** is to ensure fast transmission speed and reliable transmission (2).

The **Collect layer** is responsible for gathering data from devices and users. The information can be very sensitive since it relates to user privacy, especially medical or financial IoT applications. So, the most important requirement of the **Collect layer** is a security mechanism (3).

In addition, the system scalability and resilience should be concerned as important factors in IoT platform design (4).

There are five popular IoT protocols namely Hypertext Transfer Protocol (HTTP), Constrained Application Protocol (CoAP), Extensible Messaging and Presence Protocol (XMPP), Advanced Message Queuing Protocol (AMQP), and Message Queuing Telemetry Protocol (MQTT) [4].

Regarding to limited processing ability and power consumption (1): There are many studies comparing the advantages and disadvantages of the above-mentioned protocols [1], [5], but for the communication in constrained networks bandwidth, MQTT and CoAP are proposed to be used [6]. Furthermore, regarding energy consumption, MQTT is lower than CoAP [7], so MQTT is the most favorite protocol used by developers [8].

Regarding to transmission rate (2): MQTT is twice as fast as CoAP [9]. MQTT has three levels of Quality-of-Service (QoS) from 0 to 2. Selecting these QoS levels is a trade-off between the reliability of packet transmission (rate of loss on the link), transmission rate, and energy consumption. The transmission rate of QoS-0 is the fastest, but this has the lowest reliability [10]; whereas, the opposite is the QoS-2. On the one hand, the energy consumption of the QoS-0 level is only about 50% of the QoS-2 level [11].

Regarding to security mechanism (3): MQTT has many limitations [10]. Scenarios that attack the Integrity, Availability, and Authentication and Authorization mechanisms of MQTT are outlined in [12].

Regarding to system scalability and resilience (4): MQTT protocol is used in popular IoT frameworks including well-known companies (e.g. IBM, Amazon and Microsoft) [13]. The MQTT protocol uses a pub/sub architecture [14], with the MQTT broker at the center. The MQTT subscriber (client), connects to the broker and sends messages to topics. Brokers rely on topic to route the packet, meaning that subscribers who subscribe a topic will receive all messages sent to that topic. This easily leads to a single point failure [15] error at the central broker location. In addition, MQTT is created for transmission purposes. Therefore, MQTT broker does not provide message storage as well as guarantee the order of messages when it reaches the receiver [16]. This is also an important issue to be considered as the weakness of the system using MQTT.

Addressing the MQTT issues, the aim of this paper is to propose a new IoT platform, called BMP, built on broker-less and microservice architecture. We used gRPC protocol to build the broker-less architecture at the **Collect layer** and **Connect layer** for gateway devices. This architecture provides a peer-to-peer communication between sender and receiver in BMP instead of using a central broker to coordinate topic-based messages. Collected messages will be sent to the message queue (Kafka) which provides a caching message mechanism to reduce lost messages when system error occurs. Furthermore, the microservice architecture ensures fault tolerance, scaling horizontally, availability, and carrying capacity of the BMP [17]. Moreover, we define a management model to manage the components in the BMP including users, devices, and communication channels. The management model is built on the RBAC (role-based access control) model combined with the single sign-on and the user organization to achieve authentication and authorization. To engage further reproducibility or improvement in this topic, finally, we share the completely code solution which is publicized on the our Github.

The rest of the paper is organized as follows. In Section 2 we provide knowledge about the technology used in the paper. Section 3 discusses related work. We introduce our IoT Platform and describe the proof-of-concept in Sections 4 and 5, respectively. Section 6, we discuss our test results. Finally, we summarize the paper and provide the potential directions for future work.

II. Background

A. gRPC

The gRPC[2] is an open-source framework developed by Google for implementing RPC (Remote Procedure Calls) API via HTTP/2. The gRPC provides a new reality for RPC by making it interoperable, modern, and efficient using technologies such as Protocol buffers and HTTP/2. HTTP/1.1 protocol was born in 1997. With each request sent from the client to the server, a TCP connection is created. Connection processing must go through a three-way handshake. This takes a lot of time. In addition, the headers in each request are plain text with lots of data fields and are not compressed and the headers usually have the same data. Therefore, the header occupies a significant size and is duplicated many times in the requests, leading to consuming a lot of bandwidth. HTTP/2 protocol was born in 2015 created by Google to overcome the above problems. HTTP/2 protocol supports multiplexing so it is possible to send multiple requests in parallel on an established TCP connection. This reduces bandwidth suitable for devices with limited hardware. Protocol buffers are a popular technology for data structures developed and used in communication between google services. The developer will define the data structure in the *.proto* file. The Protoc compiler then compiles the .proto file into any language it supports. At runtime, the data is compressed and normalized to binary. The gRPC is fully compatible with embedded devices [18]. This protocol provides four communication types including unary, server-streaming, client-streaming, and bidirectional streaming [19]. In Unary, the client sends a request, then the server sends back a response. In server-streaming, the lient sends a request to the server, then the server sends back multiple responses on the same TCP connection. The order of messages for each stream is guaranteed to be the same between client and server. In client-streaming, the client sends multiple requests to the server, then the server sends back only one response to the client on the same TCP connection.In bidirectional-streaming, the client sends multiple requests to the server, then the server sends back multiple responses on the same TCP connection without waiting for response time.

B. Kafka Message Queue

Kafka[3] is a distributed messaging system. It can transfer a vast number of messages in real-time. When the receiver has not received the message, this message is still stored on the message queue and the disk. This feature allows reducing lost messages when a system error occurs. The structure of Kafka includes the following main components [20]: producer, topic, partitions, consumer, broker, and zookeeper. A producer can be any application that publishes messages to a topic.A topic is a category or feed name where the record is published. The topics are divided into different segments, which are called partitions.A consumer can be any application that subscribes to a topic and consumes messages. Kafka cluster is a set of servers, each of which is called a broker. Zookeeper used to manage and arrange brokers. Kafka-Pixy[4] is a dual API (gRPC and REST) proxy for Kafka with automatic consumer group control.In this paper, we use Kafka-pixy open source to communicate between Kafka message queue and other microservice in BMP by gRPC protocol. Detailed source code can be found at the link[5][6].

C. Oauth and Single Sign-On

Oauth version 2.0[7] stands for Open with Authentication or Authorization. OAuth was born to solve the above problem and beyond, this is an authentication method that helps applications

[2] https://grpc.io/

[3] https://kafka.apache.org/

[4] https://github.com/mailgun/kafka-pixy

[5] https://github.com/thanhlam2110/Kafka-gRPC-Producer

[6] https://github.com/thanhlam2110/Kafka-gRPC-Consumer

[7] https://oauth.net/2/

to share resources without sharing username and password information.

Single Sign-On (SSO) is a mechanism that allows users to access many websites and applications without simply logging in once. Once identified in an application A, it will also be identified in application B without repeating the login operation. This feature is suitable for IoT applications because a user can be a customer of many different IoT service providers [21]. In this paper, we use open source CAS Apereo which is modified by us to implement SSO service and Oauth protocol. Detailed source code can be found at the link[8]

D. Microservice Architecture

Microservices are an approach to developing an application using a set of small services, a service that will run its own process independently, usually an HTTP resource API. This modern architecture allows for the creation of larger, more complex, and scalable applications [22]. Applications using the microservixe architecture are very easy to maintain because the modules are completely separate from each other. It also provides high reliability since a service failure does not affect other services at all [23]. The microservice architecture is also easily extensible. The only weakness of the microservice architecture is the complex deployment process [24].

III. Related Work

A. Broker-less Architecture in IoT

Lu et al. in [25] described a solution to collect the data while meeting the confidentiality requirements via fog computing-enhanced IoT devices. They used a one-way hash string to authenticate IoT devices, then applied the Chinese Remainder Theorem to aggregate data generated by different IoT devices. They also took advantage of Uniform Paired Encryption to provide data security. Some interests of the proposed solution include i) fault tolerance, ii) solving heterogeneity problems in partial mode. Moreover, Lam et al. [26], [27] applied the broker-less architecture in their IoT platforms. However, their resolution applied a common privacy rule, differential privacy, to all IoT devices causing unintentional amounts of sensitive data to be accepted. Applying the access control model can solve this drawback [28], [29], where only authorized objects can accessed data [30], [31].

Furthermore, in the industrial environments, some research directions have applied IoT based on decentralized architecture to meet the requirements for specific system. For example, in [32], the authors focused on the varying IoT protocols in a distributed control platform by concentrating on the times and limits of the transit for different parameter options. Standardized system configuration exploits the embedded Raspberry Pi development board with open source protocol implementation for efficacy measurement, including scalability effectiveness with various data consumers in an automated network chemical. Leveraging the on-board computing resources of distributed embedded systems in the edge computing paradigm for industrial automation was argued in [33].

B. IoT Platform based on Microservice

The microservices architecture is introduced to address the traditional monolithic issues. For example, microservices architecture was introduced to fill this gap [34]. This article focuses on summarizing prior work that used the microservices benefits in designing their architecture rather than detailing how it works and the direction of development.

For the framework, Amazon offers AWS IoT Greengrass[1] as a solution to migrate analytics capabilities directly to edge devices (e.g., smartphones). Microsoft Azure provides similar functionality to Azure Stream Analytics on IoT Edge[2], enabling users to use near-real-time analytic functionalities by using Azure Stream Analytics on IoT-applied devices. However, these frameworks neither provide any functionality to move Lambda calculation to and fro between cloud edge devices nor mix with exterior stream processing engines.

For the application, Lam et al. [35], [36], [37] applied microservice architecture to define the healthcare and IoT-Platform applications. In addition, Maia and associates [38] presented IRRISENS, which is designed based on microservices architectures used in agricultural environments to sense soil, crop, and atmospheric parameters.

IV. BMP Architecture

The BMP is designed according to microservice architecture, and broker-less architecture including **Thing layer**, **User layer**, **Edge layer** and **Cloud layer**, as shown in Fig. 1:

The **Thing layer** is the group of sensor or physical devices.These devices are responsible for measuring environmental parameters such as humidity, temperature or patient health parameters, etc. depending on the IoT application.

In each of these devices, we implement two services i) collection data service (client) and ii) control service (client). The former is responsible for streaming data collected from the environment according to a predetermined collection data server (server) in the IoT Platform. This data streaming is only performed when the client is authenticated and checked the things' role by the Single Sign-On service. The latter receives control commands from the control service (server) through the message queue system. The **Edge layer** including gateway devices. The Edge layer includes gateway devices. In each gateway, we implement two services: *collect data service client-side (CDC)* and *control service client-side (CSC)*.The CDC receives data from devices in the **Thing layer**, then sends it to the **Cloud layer**. The CSC receives control commands from **User layer** throughout **Cloud layer**.

The **User layer** is the groups of users, with different roles, registered to use IoT services. Users can control and monitor the device state by using the control service and the collect data service, in turn. This is performed after passing at the authentication phase and verifying at the authorization phase by the Single Sign-On service and RBAC model, respectively. Besides, users can manage device information (for instance, create, delete, disable, active, or manage) their child users

[8] https://github.com/thanhlam2110/cas-overlay-template

[1] https://docs.osgi.org/specification/osgi.cmpn/7.0.0/service.event.html
[2] https://maven.apache.org/

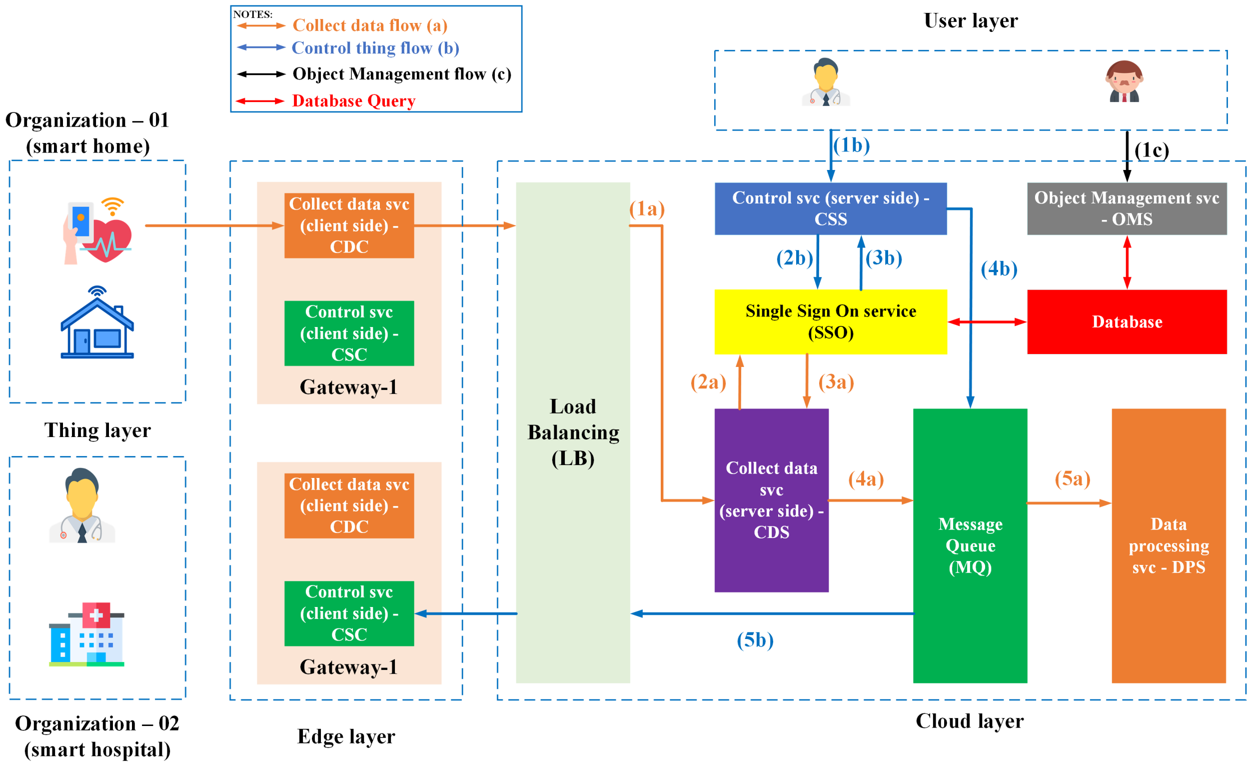


Fig. 1. BMP Platform Architecture.

(register, disable, active) through the Object Management service. Furthermore, users can create management information about devices and communication channels. This information is metadata and stored in a database helps users easily manage their devices and be aware clearly of where they send your messages. For example, user A can create/delete/update his device and communication channel information. Moreover, users are organized hierarchically according to a tree model so high-level users (aka *parent-user*) can manage low-level users (aka *child-user*). For example, user A (high-level) can enable or disable user B's status (low-level). If user B is disabled, user B can not interact with BMP. This feature is similar to the company model with user A as manager role and user B as staff role. In addition to the user and things layer services, the IoT platform has a Load balancing service to balance the load for the whole system. Besides, the Message Queue service is responsible for routing control packets from user to things and data collected from the things layer. Message Queue also stores messages going through the IoT Platform, ensuring that it could receive messages after recovery even the service fails. Data processing service analyzes data in-depth level.

The **Cloud layer** is the main processing part of BMP including *load balancing (LB)*, *collect data service server-side (CDS)*, *control service server-side (CSS)*, *single sign-on service (SSO)*, *message queue (MQ)*, *object management service (OMS)*, *data processing service (DPS)*, and *database (DB)*. The *LB* is responsible for distributing incoming traffic to ensure the **Cloud layer** is not overloaded. The *CDS* receives messages from CDC. These messages include Oauth access token and data collected from the **Thing layer**. The access token is verified by the *SSO*. If the access token is valid the data will be sent to the *MQ*, otherwise, the data will be discarded. The *CSS* receives messages from users in the **User layer**. These messages include an Oauth access token and control command. Similarly, the access token is verified by the *SSO*. If the access token is valid the control command will be sent to the *MQ*, otherwise, the the control command will be discarded. Thanks to gRPC's multiplexing fearture and client-stream method, we only need to establish an http/2 connection between CSC and CSS or CDC and CDS to send the access token and a bunch of data at the same time. This helps to reduce system bandwidth as well as power consumption on the gateway. The *MQ* is responsible for distributing data to the *DPS* for deeper analysis or the *DB* for storage and control commands to the CSC in the **Edge layer**.

V. Implementation

In this paper, we have implemented the services outlined in the proposal section including: single-sign on service (SSO)[9], collect data service (CDC and CDS)[10], and object management service (OMS)[11] . The model for the development and interaction between the symbolic services in Fig. 1 is as follows: (1a-5a) the collect data flow, (1b-5b) the control thing flow, and (1c) the object management thread.

The **collect data flow** details are shown in Fig. 2. The **Things layer** collects data from sensors and periodically sends them back to the gateway at the **Edge layer** (sending data periodically to help save energy). The CDC on gateway creates

[9] https://github.com/thanhlam2110/iccs-sso-service
[10] https://github.com/thanhlam2110/iccs-collect-data-svc
[11] https://github.com/thanhlam2110/iccs-object-management-svc

a single gRPC connection to both send access token, thing-id and streaming data to CDS - *process 1a* in Fig. 1. This is the multiplexing feature of the gRPC protocol, which was introduced in section 2.1. So the data sent from CDC to CDS is the combination of *access token*, *thing-id*, and *collected data from sensor*. Next, the CDS sends the *access token* and *thing-id* to the SSO service (process 2a) to verify access token and check role, the test result is returned (process 3a). If the token is authenticated and the role is verified, CDS will stream the bulk of *collected data* to MQ (process 4a), whereas the CDS will drop all data and return the error to CDC. MQ will distribute data to DPS for in-depth analysis (process 5a).

The **control thing flow** details are shown in Fig. 3. The user creates a single gRPC connection to both send the access token, things-id and control command to the CSS (process 1b). The CSS sends the *access token* and *thing-id* to the SSO service to verify access token and check role. Role-based access control integrated on SSO service uses user-id (get from check access token process) and thing-id to verify the device which the user wants to control belongs to the user or not (process 2b). Next, the check result is returned to the CSS (process 3b), if it is valid, the control command will be sent to the MQ (process 4b). The CSC on gateway subscribes to the channel (aka Kafka topic) on the MQ which is created by the user to be dedicated to receive control commands. When the MQ receives the control command, the CSC also receives it. (process 5b).

The **object management flow** provides APIs for users to interact with the database. The Single Sign On service will also use the data stored here to authenticate and verify the roles of users and things. We implement RBAC combined with Single Sign-On to provide role-based authentication and authorization for Things and Users. However, the concepts of users' roles and things' roles are somewhat different.

We define the users' role including permissions that affect their devices and child users.

About users' role on their child user, a parent user can create/delete/enable/ disable child user. For example, user A can create management information for user B who is the visitor to user A's house. In this case, user A is the *parent-user* and user B is the *child-user* of A. With management information, user B is allowed to control some devices in user A smart home. When user B leaves, user A deletes user B's information, which means user B can not interact with user A's smart home. We implement the user management by the model tree with the *child-user* has an attribute user_parent_id field equal to the *parent-user*'s username. This organization user model allows BMP to be flexibly applied to the business of implementing hierarchical user management.

About users' role on their devices, the user can create/delete management information, enable/disable, or assign/unassign (to another user) their devices. For example, user A can choose some devices that user B can not control when user B visits user A's house by unassigning these devices to user B. To do this, we provide APIs that allow user-id and thing-id mapping. Then, SSO will check this information when a user wants to control devices. In addition, OMS also provides APIs for users to create/delete communication channels management information. This channel information allows users to assign their devices to send and receive messages on it. This function helps users be aware clearly where they share data.

VI. Evaluation

The BMP design is based on microservice and broker-less architecture. To evaluate BMP, we deploy the **Cloud layer** on the Amazon EC2[10] platform with each service equivalent to a virtual machine with 1GB RAM and 1 vCPU configuration. For the gateway, we deploy CDC and CSC on the Raspberry Pi 3 model B[11] module with CPU quad-core, 1.2 GHz and 1GB RAM.

In this paper, we perform three test scenarios to evaluate BMP with aspects transmission rate, power consumption, and security mechanism.

Scenario 1: We measure the Round Trip Time (RTT) from when CDC streams *data* until it is received by the message queue. In addition, we also look at the error rate (number of messages lost per total number of messages). We assume the *data* in this test case is *data* = *access token* + *string*, with *string* = "hello number" + for (i=1; i¡=number_of_messages;i++). Measured results are shown in Table I.

The test results are quite good as we only deploy the evaluation on low-configuration servers. All messages are received in full and in order. By using gRPC which allows sending multiple messages on one connection, we achieve very fast transfer speed while still satisfy the whole process of checking and validating before streaming data as described in Section 5.

Scenario 2: We measure CPU and RAM usage to evaluate power consumption when streaming *data*. We use the htop[12] tool, a tool that allows real-time monitoring of system processes. Similarly test scenario 1, We assume the *data* in this test case is *data* = *access token* + *string*, with *string* = "hello number" + for (i=1; i¡=number_of_messages;i++).

As a result in Table II, at the no-load and 10000 messages levels, we recorded the value as 0, which means that the resource usage of the services is shallow (no-load) or takes place in a concise time (10000 messages); therefore, the measuring tool hardly records any change in resource usage. For the high load case (50000-100000 messages), the result shows that the resource consumption is very low. Furthermore, in practice, it is not always the thing that streams a large amount of data like in this scenario. This shows that our adoption of gRPC to the IoT Platform gives outstanding results, suitable for devices with low hardware.

Scenario 3: We capture the message that is transmitted between sender and receiver to analyze security risk. We use Wireshark[3] to capture the transmitted message. Wireshark is network packet analyzer software. Its job is to capture all network packets and then display the data of that packet in the most detail. We set up the test scenario to compare the security mechanism between MQTT and gRPC protocol. We use Wireshark to capture messages that are transmitted between the

[10] https://aws.amazon.com/
[11] https://www.raspberrypi.org/products/raspberry-pi-3-model-b/
[12] https://vitux.com/how-to-use-htop-to-monitor-system-processes-in-ubuntu/
[3] https://www.wireshark.org/

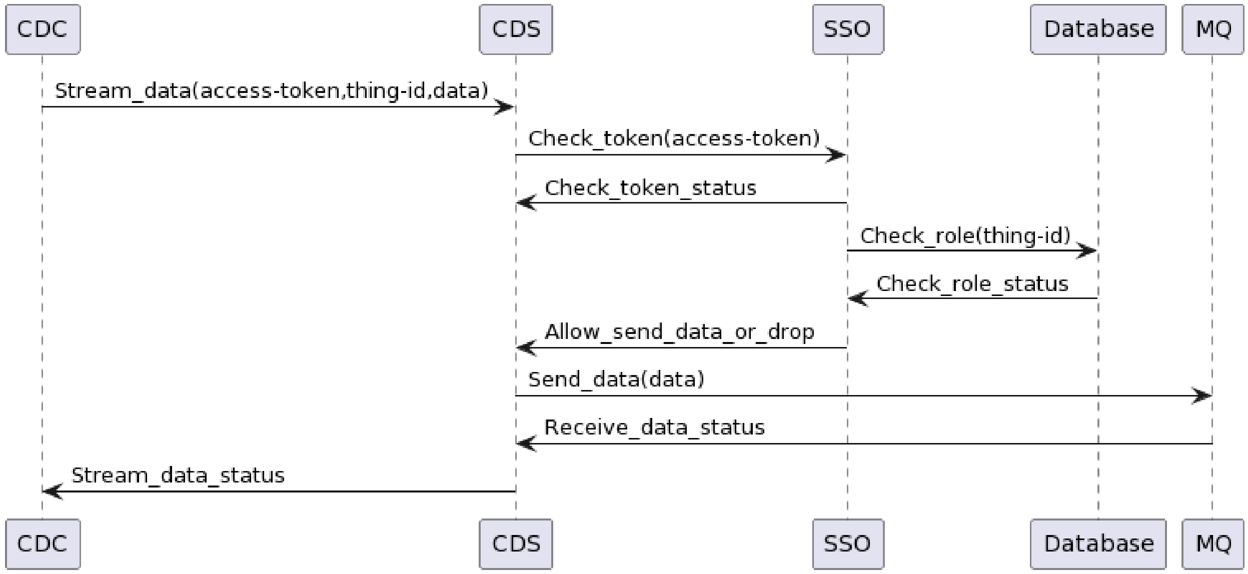


Fig. 2. The Collection Data Flow.

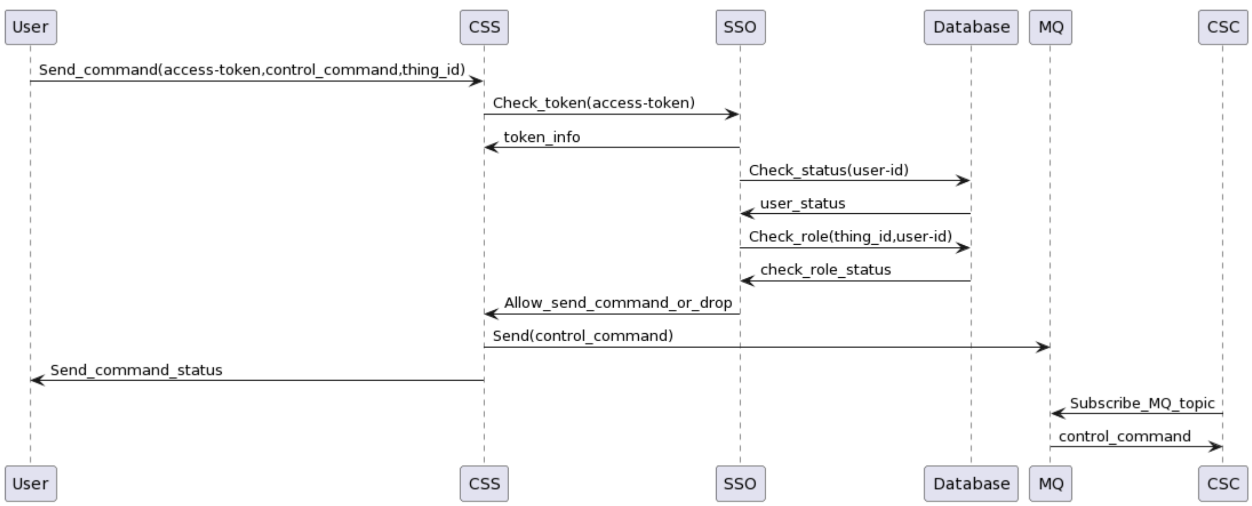


Fig. 3. The Control Flow.

TABLE I. The Result of the First Scenario

Number of messages	100	1000	5000	10000	50000
RTT (s)	1.5	4.07	14.79	27.05	138.02
Error (%)	0	0	0	0	0

sender and receiver and then analyze these messages. The test model is shown in Fig. 4.

The analyzed message result of MQTT and gRPC protocol is shown in Fig. 5 and Fig. 6, respectively.

According to Fig. 5 and Fig. 6, while MQTT protocol easily gets topic information as well as message content, in this case MQTT topic is *"thanh-lam"* and message is *"Hello Thanh Lam"*, gRPC protocol provides packet encryption. This result proves that the gRPC protocol has a better security mechanism than the MQTT protocol.

VII. Conclusion

In this paper, we propose the IoT Platform, called BMP, using the broker-less and microservice architecture. The broker-less architecture helps our IoT Platform reduce a single point of failure in comparison with brokering architecture. Furthermore, microservice architecture helps BMP easily scale out, reducing downtime when errors occur. The combination of broker-less and microservice architecture also helps 3rd parties easily integrate with BMP. Third-party developers only need to call the API and don't make major changes to the

TABLE II. THE RESULT OF THE SECOND SCENARIO

Number of messages		No load	10000	50000	100000
Raspberry Pi	CPU	0	0	0.3%	0.7%
	RAM	0	0	2.5%	2.7%
Collection data service (server)- CDS	CPU	0	0	8.2%	10.5%
	RAM	1.7%	1.7%	1.7%	1.7%

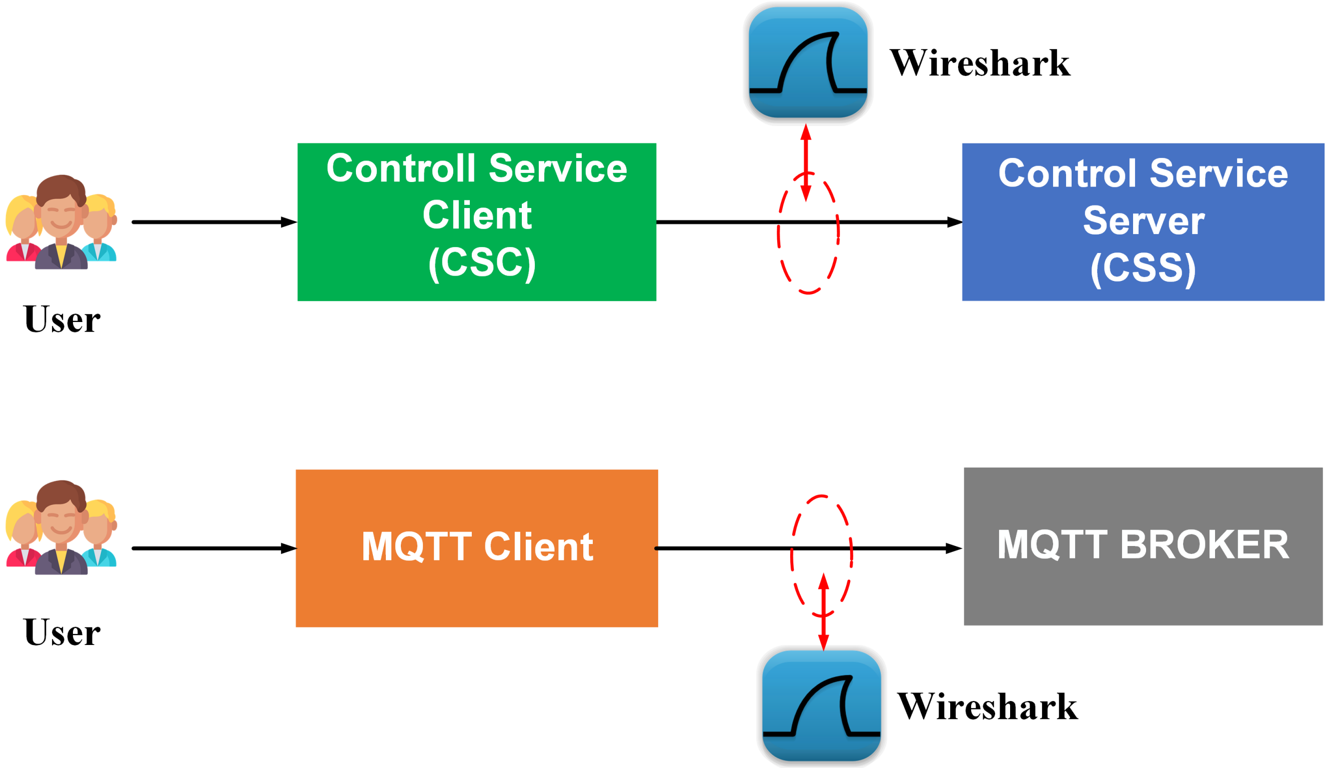


Fig. 4. Capture Transmission Message using Wireshark.

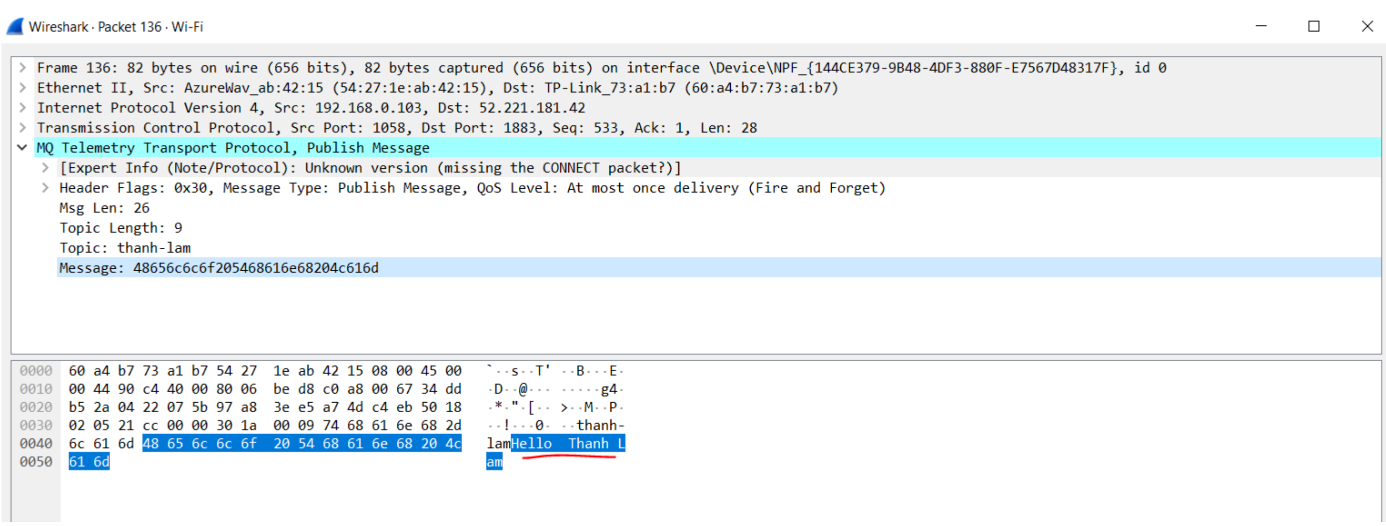


Fig. 5. MQTT Protocol Message without Encrypted Format.

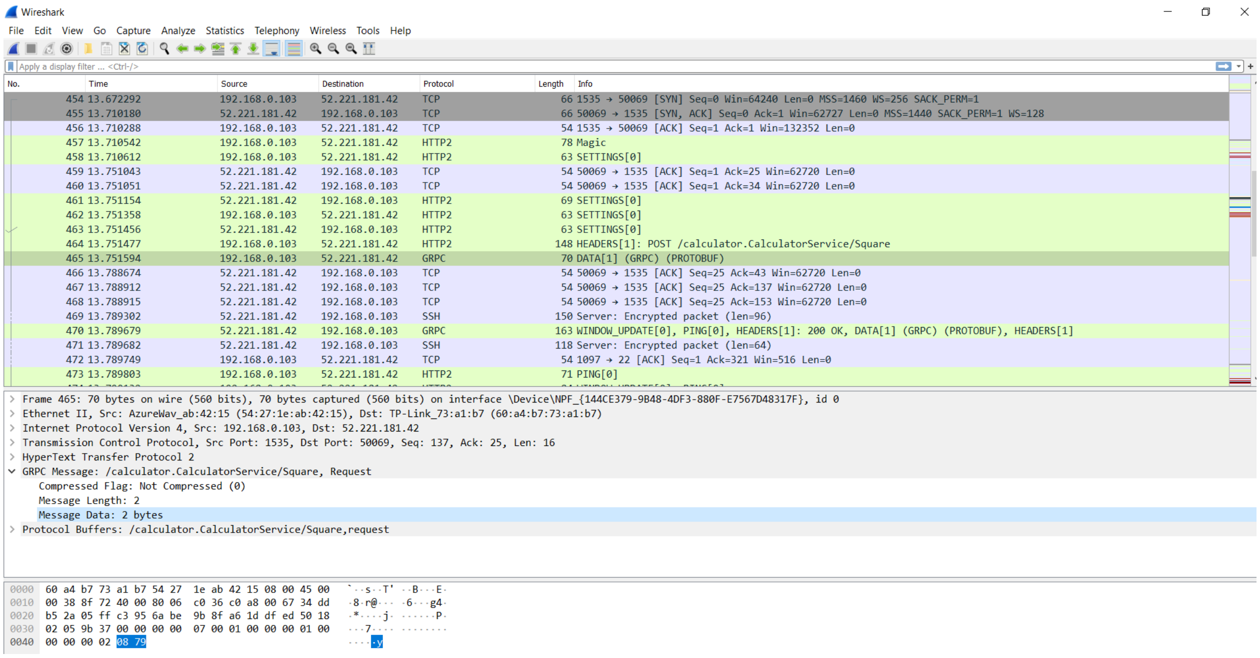


Fig. 6. gRPC Protocol Message with Encrypted Format.

BMP's architecture. Moreover, we chose gRPC as the main protocol for BMP because of its advantage in privacy, power consumption, speed, and reliable transmission in comparison with MQTT. In addition, the gRPC protocol works peer-to-peer which is suitable for broker-less implementation. BMP also provides authentication and authorization mechanisms for users, devices, and communication channels thanks to the combination of single sign-on (SSO) and role-based access control (RBAC). The BMP allows users to manage their devices and channels to aim at users clearly aware when they share data. The hierarchical organization of users as a tree model helps enhance flexibility. In the future, we will focus on the decentralized identity for IoT users and devices by applying blockchain.

REFERENCES

[1] M. Bansal *et al.*, "Application layer protocols for internet of healthcare things (ioht)," in *2020 Fourth International Conference on Inventive Systems and Control (ICISC)*. IEEE, 2020, pp. 369–376.

[2] T. Chou, *Precision-Principles, Practices and Solutions for the Internet of Things*. McGraw-Hill Education, 2017.

[3] V. Karagiannis, P. Chatzimisios, F. Vazquez-Gallego, and J. Alonso-Zarate, "A survey on application layer protocols for the internet of things," *Transaction on IoT and Cloud computing*, vol. 3, no. 1, pp. 11–17, 2015.

[4] A. Niruntasukrat, C. Issariyapat, P. Pongpaibool, K. Meesublak, P. Aiumsupucgul, and A. Panya, "Authorization mechanism for mqtt-based internet of things," in *2016 IEEE International Conference on Communications Workshops (ICC)*. IEEE, 2016, pp. 290–295.

[5] S. Verma and M. A. Rastogi, "Iot application layer protocols: A survey."

[6] S. P. Jaikar and K. R. Iyer, "A survey of messaging protocols for iot systems," *International Journal of Advanced in Management, Technology and Engineering Sciences*, vol. 8, no. II, pp. 510–514, 2018.

[7] M. Martí, C. Garcia-Rubio, and C. Campo, "Performance evaluation of coap and mqtt_sn in an iot environment," in *Multidisciplinary Digital Publishing Institute Proceedings*, vol. 31, no. 1, 2019, p. 49.

[8] G. C. Hillar, *MQTT Essentials-A lightweight IoT protocol*. Packt Publishing Ltd, 2017.

[9] B. H. Çorak, F. Y. Okay, M. Güzel, Ş. Murt, and S. Ozdemir, "Comparative analysis of iot communication protocols," in *2018 International symposium on networks, computers and communications (ISNCC)*. IEEE, 2018, pp. 1–6.

[10] S. Lee, H. Kim, D.-k. Hong, and H. Ju, "Correlation analysis of mqtt loss and delay according to qos level," in *The International Conference on Information Networking 2013 (ICOIN)*. IEEE, 2013, pp. 714–717.

[11] J. Toldinas, B. Lozinskis, E. Baranauskas, and A. Dobrovolskis, "Mqtt quality of service versus energy consumption," in *2019 23rd International Conference Electronics*. IEEE, 2019, pp. 1–4.

[12] J. J. Anthraper and J. Kotak, "Security, privacy and forensic concern of mqtt protocol," in *Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM), Amity University Rajasthan, Jaipur-India*, 2019.

[13] F. Carranza *et al.*, "Brokering policies and execution monitors for iot middleware," in *Proceedings of the 24th ACM Symposium on Access Control Models and Technologies*, 2019, pp. 49–60.

[14] D. Soni and A. Makwana, "A survey on mqtt: a protocol of internet of things (iot)," in *International Conference On Telecommunication, Power Analysis And Computing Techniques (ICTPACT-2017)*.

[15] P. Lv, L. Wang, H. Zhu, W. Deng, and L. Gu, "An iot-oriented privacy-preserving publish/subscribe model over blockchains," *IEEE Access*, vol. 7, pp. 41 309–41 314, 2019.

[16] H. C. Hwang *et al.*, "Design and implementation of a reliable message transmission system based on mqtt protocol in iot," *Wireless Personal Communications*, vol. 91, no. 4, pp. 1765–1777, 2016.

[17] M. Ali, S. Ali, and A. Jilani, "Architecture for microservice based system. a report," 2020.

[18] N. Karcher, R. Gebauer, R. Bauknecht, R. Illichmann, and O. Sander, "Versatile configuration and control framework for real time data acquisition systems," *arXiv preprint arXiv:2011.00112*, 2020.

[19] K. Indrasiri and D. Kuruppu, *gRPC: up and running: building cloud native applications with Go and Java for Docker and Kubernetes.* " O'Reilly Media, Inc.", 2020.

[20] G. Shapira, T. Palino, R. Sivaram, and K. Petty, *Kafka: the definitive guide.* " O'Reilly Media, Inc.", 2021.

[21] T. T. L. Nguyen, H. K. Vo, H. H. Luong, H. T. K. Nguyen, A. T. Dao, X. S. Ha *et al.*, "Toward a unique iot network via single sign-on protocol and message queue," in *International Conference on Computer*

Information Systems and Industrial Management. Springer, 2021, pp. 270–284.

[22] F. Tapia, M. Á. Mora, W. Fuertes, H. Aules, E. Flores, and T. Toulkeridis, "From monolithic systems to microservices: A comparative study of performance," *Applied Sciences*, vol. 10, no. 17, p. 5797, 2020.

[23] L. N. T. Thanh *et al.*, "Ioht-mba: An internet of healthcare things (ioht) platform based on microservice and brokerless architecture," *International Journal of Advanced Computer Science and Applications*, 2021.

[24] K. Gos and W. Zabierowski, "The comparison of microservice and monolithic architecture," in *2020 IEEE XVIth International Conference on the Perspective Technologies and Methods in MEMS Design (MEMSTECH)*. IEEE, 2020, pp. 150–153.

[25] R. Lu, K. Heung, A. H. Lashkari, and A. A. Ghorbani, "A lightweight privacy-preserving data aggregation scheme for fog computing-enhanced iot," *IEEE Access*, vol. 5, pp. 3302–3312, 2017.

[26] L. N. T. Thanh *et al.*, "Toward a unique iot network via single sign-on protocol and message queue," in *International Conference on Computer Information Systems and Industrial Management*. Springer, 2021.

[27] ——, "Toward a security iot platform with high rate transmission and low energy consumption," in *International Conference on Computational Science and its Applications*. Springer, 2021.

[28] H. X. Son, M. H. Nguyen, H. K. Vo *et al.*, "Toward an privacy protection based on access control model in hybrid cloud for healthcare systems," in *International Joint Conference: 12th International Conference on Computational Intelligence in Security for Information Systems (CISIS 2019) and 10th International Conference on EUropean Transnational Education (ICEUTE 2019)*. Springer, 2019, pp. 77–86.

[29] H. X. Son and E. Chen, "Towards a fine-grained access control mechanism for privacy protection and policy conflict resolution," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 2, 2019.

[30] S. H. Xuan *et al.*, "Rew-xac: an approach to rewriting request for elastic abac enforcement with dynamic policies," in *2016 International Conference on Advanced Computing and Applications (ACOMP)*. IEEE, 2016, pp. 25–31.

[31] H. X. Son, T. K. Dang, and F. Massacci, "Rew-smt: a new approach for rewriting xacml request with dynamic big data security policies," in *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*. Springer, 2017, pp. 501–515.

[32] S. K. Panda, M. Majumder, L. Wisniewski, and J. Jasperneite, "Real-time industrial communication by using opc ua field level communication," in *2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, vol. 1. IEEE, 2020, pp. 1143–1146.

[33] D. L. Tran, T. Yu, and M. Riedl, "Integration of iiot communication protocols in distributed control applications," in *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, 2020, pp. 2201–2206.

[34] L. Bixio, G. Delzanno, S. Rebora, and M. Rulli, "A flexible iot stream processing architecture based on microservices," *Information*, vol. 11, no. 12, p. 565, 2020.

[35] L. N. T. Thanh *et al.*, "Uip2sop: A unique iot network applying single sign-on and message queue protocol," *IJACSA*, vol. 12, no. 6, 2021.

[36] ——, "Sip-mba: A secure iot platform with brokerless and microservice architecture," *International Journal of Advanced Computer Science and Applications*, 2021.

[37] N. T. T. Lam, H. X. Son, T. H. Le, T. A. Nguyen, H. K. Vo, H. H. Luong, T. D. Anh, K. N. H. Tuan, and H. V. K. Nguyen, "Bmdd: A novel approach for iot platform (broker-less and microservice architecture, decentralized identity, and dynamic transmission messages)," *International Journal of Advanced Computer Science and Applications*, 2022.

[38] E. Simeoni, E. Gaeta, R. I. García-Betances, D. Raggett, A. M. Medrano-Gil, D. F. Carvajal-Flores, G. Fico, M. F. Cabrera-Umpiérrez, and M. T. Arredondo Waldmeyer, "A secure and scalable smart home gateway to bridge technology fragmentation," *Sensors*, vol. 21, no. 11, p. 3587, 2021.

CNN-LSTM Based Approach for Dos Attacks Detection in Wireless Sensor Networks

Salim Salmi*
Engineering, Systems and Applications Laboratory
National Schools of Applied Sciences
Fez, Morocco

Lahcen Oughdir
Engineering, Systems and Applications Laboratory
National Schools of Applied Sciences
Fez, Morocco

***Abstract*—A denial-of-service (DoS) attack is a coordinated attack by many endpoints, such as computers or networks. These attacks are often performed by a botnet, a network of malware-infected computers controlled by an attacker. The endpoints are instructed to send traffic to a particular target, overwhelming it and preventing legitimate users from accessing its services. In this project, we used a CNN-LSTM network to detect and classify DoS intrusion attacks. Attacks detection is considered a classification problem; the main aim is to clarify the attack as Flooding, Blackhole, Normal, TDMA, or Grayhole. This research study uses a computer- generated wireless sensor network-detection system dataset. The wireless sensor network environment was simulated using network simulator NS-2 based on the LEACH routing protocol to gather data from the network and preprocessed to produce 23 features classifying the state of the respective sensor and simulate five forms of Denial of Service (DoS) attacks. The developed CNN-LSTM model is further evaluated on 25 epochs with accuracy, Precision score, and Recall score of 0.944, 0.959, and 0.922, respectively, all on a scale of 0-1.**

Keywords*—*Denial of Service (DoS); Wireless Sensor Networks (WSN); Convolutional Neural Network (CNN); Long Short-Term Memory (LSTM)

I. Introduction

Wireless Sensor Networks is regarded as one of the prominent research topics. The technology is an ideal solution for numerous applications in various fields like telecommunication, military, healthcare, research, and agriculture, amongst others [1]. Aziz et al. [2] reported the application of wireless sensor networks in detecting natural disasters such as earthquakes, flooding, or volcanoes. The widespread WSNs usage has introduced many security threats in the implementation and deployment phase. Wireless sensor networks are susceptible to different attacks due to unique constraints like storage capacity, restricted processing power, and battery power capacity.

People worldwide rely on networking systems to bring new ideas and answers to their issues and help them meet their basic requirements. New and most often used technological innovations include sensors that allow users to receive remote data and utilize it for their specific purpose. Sensors are being used by Internet of Things (IoT) devices [28], which are becoming more popular. Recently researchers' intention over Wireless Sensors Network (WSN) increased, and several research publications have been added over the research repositories. Despite the advantages of WSN, several security loopholes can be exploited to receive DoS attacks. While using WSN applications, users can face several types of security threats that can cause data breaches [29]. Researchers have been attempting to develop new security solutions in order to prevent DoS attacks from succeeding in their endeavors. Several technological advancements have helped develop novel approaches to infiltrate and prevent such attacks. Still, deep learning has brought about the most effective approaches for preventing such security risks and DoS attacks [30].

DoS attacks are attacks on a service (network or application) that overload the service and prevent it from delivering services to the rest of the network or application's users. When a DoS assault is launched, it floods your site or the supporting infrastructure with a large amount of traffic from various sources, often preventing access to the site for the duration of the attack. Cloudflare, for example, is one of the services that provide DoS protection for websites. When it comes to defending against DoS attacks, it might be pretty challenging. Because it is coming at you from all over the Internet and all over the globe, there is almost no way to block the transmission of that deluge of illicit material. You have no control over it. Fortunately, specific DoS attacks may be detected and blocked upstream from the target (with the assistance of the ISP/backbone that hosts the target/victim). In contrast, others transmit data indistinguishable from a genuine user [20].

With limited resources, inadequate infrastructure, and a massive quantity of WSN use on our hands, we were forced to deal with a slew of security challenges. Assaults on the World Wide Web (WSN) are commonly targeted by Distributed Denial of Service (DoS) attacks. DoS attacks may be identified and avoided by several security measures that have been put in place by researchers, but preventing them is not a straightforward task. In order to safeguard WSN against such assaults, researchers are deploying dependable and easy-to-use security measures based on deep learning techniques.

This study investigates the defense mechanism for denial-of-service attacks in wireless sensor networks. The results of this deep learning technique were evaluated on a specialized wireless sensor network dataset called WSN-DS, having numerous normal and numerous attack circumstances to authenticate their efficiency in detecting Denial of Service attacks. The denial of service attacks can take place at any of the layers of the TCP/IP protocol stack [3],[4]. Presented in Table I are the different types of denial of service attacks available in each layer of the TCP/IP protocol stack. However, there is a range of DoS attacks that exist at each layer.

TABLE I. DoS ATTACKS ON TCP/IP PROTOCOL STACK [5]

Protocol Layer	Attacks
Physical Layer	Droplet Attack
	Jamming
	Node Tampering and Obliteration
Data Link Layer	Denial of Sleep
	Power Exhaustion
	Unfairness
Network Layer	Wormhole
	Blackhole
	Homing
	Spoofing, routing control traffic, replaying
	Misdirection
	Selective Forwarding
	Acknowledge Spoofing
	Sybil
Transport Layer	Desynchronization Attack
	SYN Flood
Application Layer	Overwhelming Sensors
	Reprogramming
	Path-based DoS

II. LITERATURE REVIEW

In recent years, there has been a rise in published studies on Wireless Sensors Networks (WSN). Despite the benefits of WSN, it is vulnerable to DoS assaults because of several security flaws. Users using WSN services may be exposed to various security risks, some of which may result in data breaches. DoS attacks are becoming more common, and researchers prevent them from succeeding. There have been several technological breakthroughs that have made it easier to penetrate and protect against these assaults. However, the most successful techniques to avoid such security threats and DoS assaults have been developed using deep learning.

Numerous studies have detected and classified attacks in overall security architecture and wireless sensor network attacks. The study presented by Alsheikh [6] discussed different algorithms, applications, and strategies of machine learning in a wireless sensor network. The study also highlighted some notable challenges facing the performance of wireless sensor networks, such as quality of service (QoS), query processing, security, energy awareness, and event identification, though the study only highlights the qualitative evaluation of this work.

In the work of Gundunz et al. [5], a survey of machine learning solutions for identifying denial of attacks was presented. This study reviewed the DoS discrepancy available at each layer of the TCP/IP protocol stack and concentrated on the network layer attacks.

Sudar et al. [20] proposed an ML model in SDN to identify DoS attacks in KDD99 dataset . They have used SVM and Decision tree algorithm to detect the attacks due to its accurate classification and less complexity. They claimed that the proposed algorithm (SVM) gives a good performance level of 80% .

Anomaly detection in big data analytics addressed by [21]. based on a big data analytics framework , in which the authors handled structured and unstructured data streams and batch processing techniques. The authors used the WIDE backbone dataset gathered in real time . They recognized 5 types of attacks, which are DoS attacks, HTTP flashcrowd attacks, flooding attacks, abnormal UDP and TCP using machine learning. The attack was identified by using 5 supervised machine learning techniques: Decision Trees (DT), Na¨ıve Bayes (NB), Neural Networks (NN),Support Vector Machines (SVM) and Random Forest (RF).

Almomani et al. [22] used eight different machine learning models in detecting DoS attacks which are: Naive Bayes (NB), Decision Trees (DT), Random Forests (RF), Support Vector Machine (SVM), J48, Artificial Neural Networks (ANN), K-Nearest Neighbor (KNN) and Bayesian Networks (BN). They used the WSN-DS dataset for their experiment and performed feature selection based on expert survey. The authors reported that the Random Forest algorithm achieved the best results with a True positive of 99.7% accuracy, out-performing the ANN model with a True positive of 98.3%.

In [23], the authors have proposed a method provides two level of security , they have implemented suspicious detection module In the first level of security , and they imposed machine learning based C4.5 decision tree model in the second level. First inbound traffic is handled by a suspicious data detection engine.If traffic is suspected to be an attack based on entropy values, a temporary alert is generated and sent to OpenFlow switches the controller to save that particular flow. This module facilitates early detection of attacks This module results once again through Level 2 security. This module provides results by analyzing additional characteristics of the traffic. The output of this module is considered the final result. This module helps detect attacks with a low false positive rate. If it is an attack, this module sends an alert to drop packets and remove the flow from the flow table. By using these two levels autors can help for early detection of DoS attack with low false alarm rate .

Wu et al. [24] proposed a CNN+RNN hierarchical neural network, which they named LuNet. It consists of multiple layers of CNN and RNN, both networks learn together from their input data. Their proposed model was tested on the NSL-KDD and UNSW-NB15 datasets [25]. They performed binary and multi-class classification and achieved maximum accuracies of 99.36% and 99.05%, respectively. Both results are in the NSL-KDD dataset.

This research [27] aims to evaluate the effectiveness of machine learning classification algorithms in detecting flooding,grey hole, and black hole distributed denial of service attacks in wireless sensor networks. We conducted our review using a WSN-based dataset, referred to as WSN-DS, and took the accuracy and speediness measures into account. The results show that the J48 approach is the most accurate and fastest way for identifying grey hole and black hole attacks. At the same time, the Random Tree method is the most accurate and fastest method for detecting flooding assaults. The J48 approach is the most efficient for speed, requiring an average of 0.54 seconds of processing time per sample.

III. SECURITY OBJECTIVES IN WIRELESS SENSOR NETWORKS (WSNs)

In wireless sensor networks, the security objectives are essential aspects of WSNs that must be addressed to avoid security compromise of any kind. There has been an ever-growing application of WSNs in penetrating security environments;

nodes are the network interface through which the attack nodes destroy the network. Routing is regarded as a trust-based process within nodes; the process serves as a good platform for attackers to disrupt the network. Security investigations in networks are carried out individually; thereby, networks are usually designed without pre-planning and are employed for a short period. Therefore, it is imperative to implement countermeasures to secure the wireless sensor networks from security attacks.

DoS is one of the most common attacks in wireless sensor networks. Figure 1 presents the wireless sensor networks with denial of service attacks. One profound effect of DoS involves refraining the radio from switching into sleep mode and draining the system battery completely. In the normal operating conditions, operating situations, the energy consumption ratio in the sensor reduces the battery capacity in months, while DoS reduces the battery in days by keeping the transmitter system incorporated in the sensor nodes [7],[8] and [9].

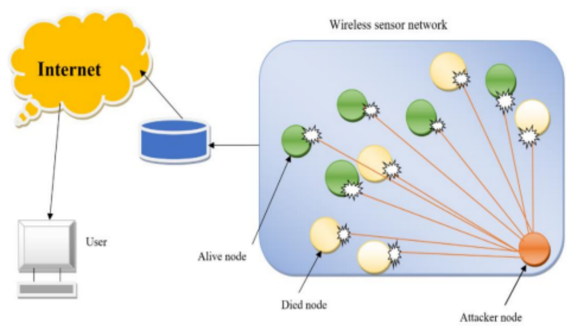


Fig. 1. WSNs with DoS Attack [10].

The security goals are based on the well-known triangle of CIA, namely confidentiality, integrity, and availability of information safety, and this describes what they represent in wireless sensor networks [11].

A. *Confidentiality*

In wireless sensor networks, the two most critical requirements are security and efficiency. There are several applications of WSNs, namely medical, military, research, agriculture, environmental monitoring and others. It is essential to avoid data leakage from sensor networks to neighbouring networks to avoid data confidentiality breaches. Securing the confidentiality of data is essential in protecting the data from attacks like spying [12]. The standard security measure in concealing confidential data is encryption before data transmission with a secret key acknowledged only by a particular receiver. Secure communication channels are established between source and sink, and other secure channels are triggered later if required [13].

B. *Integrity*

Wireless sensor nodes are susceptible to different security attacks threatening the reliability of the data, mainly in the interruption of the flow of information or data fraud [14]. In Sensor networks, transmitted data is considered by nodes to choose the right moves; this further confirms the importance of data integrity. There are two main parts of transmitted data, namely, updated or deleted. To secure data information, data transmitted from the node should arrive at the destination without an alteration in the transmission. The most suitable means of providing data integrity is wireless sensor networks are by checking the data at the receiver end [15].

C. *Availability*

Wireless sensor network nodes should continue operating excellently and not disturbed even when attacked. The implementation of sensors ensures the accessibility of authorized when the data is needed. Information gathered from wireless sensor networks is essential only if the correct user gain access to it at the appropriate time. It is known that WSNs is used in numerous fields, loss of information may lead to damaging consequences. In all the attacks, the most common attack intended at data availability is a denial-of-service attack [16].The CNN-LSTM model was trained using 10 and 25

IV. Methodology

This section covers the steps involved in the data acquisition process, attack detection and classification process, algorithms used, and model design for the research. The block diagram is presented in Figure 2 for the proposed system, which used a CNN-LSTM network to detect and classify intrusion attacks. The model layer explanation, dataset properties, data process, model training, and many other methodologies are discussed in this section. After this stage comes to the main model development stage, then the inference stage, where the performance evaluation for the model was determined.

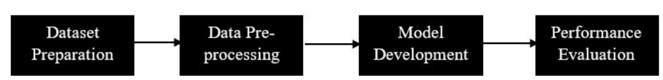


Fig. 2. Block Diagram of the Proposed Intrusion Detection Model.

A. *DataSet*

This research study uses a computer-generated wireless sensor network-detection system dataset developed by Almomani et al. [17]. The wireless sensor network environment was simulated using network simulator NS-2 based on the LEACH routing protocol to gather data from the network and preprocessed to produce 23 features classifying the state of the respective sensor and simulate five forms of Denial of Service (DoS) attacks, namely; Flooding, Blackhole, Normal, TDMA, and Grayhole. WSN dataset was gathered as an intrusion detection dataset tailored towards machine learning and deep learning techniques to identify and classify Denial of Service attacks. 365788 occurrences of records were extracted; it has 19 different attributes. The simulation parameters of the WSN dataset is presented in Table II.

TABLE II. WSN DATASET PARAMETERS [17].

Parameter	Value
Cluster Number	5
Location of the Sink	(50, 175)
Packet header size	25 bytes
Data packet size	500 bytes
Network area dimension	100m×100m
Routing protocol	LEACH
Simulation time	3600s
Nodes number	100 nodes

The WSN dataset has the following data points; Normal has 332040, Grayscale has 13909, Blackhole has 10049, TDMA has 6633, and flooding has 3157 data points.

B. DoS attacks types Description

DoS attacks types are described below [26]:

1) Black Hole attacks: the attacker plays the CH role. Then the attacker will keep dropping packets and not forwarding them to the sink node.
2) Grayhole attacks: the attacker advertising itself as a CH for other nodes. After the forged CH receives packets it selectively or randomly discarding packets, therefore it will prevent the legitimate packets to be delivered.
3) Flooding attacks: flooding attacks targeting LEACH protocol by sending a large number to the sensor to advertise itself as an advertising CH. This will lead to consuming energy, memory, and network traffic.
4) Scheduling attack: It occurs during the setup phase when CHs set up TDMA schedules for the data transmission time slots. The attacker will change the behavior of the TDMA schedule from broadcast to unicast to assign all nodes the same time slot to send data. This will cause a packet collision which leads to data loss.

C. Data Preprocessing

WSN dataset has been employed for testing and assessing intrusion detection techniques. It possesses a good understanding of different intrusion behaviours. Figure 3 presents the importing procedure of the WSN dataset; the dataset was imported to SQL server to implement different statistical measurements values such as occurrences distribution, classes of attacks, and percentage of the occurrences.

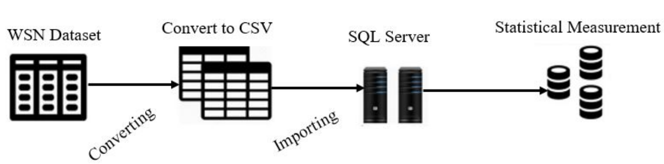


Fig. 3. Importing Procedure of WSN Dataset.

D. Data Split

The dataset was split into two sections: A training set dedicated to training the detection algorithm and a testing set that is completely hidden from the training process. The two subsets use the 80:20 approaches. 80% of the total dataset is used for the training and validation set, while 20% is used for the test set.

V. MODEL DESIGN AND DEVELOPMENT

To detect the intrusion attacks within the WSN dataset, a neural network that uses multi-layers that are interlinked together was used. The model used an eight-layered neural network structure to implement this study. These numerous layer neural networks used the same activation function (ReLU). The multiple layer networks learn over the input data using a selected kernel filter to extract essential features seen as necessary in the intrusion detection system. The developed model has one last layer, the dense layer; softmax activation function was considered for this layer due to its capability to hand classification of multi-classes. The summary of the developed CNN-LSTM model is illustrated in Table III and Table IV.

TABLE III. SUMMARY OF THE CNN MODEL PARAMETERS

Layer(type)	Output Shape	Param #
conv1d (Conv 1D)	(None, 18, 64)	256
conv1d (Conv 1D)	(None, 18, 64)	12352
max_pooling1d (MaxPooling1D)	(None, 9, 64)	0
flatten (Flatten)	(None, 576)	0
dense (Dense)	(None, 64)	36928
dropout (Dropout)	(None, 64)	0
dense_1 (Dense)	(None, 5)	325

TABLE IV. SUMMARY OF THE CNN-LSTM MODEL PARAMETERS

Layer(type)	Output Shape	Param #
conv1d_2 (Conv 1D)	(None, 18, 64)	256
conv1d_3 (Conv 1D)	(None, 18, 64)	12352
max_pooling1d_1 (MaxPooling1D)	(None, 9, 64)	0
conv1d_4 (Conv 1D)	(None, 9, 128)	24704
conv1d_5 (Conv 1D)	(None, 9, 128)	49280
max_pooling1d_2 (MaxPooling1D)	(None, 4, 128)	0
conv1d_6 (Conv 1D)	(None, 4, 256)	98560
conv1d_7 (Conv 1D)	(None, 4, 256)	196864
max_pooling1d_3 (MaxPooling1D)	(None, 2, 256)	0
lstm (LSTM)	(None, 70)	91560
Dropout_1 (Dropout)	(None, 70)	0
dense_2 (Dense)	(None, 5)	355

A. Model Architecture

This section describes the steps taken in achieving the intrusion detection technique. The model takes in an input having an unknown type of attack; the second step involves processing the input data by converting it to an acceptable model format. Then the model carries out a detection process by comparing the features of the present input attack data with the learned features of different kinds of attacks it has been trained with. If the model refuses to detect an attack, the system will return to step 2. If the model detects an attack, then the classification process takes place to ascertain the actual type of attack.

B. Model Hyperparameter Setting

The model hyperparameters are a set of values well-defined to improve the training process of the developed model and its general performance.

The model hyperparameters acknowledged in this study include activation function, epoch, learning rate, verbose, patience, optimization technique, and loss function, as presented in Table V.

These hyperparameters are set at optimal values after many rounds of random search to enhance model optimization. The number of the epoch is the number of times the training data is exposed to the model while training; it is the total number of iterations the whole training data passes through the developed model.

The CNN-LSTM model was trained using 10 and 25 epochs. Activation function was introduced into the model training to incorporate non-linearity effects into the developed model due to the non-linear type of data used. The two activation function used for this study is the softmax function and Rectifier Linear Unit (ReLU).

The softmax function is employed as an activation function in the output layer; it was the selected activation function in the output layer due to its excellent performance when used as a classifier.

Conversely, ReLU is an element-wise activation function; it is fast and straightforward to implement. Also, ReLU is computationally efficient to compute than other kinds of activation functions.

An exponential decay where the learning rate reduces exponentially, a learning rate of 0.001 was optimal for this study.

TABLE V. MODEL HYPERPARAMETER

Hyperparameter	Value
Epoch	10, 25
Activation Function	ReLU, Softmax
Loss Function	Categorical Cross Entropy (CCE)
Optimization algorithm	Adam
Learning rate	0.001
Verbose	1

C. Model Optimization

These are processes employed in ensuring the developed model reach a consistent and efficient level to achieve peak performance.

The adaptive moment estimation (Adam) is the optimizer used to minimize the loss function in this work. Adam is an efficient stochastic optimization that only requires a first-order gradient with its memory requirements. Adam was selected as the preferred choice of optimizer due to requiring a stationary objective.

Categorical Cross-Entropy (CCE) was employed in this study to ensure a better classification process in the CNN-LSTM model. CCE was selected for this work due to its improved choice for cost function, and Ho and Wookey [18] described CCE mathematically using Equation (1).

$$J_{CCE} = -\frac{1}{M}\sum_{K=1}^{K}\sum_{m=1}^{M} W_K \times Y_m^k \times \log\left(h_\Theta\left(x_m, k\right)\right) \quad (1)$$

where M represent the number of training examples, W_K represent the weight for class k, Y_m^k represent the target label for training example m for class k, K represent the number of classes, x_m represent the input for training example m , h_Θ represent the model with neural network weights Θ.

D. Model Implementation and Environment

The study was implemented using Python 3.7.7. Python language was selected as there is a lot of support from an active community for image classification using TensorFlow with Keras [19]. The study was started and completed on a laptop running on core i7, 8GB DDR RAM, a web IDE for Python (Google Colab) with Windows 10 operating system.

E. Performance Evaluation Metrics

The developed model was evaluated using various performance metrics. The assessment metrics used to estimate the model's performance include precision, accuracy, recall, and f1-score.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

where TP is the True Positives, TN is the Tue Negatives, FP is the False Negatives, FN False Negatives.

VI. THE EXPERIMENTAL RESULTS

This section describes the implementation of the intrusion detection model of all classes of attack on the network using the CNN-LSTM model. All the research details, results, and discussion of each experiment are presented. The results of the experiment are shown in different graphs and tables.

A. Detection of Attacks using the Collected Dataset

Attacks detection is considered a classification problem; the main aim is to clarify the attack as Flooding, Blackhole, Normal, TDMA, or Grayhole. Presented in Table VI are the classes of attacks in the dataset and their percentage distribution. Figure 4 presents the graphical representation of all the five kinds of attacks present in the dataset and their distribution.

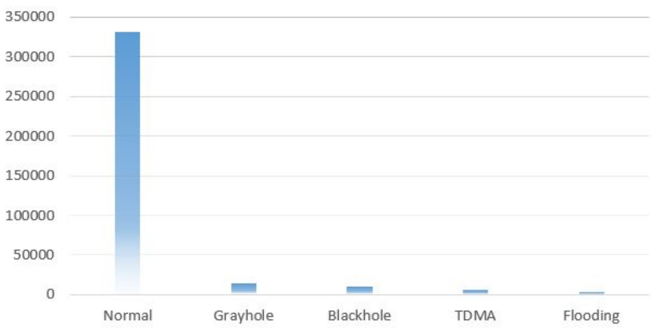


Fig. 4. WSN Dataset Distribution.

TABLE VI. DoS Attacks and Percentage Distribution

S/N	DoS Attacks	Distribution%
1	Normal	90.774
2	Grayhole	3.802
3	Blackhole	2.747
4	TDMA	1.813
5	Flooding	0.863

B. Intrusion Detection Model

The WSN dataset was used to train the CNN-LSTM model. The developed model gives a promising outcome in the attack detection process. The model successfully classifies the given attacks with a training accuracy of 91% on ten epochs and 97% on 25 epochs using a learning rate of 0.001.

The softmax activation function was selected in the output layer of this model due to its capability to handle multi-classification excellently well. The execution time achieved by the CNN-LSTM model on 10 and 25 epochs was 805 and 1103 secs, respectively.

C. Training Phase

During this phase, the training set was employed to train the intrusion detection model. The developed CNN-LSTM model recorded a training loss of 41.7%, training accuracy of 91.07%, validation loss of 47.01% and validation accuracy of 89.44% on 10 epochs, as illustrated in Figures 5 and 6.

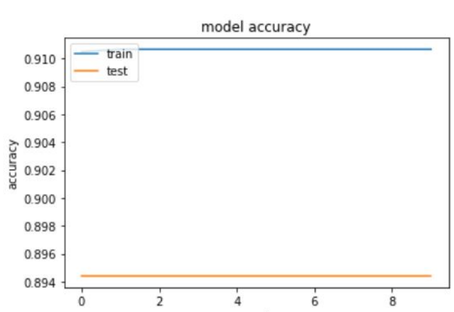


Fig. 5. Snippet of Model Training Accuracy with 10 Epochs.

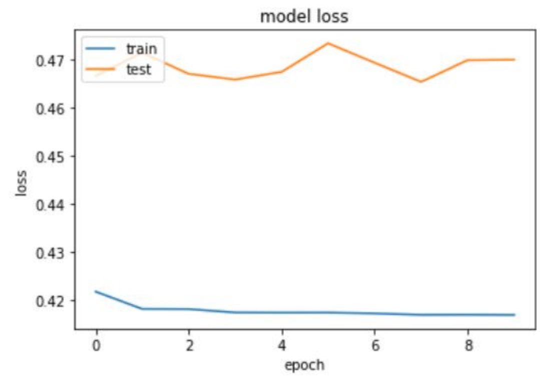


Fig. 6. Snippet of Model Training Loss with 10 Epochs.

The developed CNN-LSTM model recorded a training loss of 8.91%, training accuracy of 96.57%, validation loss of 11.47% and validation accuracy of 94.36% on 25 epochs, as illustrated in Figures 7 and 8.

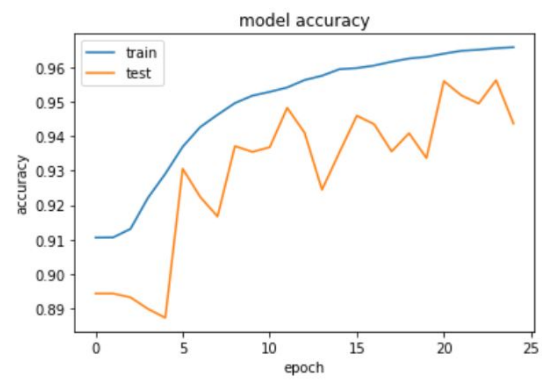


Fig. 7. Snippet of Model Training Accuracy with 25 Epochs.

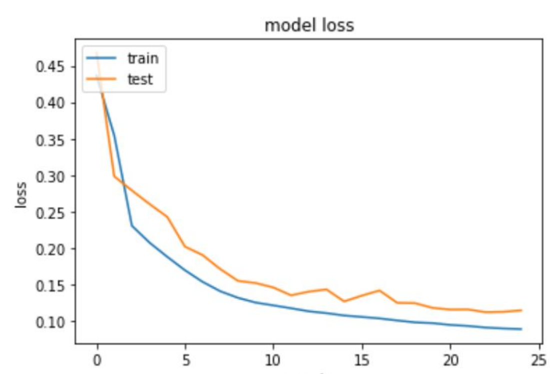


Fig. 8. Snippet of Model Training Loss with 25 Epochs.

D. Performance Evaluation of the Intrusion Model

The entire test set of the overall dataset was tested on the intrusion detection model. The test samples for each attack class were randomly selected. The CNN-LSTM intrusion detection model is evaluated to give the accuracy, Precision score, and Recall score of 0.89, 0.894, and 0.894, respectively, on ten epochs, all on a scale of 0-1. The developed CNN-LSTM model is further evaluated on 25 epochs with accuracy, Precision score, and Recall score of 0.944, 0.959, and 0.922, respectively, all on a scale of 0-1.

VII. Conclusion

Intrusion Detection System is an essential tool used in cyber-security to determine and track intrusion attacks. The rising development of information technology lately has further increased the usage of computer networks for several applications such as finance, business, industry, health and other various aspects of human life. Therefore, developing and deploying secure and reliable networks are critical to information technology administrators. This rapid development of information technology has produced several threats to building a robust and reliable network. There are many kinds of attacks threatening the confidentiality, integrity, and availability of computer networks. Some of these are Flooding, Blackhole, Normal, TDMA, or Grayhole, and they are regarded as harmful attacks.

The DOS attacks are the most common harmful attacks that temporarily denies several services of the end-users, consume computer and network resources. To avoid DoS attacks on computer networks, it is very important to detect and identify the actual type of attacks invading the network. This study developed a neural network model that detects the type of attack affecting the overall system network.

Wireless Sensor Networks Dataset (WSN) having five types of attacks was used in this study. The CNN-LSTM learning model was trained over 10 and 25 epochs with a 0.001 learning rate to ideally detect and classify the attacks. The overall learning algorithm registered a training accuracy of 96.57%; the detection model detected the five kinds of attacks available successfully. The CNN-LSTM intrusion detection model is evaluated to give the accuracy, Precision score, and Recall score of 0.89, 0.894, and 0.894, respectively, on ten training epochs, all on a scale of 0-1. The developed CNN-LSTM model is further evaluated on 25 training epochs with accuracy, Precision score, and Recall score of 0.944, 0.959, and 0.922, respectively, all on a scale of 0-1. The model has successfully extracted essential features of the five kinds of attacks considered.

This study is suitable for detecting intrusion attacks of computer networks, thereby enabling a secured environment for the system's proper functioning.

References

[1] Alsulaiman, L. and Al-Ahmadi, S., 2021. Performance Evaluation of Machine Learning Techniques for DoS Detection in Wireless Sensor Network. arXiv preprint arXiv:2104.01963.

[2] Aziz, N.A.A. and Aziz, K.A., 2011, February. Managing disaster with wireless sensor networks. In 13th International Conference on Advanced Communication Technology (ICACT2011) (pp. 202-207). IEEE.

[3] López, J. and Zhou, J. eds., 2008. Wireless sensor network security (Vol. 1). Ios Press.

[4] Das, S.K., Kant, K. and Zhang, N., 2012. Handbook on securing cyber-physical critical infrastructure. Elsevier.

[5] Gunduz, S., Arslan, B. and Demirci, M., 2015, December. A review of machine learning solutions to denial-of-service attacks in wireless sensor networks. In 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA) (pp. 150-155). IEEE.

[6] Alsheikh, M.A., 2014. S. lin, D. Niyato and H.-P. Tan,". Machine Learning in Wireless Sensor Networks: Algorithms, Strategies, and Applications", IEEE Communications Survers & Tutorials, 16, pp.1996-2018.

[7] Juneja, V. and Gupta, D.V., 2018, August. Security against vampire attack in ADHOC wireless sensor network: detection and prevention techniques. In International Conference on Wireless Intelligent and Distributed Environment for Communication (pp. 25-38). Springer, Cham.

[8] Peng, S., Zhou, Y., Cao, L., Yu, S., Niu, J. and Jia, W., 2018. Influence analysis in social networks: A survey. Journal of Network and Computer Applications, 106, pp.17-32.

[9] Zhang, D., Ge, H., Zhang, T., Cui, Y.Y., Liu, X. and Mao, G., 2018. New multi-hop clustering algorithm for vehicular ad hoc networks. IEEE Transactions on Intelligent Transportation Systems, 20(4), pp.1517-1530.

[10] Osanaiye, O.A., Alfa, A.S. and Hancke, G.P., 2018. Denial of service defence for resource availability in wireless sensor networks. IEEE Access, 6, pp.6975-7004.

[11] Neogy, S., 2015, June. Security management in wireless sensor networks. In 2015 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA) (pp. 1-4). IEEE.

[12] Di Pietro, R., Michiardi, P. and Molva, R., 2009. Confidentiality and integrity for data aggregation in WSN using peer monitoring. Security and Communication Networks, 2(2), pp.181-194.

[13] Perrig, A., Szewczyk, R., Tygar, J.D., Wen, V. and Culler, D.E., 2002. SPINS: Security protocols for sensor networks. Wireless networks, 8(5), pp.521-534.

[14] Acharya, R. and Asha, K., 2008, December. Data integrity and intrusion detection in wireless sensor networks. In 2008 16th IEEE International Conference on Networks (pp. 1-5). IEEE.

[15] Talzi, I., Schönborn, S., and Tschudin, C. 2008 "Providing data integrity in intermittently connected wireless sensor networks," in 5th International Conference on Networked Sensing Systems, 2008, IEEE, pp. 11–18 .

[16] Pelechrinis, K., Iliofotou, M. and Krishnamurthy, S.V., 2010. Denial of service attacks in wireless networks: The case of jammers. IEEE Communications surveys & tutorials, 13(2), pp.245-257.

[17] Almomani, I., Al-Kasasbeh, B. and Al-Akhras, M., 2016. WSN-DS: A dataset for intrusion detection systems in wireless sensor networks. Journal of Sensors, 2016.

[18] Ho, Y., and Wookey, S. (2019). The real-world-weights cross-entropy loss function: Modelling the costs of mislabeling. IEEE Access, Vol. 8, pp. 4806-4813.

[19] Ketkar, N., 2017. Introduction to Keras. In Deep learning with Python (pp. 97-111). Apress, Berkeley, CA.

[20] Sudar, K. Muthamil, et al. "Detection of Distributed Denial of Service Attacks in SDN using Machine learning techniques." 2021 International Conference on Computer Communication and Informatics (ICCCI). IEEE, 2021.

[21] SUDAR, K. Muthamil, NAGARAJ, P., DEEPALAKSHMI, P., et al. Analysis of Intruder Detection in Big Data Analytics. In : 2021 International Conference on Computer Communication and Informatics (ICCCI). IEEE, 2021. p. 1-5.

[22] Almomani, Iman M., and Mamdouh Alenezi. "Efficient Denial of Service Attacks Detection in Wireless Sensor Networks." J. Inf. Sci. Eng. 34.4 (2018): 977-1000.

[23] MUTHAMIL SUDAR, K. et DEEPALAKSHMI, P. A two level security mechanism to detect a DoS flooding attack in software-defined networks using entropy-based and C4. 5 technique. Journal of High Speed Networks, 2020, vol. 26, no 1, p. 55-76.

[24] WU, Peilun, GUO, Hui, et BUCKLAND, Richard. A transfer learning approach for network intrusion detection. In : 2019 IEEE 4th international conference on big data analytics (ICBDA). IEEE, 2019. p. 281-285.

[25] MOUSTAFA, Nour et SLAY, Jill. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In : 2015 military communications and information systems conference (MilCIS). IEEE, 2015. p. 1-6.

[26] AL-AHMADI, Saad. Performance evaluation of machine learning techniques for DOS detection in wireless sensor network. International Journal of Network Security & Its Applications (IJNSA) Vol, 2021, vol. 13.

[27] Wazirali, R., & Ahmad, R. (2022). Machine learning approaches to detect DoS and their effect on WSNs lifetime. Comput. Mater. Contin., 70(3), 4922-4946.

[28] Kopetz, H. (2011). Internet of things. In Real-time systems (pp. 307-323). Springer, Boston, MA.

[29] Abidoye, A. P., & Obagbuwa, I. C. (2018). DDoS attacks in WSNs: detection and countermeasures. IET Wireless Sensor Systems, 8(2), 52-59.

[30] Zaib, M. H., Bashir, F., Qureshi, K. N., Kausar, S., Rizwan, M., & Jeon, G. (2021). Deep learning-based cyberbullying early detection using distributed denial of service flow. Multimedia Systems, 1-20.

A Deep Transfer Learning Approach to Enhance Network Intrusion Detection Capabilities for Cyber Security

Abhijit Das
Research Scholoar, VTU-Belagavi, CSE, PESITM, Shivamogga
Asst. Prof., Dept. of CSE, BNMIT, Bangalore, India

Pramod
Associate Professor, Dept. of ISE
PES Institute of Technology & Management
Affilated to VTU, Shivamogga, India

Abstract*—Cyberattacks are on the rise, making technology companies increasingly prone to data theft. Recent research has focused on constructing cognitive models for traffic anomaly detection in a communication network. Many of these experiments resulted in data packets recorded by technologies like Wireshark. These datasets provide high-dimensional data relating to benign and malicious data packets. Recent research has mostly focused on developing machine learning, and deep learning systems o detect attack data packets in a network. Also, machine learning algorithms are currently trained to detect only known threats. However, with the growth of new cyberattacks and zero-day attacks, current algorithms are unable to detect unknown attacks. This research focuses on detecting rare attacks using transfer learning from a dataset of known attacks. Deep learning outperforms explicit statistical modelling approaches by at least 21% for the dataset used. A preliminary survey of candidate deep learning architectures has been performed before testing for transferability and proposes a Convolutional Neural Network architecture that is 99.65% accurate in classifying attack data packets. The suggested CNN architecture trained with a known attack and then tested its performance on unknown attacks to assess transferability. For this model to extract sufficient information for transferability, the training samples must have more information. Only 20% of the dataset represents current threat data. Several strategies, such as innovative synthetic dataset-based training and bootstrapped dataset training, have been developed to overcome small training sets. A subset of training attacks is determined to optimise learning potential. This study finds training-testing attack pairings with good learning transferability. The most robust and stable relationships are found in DoS attack training-testing pairings. This study also presents model generalisation hypotheses. The dataset features and attack characteristics were analysed using the Recursive Feature Elimination (RFE) algorithm to validate the results.

Keywords*—*Transfer learning; convolutional neural network; intrusion detection system; cyber security; machine learning and deep learning

I. Introduction

A neural network uses knowledge learned from previous training to improve generalization on another similar task in transfer learning. The primary goal of this study is to leverage learning transferability to select attacks from the training dataset that causes the model to generalize for other attacks, which can be extended to rare attacks.

In computer networks that employ machine learning algorithms for network security, it is difficult to provide guarantees about the kind of attacks that the network can expect to see and is susceptible to, especially with the rise in the number of novel attacks. Current machine learning algorithms are usually trained to detect a set of known attacks or learn from attacks performed in the past. Therefore, it is hard to predict the range of attacks against which the employed machine learning algorithm is robust. In the literature, there is proof of concept for both classical ML and DL approaches for anomaly detection [1].

Transferability studies can help us understand the range of attacks against which the system is actually robust based on the attacks that the algorithm has been trained to detect and how those attacks enable the model to generalize for other attacks. This is especially useful when the trained model is able to generalize to unknown attacks without being trained on them explicitly because this provides a more extensive range of attacks that the network may be secure against attacks. Studying these attack correlations indicates the attacks that the model can detect and the accuracy of these unknown attacks. This helps us evaluate the risks of what the model can predict if the observed correlations are consistent and what kind of protection the algorithm provides at what computational cost. Transferability studies could discover that the model scales to other different attacks it has learned to detect without being trained on them explicitly. It assures detection for an even broader range of attacks than the ones it has seen. For groups of similar attacks or correlated attacks, transfer learning enables us to find representative attacks from that group. This helps us to train the model with a smaller number of attacks yet achieve the same level of security as it would have if rained with all the attacks in the group.

The increasing machine learning applications have also seen deep learning models going into hardware chips. The advantage of transferability is that we could identify smaller training sets yet make the model generalize well for a broader range of attacks. This decreases training times, making the process of training computationally efficient and having lower memory requirements. This is important, especially if this algorithm is deployed on a resource-constrained device while performing at par with a model trained on a full training set.

Increasing research in this domain has led to an increase in network traffic datasets. Some of the common datasets used are CAIDA 2007, DARPA 98, KDD 99, CSE-CIC-IDS 2018, and UNSW-NB15, to name a few. This work uses the traffic dataset CICIDS 2017 for training the proposed model. Traffic monitoring software like Wireshark has been used for logging and monitoring network traffic packets for creating these datasets. While these datasets provide a complete description of network packets, they have a high dimensional feature space. In a preliminary study of a potentially suitable neural network, a model has been developed to extract meaningful data from high dimensional training data and a limited number of attack data packets. After identifying a suitable network, the work focuses on attack correlations and comparisons for why training the model with one attack scales for another and if these correlations are symmetric or asymmetric. This study describes the work's contributions in three sections: first, Developing a DNN architecture; second, Testing the proposed architecture for transferability and third, hypothesis for attack correlations.

Developing a DNN architecture: A comparative study of a Hidden Markov Model has been carried out as a statistical modelling-based approach and a candidate Convolutional LSTM Deep Neural Network (CLDNN) architecture and shows that the deep learning approach scales better for the used dataset. Further, deep learning models have been explored for the classification task. The proposed model's candidate DNN architectures include a Convolutional LSTM Deep Neural Network (CLDNN), Convolutional Neural Network with BGRU recurrent layers, and One Class Neural Networks (OCNN), and a Convolutional Neural Network (CNN). The performance of these models has been evaluated when trained and tested on all attacks in the dataset. Also, the learning transferability has been evaluated. All the above architectures had overall accuracies of above 90%; however, this dataset is highly unbalanced, with a majority of the dataset consisting of benign data packets. The models are able to detect benign attack packets with high accuracy, thereby boosting the overall accuracy of the models. Therefore, a major criterion for architecture selection is not the overall accuracy but the percentage of attack data packets it is able to classify correctly. Attack transferability has been studied with some of these architectures; however, none of the candidate models other than the proposed architecture showed very strong correlations between training-testing attack pairs. The proposed architecture is the model that shows maximum attack data classification accuracy for the multi-class classification problem. This architecture is able to perform deep feature extraction using flow-based features not only for adequately sampled attacks but also for severely underrepresented attacks.

Testing the proposed architecture for transferability: Attack transferability has been tested with the proposed model architecture by training the model with one kind of attack from the training dataset and testing it on another attack. The entire dataset consists of 20% of attack data that are further divided into fourteen categories of attacks. Therefore, the representation of each attack class is low. For learning transferability, the model requires a larger number of attack data packets. Two techniques have been used to increase the number of attack data packets to address this problem: 1) Using SMOTE generated data and 2) Using a bootstrapped dataset. A Synthetic Minority Oversampling Technique (SMOTE) was used for the first method to generate more attack data packets in each attack class synthetically. For the second method, the original attack data is resampled, shuffled and added to the training dataset to match the number of benign data packets. Furthermore, exploring the possibility that the model may exhibit higher testing accuracies for a particular attack if it is trained with a subset of other attacks in the dataset. An attack boosting algorithm is employed to select a subset of attacks from the training dataset to maximize the classification accuracy when tested on a particular attack. **Hypotheses for correlated attacks:** The study of attack correlations reveals training testing attack pairs that exhibit strong correlations with each other. A hypothesis has been proposed for these observed attack correlations. The attack characteristics have been studied and important features to validate the hypotheses. The Recursive Feature Elimination (RFE) algorithm is used to identify the dominant features for learning. This not only reduces the dimensionality of the training data but also reduces training times. This study makes inferences about the properties of the features selected and the number of features selected for a pair of correlated attacks. This results in identifying usable attacks in this dataset for transfer learning and explaining why they scale for certain attacks.

II. Background and Related Work

The dataset used is the Canadian Institute of Cybersecurity Intrusion Detection Systems dataset: CICIDS 2017. Overall, the dataset consists of 8 separate CSV files, with data corresponding to attacks simulated in 8 sessions included. The entire dataset consists of data corresponding to 14 types of attacks and benign traffic. The dataset describes 78 flow features. 80% of the data entries are benign data, and 20% of the dataset consists of attack data. The dataset was created by simulating 14 different types of attacks that are shown below in Table I.

TABLE I. Dataset Composition

Classes	Attack	Percentage of data
0	BENIGN	80.3
1	Botnet	0.069
2	DDoS	4.52
3	DoS GoldenEye	0.36
4	DoS Hulk	8.16
5	DoS Slowhttptest	0.19
6	DoS Slowloris	0.2
7	FTP-Patator	0.28
8	Heartbleed	0.00038
9	Infiltration	0.0012
10	PortScan	5.61
11	SSH-Patator	0.2
12	Web Attack Brute Force	0.053
13	Web Attack SQL Injection	0.00074
14	Web Attack XSS	0.023

An attacker gains control of several machines and servers connected to the internet and uses these servers to carry out cyberattacks against a target. Due to multiple synchronized attacking machines, the attack data volumes are huge, thereby facilitating a strong attack against the target machine. Botnets are often used to carry out DDoS attacks. Denial of Service (DoS) attacks are attacks aimed at making a server unresponsive to the requests of legitimate users. They usually work by sending enormous traffic, thereby flooding the server and hitting its resource pool by requesting resources or sending obfuscated data that make the server unavailable to real requests.

Distributed Denial of Service (DDoS) attacks are DoS attacks conducted by different servers at once on a victim server. These servers have different IP addresses, thus making it difficult to track down one single IP address for attacker identification, which makes it hard to mitigate the attack. In the case of flooding attacks, the volume of data generated to flood the victim server is larger, making the DoS attack strong and harder to terminate[2]. DoS: A goldenEye is an attack tool that identifies vulnerabilities in a target server. It explores the capability of victim servers to form multiple HTTP connections, thereby using up all possible connections on that server. This attack can be operated as a distributed attack as well. DoS GoldenEye uses Keep-Alive headers and Caching Control options to keep the connection alive, preventing the target server from shutting down the connections and making the server unresponsive to non-attack requests.

DoS: HULK is HTTP Unbearable Load King attack. A flooding attack floods the target servers with a large volume of HTTP requests, requesting data or resources or sending unclear HTTP packets. This floods the target machine with HTTP data packets, and the load is unbearable for the server, which makes the server unreachable.

DoS: Slowhttptest and DoS: Slowloris are attack tools for carrying out Slow HTTP attacks. **Slow HTTP attacks** form multiple connections with the target server and try to keep the connections open. Some slow HTTP attacks keep the connection open by declaring a large amount of data to be sent and sending the data at very large intervals of time, almost equal to the timeout period. However, the server cannot close the connection, and the connection does not take time out. This keeps all available connections alive and keeps the server from responding to legitimate requests. Slow HTTP attacks can be made using Slow Header attacks, where the packet header arrives at large intervals or Slow Body attacks, where the body of the data arrives very slowly.

Brute Force attacks are targeted towards gaining access to authentication keys by trying all possible combinations of the key. File Transfer Protocol is a network protocol that can be used to transfer files. Users connect to the server using an FTP client using the username and password authentication. Brute Force FTP attacks are an attack on the username and password. Brute Force SSH gains access to valid login credentials to authenticate SSH access to a server by trying all possible combinations.

Heartbleed could be classified as a protocol based attack where the attack utilizes a packet header field required for the transport layer security protocol that causes a machine to dump out its entire memory, including confidential and protected data. It does so without leaving traces in the target machine, making it undetectable. In an infiltration attack, the attacker gains access to a protected network or system and finds vulnerabilities in the machines or devices connected to the network. After identifying vulnerabilities, the attacker attacks the machine or device to steal private information.

In an **Port Scan attacks**, the attacker scans the ports of the target server and sends requests to a range of ports on the target server. The attacker finds an active port on the server and exploits a known vulnerability of that service to attack the target server.

Web Attacks identify weaknesses in web applications to gain access to private or protected data. These weaknesses could be exploited using data entry or using injection attacks where malicious scripts are injected into data entries of otherwise harmless websites or by using brute force. For example, for a web application that uses an underlying SQL database and has an option for user inputs, malicious inputs could be given that cause the database to dump out confidential information. DDoS attacks are also a kind of web attack.

It can be observed from Table 1 the CICIDS 2017 dataset is highly unbalanced, with 80% of the dataset consisting of benign data and 20% of the dataset consisting of attack data packets. The attack data is further divided unequally into fourteen different types of attacks. Because of the unequal division, certain attack classes have a very sparse representation, for example, just 11 or 36 data packets throughout the entire dataset. This makes it difficult to generate realistic synthetic data to augment the number of data packets and limits model learning even when using a bootstrapped dataset for these classes of attacks. Therefore, when the proposed architecture has been tested for attack transferability. It eliminates those classes of attacks that cannot use for the transferability of learning.

In the literature, there are different machine learning techniques that deal with intrusion detection using flow characteristics. Alkasassbeh et al. explore MLP, Naive-Bayes, and Random Forest classification algorithms for Distributed Denial of Service attack detection and show that MLP achieves the highest accuracy [3]. Lopez et al. presents a study of machine learning techniques for traffic anomaly detection and proves that a Random Forest-based decision classifier is the best model for anomaly detection, and a Dense Neural Network is a good classifier for some types of DDoS attacks with methods to boost a number of attack samples of underrepresented attack types [1]. Vinayakumar et al. carry out a comprehensive study of DNNs and Machine Learning classifiers that learn abstract and high dimensional data representation using the KDDCup99 dataset and test their model performance on other datasets, such as NSL-KDD, UNSW-NB15, Kyoto, WSN-DS, and CICIDS 2017. They also propose a hybrid DNN framework which can be deployed in real-time to monitor network traffic and events to detect possible network attacks [4]. Sharafaldin et al. generate a dataset consisting of benign and seven types of attack data. They also evaluate the performance of machine learning algorithms to identify the best subset of features for certain types of attacks [5]. Ferrag et al. analyze RNNs, DNNs, restricted Boltzmann machines, Deep Belief Networks, CNNs, Deep Boltzmann machines and deep autoencoders for traffic data classification using the CSE-CIC-IDS2018 dataset and the Bot-IoT dataset [6].

To find out about an undiscovered attack, researchers have chosen TL-based IDS more. Zhao et al. [7] proposed an algorithm that used the NSL-KDD dataset for binary classification mapped the source and target domains into the latent space K. In their method, the R2L attack was found by looking at DoS data from the source domain. The same author made their TL even better in their next work by adding clusters [8]. They found that 500 data points are enough to train the NSL-KDD dataset well. It was shown by Taghiyarrenani et al. [9] that transfer learning can work even if the source and target data are different from each other. They were able to tell the difference between normal traffic and traffic that wasn't normal. Their results were better when the source domain had a lot of labels and the target domain didn't have as many labels.

Wu et al. [10] also used a CNN to pass on knowledge learned from one dataset to another. When they researched, the UNSW-NB15 dataset was used as the source data, and the NSL-KDD dataset was considered the target data. As the datasets were different, they used two CNNs in their method. Afterwards, a fully connected layer used the data to make the final classification for the data from NSL-KDD. Singla et al. [11] found that transfer learning can help when there isn't enough training data to learn new attacks. Their approach isn't very different from what [10] had done. They used binary classification to look for a specific attack in the UNSW-NB15 dataset. They used the remainder of the attacks using the same dataset as the original domain. With the normal DNN, they found that it did better when there was less training data for their transfer learning-based approach.

In 2020, Masum et al. [12] used the same computer vision algorithms on IDS that they used on PCs. As image processing needs two-dimensional data, they turned the intrusions dataset into two-dimensional data, and then used VGG-16 [13] to look for intrusions. However, they were able to get about 95% accuracy, which some people might think isn't enough for the real world because other algorithms have been able to get more than 95% accuracy in the past. The reason could be that the features that were made with VGG-16 might be better for image processing than intrusion detection. Second, Dhillon et al. [14] have shown how TL can be used in real-time. A CNN-LSTM algorithm was made. Then, they moved their model to a new place where they could use data that wasn't there to figure out whether it was an attack or not. Research done by them shows that there is no difference between the space for features and the space for labels in both fields. They were able to speed up their research by using TL.

Vijayakumar, Alazab et al. [15] evaluated the performance of CNN for intrusion detection. CNN was investigated in their work both directly and in combination with RNN, LSTM, and GRU algorithms. CNN and recurrent algorithms were found to perform nearly identically when used together. Even though they could get good outcomes, the computational power required by their approach was excessive.

III. METHODOLOGIES

A. Hidden Markov Models: A Statistical Modeling based Approach

HMM is a Markov Model in which the process being modelled is a Markov Process. Markov models are used to model processes that change stochastically. This study considers that the phenomenon being modelled is a first-order Markov Process where the next state of the phenomenon is only dependent on the current state of the phenomenon. This is called the Markovian Property. If q_t represents the state of a model at time instant t, then the conditional probability of its next state will be as given:

$$P(q_t = S_i|q_t - 1 = S_j, q_t - 2 = S_k, ...) = P(q_t = S_i|q_t - 1 = S_j)$$

Where S_i is the current state and S_j, S_k represents the past states.

In HMM problem, the observations are the features corresponding to a data sample. Forty flow features from the dataset are taken for each data sample and are fed to each HMM as 40 observations. Each CSV file in the dataset is treated as a separate dataset. The performance of HMMs has been evaluated on 2 class and 4 class classification problems. In those cases, 2 HMMs and 4 HMMs have been used, respectively, for training. Observations are taken at a time from the testing dataset, and each model is tested with the T observations. The model that generates the highest probability $P(O|\lambda)$ implies that the set of observations O belongs to that HMM and class.

Computation of $P(O|\lambda)$ using hidden states. Let Q be the set of hidden states:

$$O = O_1, O_2, O_3, ..., O_T \ where \ T = 40.$$

$$P(O|\lambda) = P(O_1, O_2, O_3, ., O_T|\lambda) = \sum_Q PP(O|Q, \lambda)P(Q|\lambda)$$

Where Q is the set of all possible hidden states, the complexity is of the order $O(2TN^T)$ for a general system with N states and T observations to compute $P(O|\lambda)$ using the method above. A forward-backwards algorithm is used to compute the same quantity. This reduces the computational complexity to $O(N^2T)$.

B. Neural Network Architectures

Deep Neural Network architectures have been tested to identify the best architecture to use to study attack correlations. The first two architectures that have been tested are a CLDNN and CNN with BGRU layers. These neural networks are part of a class of neural networks called CRNN, which is CNN with Recurrent Layers. Convolutional layers have been used because they capture spatial features from the dataset and gated RNNs to capture temporal patterns in the sequence. Bidirectional GRUs capture temporal patterns not only forward in time but also backward in time. In addition, the performance of a One Class Neural Network (OCNN) has been evaluated as it is good for minority data detection in large datasets. Finally, the performance of a Convolutional Neural Network without Recurrent layers has been evaluated.

1) ***CLDNN:*** The Convolutional LSTM Deep Neural Network architecture has 6 hidden layers. The first layer is a convolutional layer with 256 feature maps, a (1, 3) convolution kernel and 20% dropout. The second hidden layer is a convolutional layer with 256 feature maps and a (2, 3) convolution kernel. The third layer is a convolutional layer with 80 feature maps and a (1, 3) kernel with 20% dropout. The fourth hidden layer is a convolutional layer with 80 feature maps and a (1, 3) kernel. The fifth hidden layer is an LSTM layer with 50 cells. The sixth hidden layer is a fully-connected layer with 128 neurons. The CLDNN architecture has shown in Table II. All hidden layers have Rectified Linear Unit(ReLU) activation as:

$$g(x) = max(0, x)$$

This architecture was used to evaluate its performance on i) The multi-class classification problem when it is trained with all classes of data and tested on all classes of data and ii) The two-class classification problem when it is trained with one class of data and tested on all other classes of data. For the multi-class classification problem, the output layer has 15 output classes. For the two-class classification problem, the output layer has two output classes. In both implementations, the output layer has softmax activation. Softmax activation can be given as:

$$\sigma_i(z) = \frac{e^{z_i}}{\sum_{j=i}^{J} e^{z_j}}, \; where\ J\ is\ the\ total\ number\ of\ classes$$

TABLE II. CLDNN ARCHITECTURE

CLDNN Architecture
Conv2D 1x3x256, ReLU, Dropout(rate=0.2)
Conv2D 1x3x80, ReLU, Dropout(rate=0.2)
Conv2D 2x3x256, ReLU
Conv2D 1x3x64, ReLU
LSTM 50
Dense, 128, ReLU
Dense, 2, Softmax

All 78 features are used for training. For the two-class classification problem, the model has been trained on one kind of attack and evaluates its performance when tested on all the other attacks. In this case, all benign data is labelled zero, and all attack data is labelled one. For the multi-class classification problem, the model is trained with all 15 classes of data. The total dataset is divided into a fifty-fifty split; 50% of the data is used for training, and 50% of the data is used for testing. The loss function used is Categorical CrossEntropy, and the optimizer is Adam. The batch size is 1024. This model is trained for 50 epochs. The categorical CrossEntropy loss function is given by: $L_{CE} = -\frac{1}{N}\sum_i t_i log(p_i)$.

where t_i is the true label of the i^{th} data point, p_i is the predicted probability of the data point belonging to class t_i and N is the batch size.

*2) **CNN with BGRU layers and CNN with stacked BGRU layers**:* Gated Recurrent Units are different from traditional RNNs without gates as they are able to control the amount of previous data to carry over to the next iteration using hidden states. GRUs usually have a reset gate and update gate as opposed to a forget gate, input gate and output gate as in LSTMs. This makes training the model with GRUs faster than using LSTMs. BGRUs have connections that go backwards in time, enabling them to capture temporal patterns backwards and forward in time. The CNN with Bidirectional Gated Recurrent Units and Stacked Bidirectional Gated Recurrent Units architecture is the same as the architecture of the CLDNN used, except in these architectures, there are BGRU and stacked-BGRU layers in place of the LSTM layer. Each BGRU layer has 50 cells. For a stacked-BGRU approach, there are two BGRU layers consecutively with 50 cells in each layer. Each model is trained for 150 epochs, with a batch size of 1024. The loss function used is mean squared error, and the optimizer used is RMSProp. Table III and Table IV are the architectures of these two DNNs.

*3) **One Class Neural Networks**:* One Class Neural Networks (OCNN) are used for anomaly detection in complex datasets that require highly non-linear decision boundaries. The loss function of OCNNs is derived from the loss function of OC-SVMs. In an OC-SVM, for input data X with N input samples $\left\{_{X_i}\right\}_{i=1}^{N}$. A $\phi(X_n)$ is a reproducing kernel Hilbert

TABLE III. CNN WITH BGRU ARCHITECTURE

CNN with BGRU Architecture
Conv2D 1x3x256, ReLU, Dropout(rate=0.2)
Conv2D 2x3x256, ReLU
Conv2D 1x3x80, ReLU, Dropout(rate=0.2)
Conv2D 1x3x64, ReLU
BGRU 50
Dense, 128, ReLU
Dense, 2, Softmax

TABLE IV. CNN WITH STACKED BGRU ARCHITECTURE

CNN with Stacked BGRU Architecture
Conv2D 1x3x256, ReLU, Dropout(rate=0.2)
Conv2D 2x3x256, ReLU
Conv2D 1x3x80, ReLU, Dropout(rate=0.2)
Conv2D 1x3x64, ReLU
BGRU 50
BGRU 50
Dense, 128, ReLU
Dense, 2, Softmax

space that is a mapping from the input space to feature space. All the input data points are labelled one, and the only negative point is the origin. A hyperplane separates the origin from the mapped $\phi(X_n)$s. The hyperplane in the feature space is given by $f(X_n) = w^T\phi(X_n) - r$. v is the parameter to control the trade-off between maximizing the distance and the number of data points falsely classified as positive. r is the hyperplane bias. An OCNN has a feedforward neural network with one hidden layer and one output node. Here, w is the scalar output from the hidden to the output layer, and V is the weight matrix from the input to the hidden layer. The hidden layer has a linear or sigmoidal activation given by $g(.)$

The optimization problem for OCNNs is:

$$min_{w,V,r}\frac{1}{2}\left\|w\right\|_2^2 + \frac{1}{2}\left\|V\right\|_2^F + \frac{1}{v}\cdot\frac{1}{N}\sum_{n=1}^{N} max(0, r - \langle w, g(VX_n)\rangle) - r \tag{1}$$

(w, V) are updated using normal backpropagation. The model is trained using important features extracted from an autoencoder instead of using the raw data. The OCNN algorithm is implemented by Chalapathy et al. [16]. The performance metrics have been calculated by True Positive Rate(TPR), False Positive Rate and (FPR)ROC curve.

*4) **CNN**:* The Convolutional Deep Neural Network architecture consists 5 hidden layers. The first hidden layer is a Convolutional layer with 256 feature maps and a (1, 3) convolution kernel. The second hidden layer is a convolutional layer with 256 feature maps and a (2, 3) convolution kernel. The third hidden layer is a convolutional layer with 256 feature maps and a (1, 3) convolution kernel and a dropout of 20%. The fourth hidden layer is a convolutional layer with 80 feature maps and a (1, 3) convolution kernel. The fifth hidden layer is a fully-connected layer with 128 neurons. The output layer has 2 output nodes and sigmoid activation. Layers 1,4 and 5 have ReLU activation and layers 2 and 3 have Leaky ReLU activation with $\alpha = 0.3$. All 78 features from the dataset were used for training. This model's performance is evaluated on the multi-class classification problem. The architecture of CNN has shown in Table V. The model is trained on 70% of data and tested on 30% of the data. 10% of the training data is used as validation data. The model is trained for 100 epochs. The activation function used is Leaky ReLU and Sigmoid. The loss function used is Binary Cross-Entropy(BCE).

TABLE V. CNN ARCHITECTURE

CNN Architecture
Conv2D 1x3x256, ReLU
Conv2D 2x3x256, LeakyReLU(0.3)
Conv2D 1x3x256, LeakyReLU(0.3), Dropout(rate=0.2)
Conv2D 1x3x80, ReLU
Dense, 128, ReLU
Dense, 2, Sigmoid

C. Attack Correlations

Base Architecture Trained on Original Dataset: This section studies the transferability of learning using the proposed base architecture, training the CNN architecture with one attack, and testing on all the other attacks. A two-class classification problem has been considered to implement this. All attack data has been labelled as one class and all benign data as another class. The model has been trained on benign data and attack data corresponding to the attack whose transferability has to study. As seen in Table I, attacks like Attack 8 (HeartBleed), Attack 9 (Infiltration) and Attack 13 (Web Attack: SQL Injection) are severely underrepresented in the dataset. A synthetic and bootstrapped dataset has been discussed to address this problem. In this section, the models are trained with only the original dataset. In confusion matrices in the results section, Attacks 8, 9, and 13 are not included as there are not enough original samples in the dataset for transferability in learning. The model is trained for 20 epochs for each attack, the loss function used is Binary CrossEntropy, and the optimizer used is Adam.

Base Architecture Trained on Synthetic and Bootstrapped Dataset: The CICIDS 2017 dataset is unbalanced, with 80% of the data samples being benign data packets and 20% of the data samples belonging to attack data. The total attack data is further divided into fourteen classes of attacks; thus, each class is severely underrepresented in the dataset, which is not ideal for learning. The main goal of this section is to test the model for strong and consistent attack correlations when it is trained with a greater number of samples from the minority attack classes.

Synthetic Minority Oversampling Technique (SMOTE): More attack data has been generated per class using Synthetic Minority Oversampling Technique (SMOTE). SMOTE selects examples that are close to each other in the feature space. This technique selects one point from the minority attack class and uses K nearest neighbours algorithm to find the K nearest data points around the randomly chosen point. The number K is chosen based on how much minority classes have to oversample. In this implementation, 5 nearest neighbours have been chosen. It then randomly selects one neighbour point out of the K neighbours and creates a synthetic data point on the line joining the two selected points. Using this technique, specific regions have been identified in the feature space that can generate more samples belonging to a certain class of data. The number of attack packets generated using this technique is used to match the number of benign data packets in the dataset. This process is repeated for each attack. The model trained with this dataset is tested on: 1) SMOTE generated benign data and real attack data for evaluation of accuracy when tested on a synthetically generated benign dataset and 2) Real benign data and real attack data as this scenario best emulates a real-life scenario.

Bootstrapped Dataset: Another technique used to create a balanced training dataset is for every attack that the model is trained with, the attack data is resampled to match the number of benign data packets. This training dataset is bootstrapped dataset. This process is used to replicate attack data corresponding to all attacks. The base architecture is trained with this enhanced dataset and tested on real attack data. During training, it is validated on 20% of the attack data. During training with the SMOTE dataset as well as a bootstrapped dataset, the model is trained for 20 epochs per training attack, the loss function used is Binary CrossEntropy, and the optimizer is Adam.

Attack Boosting Algorithm: The previous sections explore transferability in learning when the model is trained with a single attack and tested on other attacks. This section aims to identify a subset of training attacks that can be used to train the model to cause it to scale for other attacks. The possibility of the model scaling for a certain test attack has been explored when trained with a combination of training attacks (that do not include the testing attack). The aim was to observe consistent and strong off-diagonal correlations and identify training attacks. To do this, the attack boosting algorithm has been used as outlined in Algorithm 1. 50% of the set is used only as training data for selecting the training attacks, and the model performance is evaluated on the remaining 50% of the set that serves as the validation data. The testing data consists of the attack for which we want to select a subset of training attacks along with benign data. The loss function used is Binary CrossEntropy, and the optimizer used is Adam.

Algorithm 1 Attack Boosting Algorithm

```
Require: Datasets
Result: Selected Training Attack List
while Accuracy increase=1 and remaining attack list is not empty do
    for single attack in remaining attack list do append single attack data to the current training set (without updating current train- ing set), train model and evaluate on validation set
        if accuracy > appending accuracy then
            appending attack=single attack
        end
    end
    if appending accuracy > current accuracy then
        current accuracy=appending accuracy
        update selected attacks list
        remove appending attack from remaining attack list
        update Accuracy increase
    end
end
```

$\frac{a+b}{C}$

D. Hypothesis for Attack Correlations

In this section, we propose hypotheses for observing the correlations as in Fig. 5.2. In doing so, we use the Recursive Feature Elimination Algorithm [17] to identify the important and dominant features for model learning. We propose hypotheses based on the number of important features selected by the algorithm as well as the properties of the attacks and features selected.

Recursive Feature Elimination is a feature selection algorithm used to identify the important features. A different ML algorithm is used to identify important features in a given feature set and wrapped by the RFE algorithm to do so. This ML algorithm at the core of the wrapper does not have to be the same as the classifier used for the classification task at hand, and it can be other algorithms like Decision Trees, Random Forests, to name a few. RFE provides us with a subset of the entire feature set that is important for learning. This is especially important for datasets with a large number of features to identify the important features for learning and to reduce training times significantly. RFE works by recursively training the model with reduced features and removing features without which the model performs with higher accuracy. RFE also ranks the features in order of importance. The top n features can be selected based on this algorithm.

The CICIDS 2017 dataset has 78 features, a relatively large number of features. The main goal is to identify a smaller subset of features that are important for the learning and scalability of the proposed model and speed up training times. Two implementations of the RFE algorithm have been done in the proposed model. A single attack has been selected from all the attacks for the first implementation. The RFE algorithm is used to identify features that the classifier uses to separate from the benign data. This process has been repeated for all attacks in the dataset in order to identify the common features between attacks to validate the proposed hypotheses. The second implementation involves taking two correlated attacks (as observed in Figure 5) and using the RFE algorithm to identify the number of features that the classifier uses to differentiate both attacks from benign data. The Decision Tree Classifier has been used as the core model for RFE. The training data for both these implementations contain the attack and benign data.

IV. Result and Discussion

A. Comparison between HMM Performance and CLDNN Performance

To observe if the statistical modelling-based approach scales better than the deep learning approach, the performance of the HMM compared with a candidate CLDNN model (Table VI). It has been observed that the CLDNN model is able to achieve an overall accuracy of 99.69% for 2 classes of data and 99.57% for 4 classes of data. In contrast, the HMM model is only 78.7% accurate for 2 classes of data and 50.8% accurate for 4 classes of data. This could be because not all the attacks in the dataset are multi-stage attacks that correspond to the hidden states of the HMMs. The performance of the HMM model deteriorates as the number of classes in the dataset is increases.

TABLE VI. HMM vs CLDNN

Number of Classes of data in dataset	HMM accuracy	CLDNN accuracy
2 Classes	78.7%	99.69%
4 Classes	50.8%	99.57%

Table VI concluded that the deep learning approach shows more promise for the multi-class classification problem. Therefore, more deep learning architectures have been explored in the following sections to identify the best architecture to study attack transferability patterns. The next section gives the architecture details of the used candidate CLDNN model.

B. Result of the Neural Network Architectures

1) ***Result of the CLDNN Architecture on the Multi-class Classification Problem:*** An attack accuracy metric is used to evaluate the performance of a model other than the overall accuracy metric [18]. This metric was used because the number of benign packets is much larger than the attack data. A gross misclassification of the attack data packets does not diminish the overall accuracy considerably. A more accurate metric of performance evaluation would be the percentage of correctly classified attack data. Therefore, the attack accuracy is defined as the ratio of the number of correctly classified attack packets in a class to the total number of attack packets in that class. This is also called true positive rate more generally.

$$AttackAccuracy = \frac{Number\ of\ correctly\ classified\ attack\ packets\ in\ class\ i}{Total\ number\ of\ attack\ packet\ packets\ in\ class\ i}$$

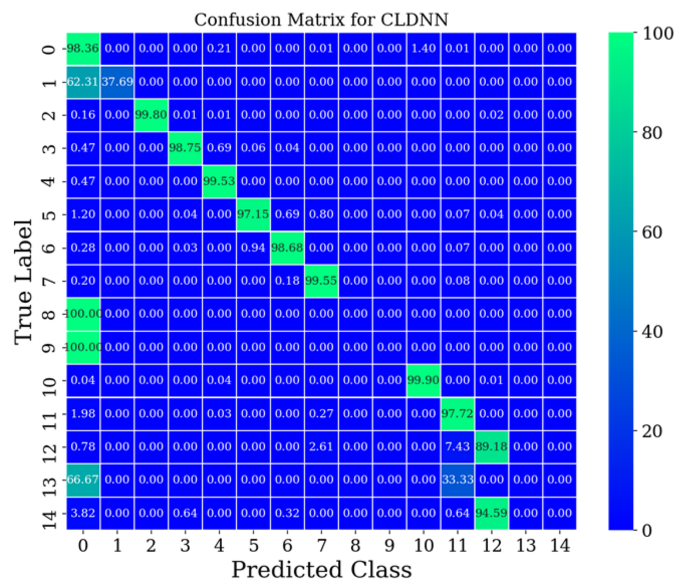


Fig. 1. Attack Accuracy for CLDNN Multi-class Classification Problem.

Fig. 1 is the confusion matrix of the performance of this CLDNN architecture. The labels on the Y-axis of the matrix indicate the true labels of the testing data. The labels on the X-axis of the matrix indicate the predicted labels of the testing data. A model with good performance observed high accuracy along the diagonal of the confusion matrix, which indicates that the model correctly classifies the attack with which it is trained with high accuracy. It has been observed that the proposed CLDNN architecture achieves an overall accuracy of 98.53%. For most classes, the model is able to classify the attacks correctly. However, it can be observed that although the model is being trained on data from Attacks 8 (HeartBleed), 9 (Infiltration), and 13 (Web Attacks: SQL Injection), it is not able to classify them with a good attack accuracy correctly. This is mainly because Attacks 8, 9 and 13 are significantly underrepresented in the dataset. It could also be that there are not a lot of dominant features that influence the learning of the model. In addition, it is also a possibility that the important features do not have a large variance in their distribution. This makes it hard for the model to learn patterns to differentiate between the data classes.

2) ***Result of the CLDNN Architecture on the Two-class Classification Problem:*** Fig. 2 is the confusion matrix for the two-class classification problem. In this confusion matrix, the labels on the Y-axis indicate the attack with which the model has been trained. The labels on the X-axis represent the class of attack that the model is tested on. A model trained on one attack is tested on all other attacks for transferability. More off, the diagonal correlations are preferred for a model that exhibits attack transferability. High classification accuracies on the diagonal represent the model's performance when it is trained and tested on the same attack. However, to observe transferability in learning, it will be best to observe high accuracies when the model is tested on an attack class that it has not been trained on. It can be observed that training this CLDNN model with one attack does not correlate to any other attack. Most of the correlations observed are on the diagonal of the confusion matrix, which means the model only performs well when tested on the training attack. It exhibits low transferability; for example, when the model is trained with Attack 12, it can identify attacks in class 14 with 91.87% accuracy. It does not exhibit other strong off the diagonal correlations. Therefore, it has been concluded from these results that the transferability of learning cannot be tested using this architecture.

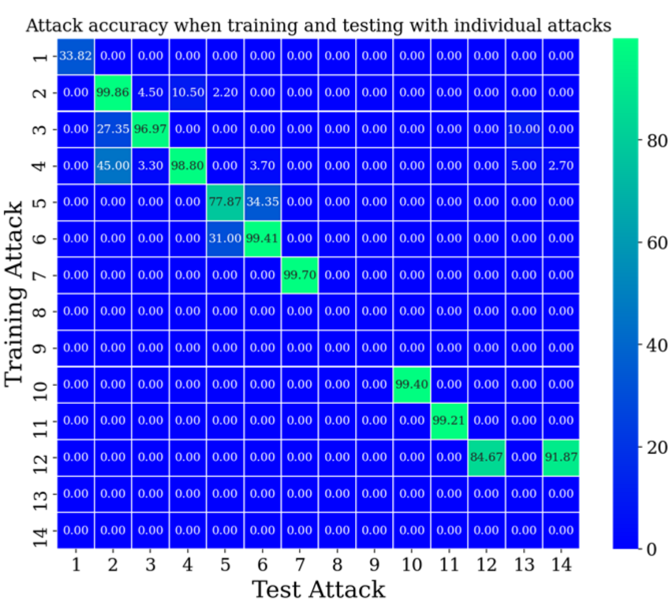


Fig. 2. Attack Accuracy for CLDNN for Two-class Classification Problem.

3) ***Results on CNN with BGRU Layers and CNN with Stacked BGRU Layers:*** These results present a comparative study of the performance of the CLDNN, CNN with BGRU layer and CNN with stacked BGRU layers for select attacks for the two-class classification problem. The performance of a CNN has also been evaluated with a BGRU layer with 10 cells. The attack accuracy was evaluated when tested on each attack. All the entries are attack accuracies in terms of percentages.

TABLE VII. Comparison of Performance of LSTM, BGRU and Stacked BGRU Architectures on All Testing Attacks when Trained with Attack 4(DoS:HULK).

Testing Attack	1	2	3	4	5	6	7	8	9	10	11	12	13	14
LSTM	0	45	3.3	98.8	0	3.7	0	0	0	0	0	0	0	2.7
BGRU (10)	0	46	0	98.33	0	0	0	0	0	0	0	0	0	2.6
BGRU (50)	0	43.7	47	98.79	9	2.1	0	0	0	0	0	0	0	3
Stacked BGRUs	0	47.7	31.3	99	3.2	0	0	0	0	0	0	0	0	3

TABLE VIII. Comparison of Performance of LSTM, BGRU and Stacked BGRU Architectures on All Testing Attacks when Trained with Attack 2(DDoS)

Testing Attack	1	2	3	4	5	6	7	8	9	10	11	12	13	14
LSTM	0	99.86	4.5	10.5	2.2	0	0	0	0	0	0	0	0	0
BGRU (10)	0	99.81	0	3.3	4	0	0	0	0	0	0	0	0	0
BGRU (50)	0	99.79	0	0	4.4	0	0	0	0	0	0	0	0	0
Stacked BGRU	0	99.84	0	0	4.1	0	0	0	0	0	0	0	0	0

TABLE IX. Comparison of Performance of LSTM, BGRU and Stacked BGRU Architectures on All Testing Attacks when Trained with Attack 3(DoS:Golden- Eye).

Testing Attack	1	2	3	4	5	6	7	8	9	10	11	12	13	14
LSTM	0	27.35	96.97	0	0	0	0	0	0	0	0	0	0	0
BGRU (10)	0	0	95.66	0	0	0	0	0	0	0	0	0	0	0
BGRU (50)	0	11	97	0	0	0	0	0	0	0	0	0	0	0
Stacked BGRU	0	30.7	96.5	0	0	0	0	0	0	0	0	0	0	0

TABLE X. Comparison of Performance of LSTM, BGRU and Stacked BGRU Architectures on All Testing Attacks when Trained with Attack 5(DoS:Slowhttptest).

Testing Attack	1	2	3	4	5	6	7	8	9	10	11	12	13	14
LSTM	0	0	0	0	77.87	34.35	0	0	0	0	0	0	0	0
BGR (10)	0	3.2	0	0	98.83	35.9	0	0	0	0	0	0	0	0
BGRU (50)	0	0	0	0	76	34.16	0	0	0	0	0	0	0	0
Stacked BGRU	0	0	0	0	90.56	34.7	0	0	0	0	0	0	0	0

From Tables VII, VIII, IX, X, and XI, the models yield high accuracies only when tested on the training attack. Small correlations can be studied, for example, when the models are trained with Attack 5 (DoS: Slowhttptest) and tested on Attack 6 (DoS: Slowloris) and vice versa. However, the prediction accuracies are still low (about 28%-35%), which is insufficient to

TABLE XI. Comparison of Performance of LSTM, BGRU and Stacked BGRU Architectures on All Testing Attacks when Trained with Attack 6(DoS:Slowloris)

Testing Attacks	1	2	3	4	5	6	7	8	9	10	11	12	13	14
LSTM	0	0	0	0	31	99.41	0	0	0	0	0	0	0	0
BGRU (10)	0	0	0	0	33	99.44	0	0	0	0	0	0	0	0
BGRU (50)	0	0	0	0	31.13	98.24	0	0	0	0	0	0	0	0
Stacked BGRU	0	0	0	0	28.91	98.24	31	0	0	0	0	0	0	0

conclude that those attacks have a strong correlation. It can also tell from Table 4.4 that training the model with Attack 4 (DoS: HULK) has the potential to scale for Attack 2(DDoS) and Attack 3 (DoS: Hulk), but this model only detects those attacks with 31%-47% accuracy. This comparative study concludes that the architectures explored so far do not exhibit very strong attack transferability. The next section will explore another deep learning architecture, a One Class Neural Network and evaluate its performance for anomaly detection.

C. Results of One Class

The implementations are the implementation by Chalapathy et al. [16][19]. The raw data is passed through an autoencoder to find important features. It is then passed through a one-layer NN, and the function it optimizes is given by equation 1. For the CICIDS dataset, the model AUROCs were computed for different seed values.

From Table XII, it can observe that the model performs as good as random guessing in 30% of the runs, marginally better than random guessing in 60% of the runs and worse than random guessing in 10% of the runs. On average, it achieves an AUROC of 0.55. Therefore, It concluded from the AUROC values that this model cannot detect anomalies with high accuracy and would not be suitable for the two-class classification task at hand. The performance of a Convolutional Deep Neural Network has been evaluated in the next section.

TABLE XII. AUROC Values for Different Seed Values

AUROC values for different seed values
0.6047022068674568
0.6040465887906412
0.5290125251557807
0.48844116825440265
0.549943844886464
0.6398034285974904
0.5154925402064912
0.561410034947294
0.6227060493594897
0.5054273774706824

D. Result of CNN Architecture on Multi-class Classification Problem

The overall accuracy of the CNN is 99.65%, whereas the CLDNN achieved an overall accuracy of 98.53%. In the confusion matrix, the diagonal elements represent the percentage of data in each class correctly classified. It has been observed from the confusion matrix that the model is able to classify Attack 8 (HeartBleed) correctly and Attack 9

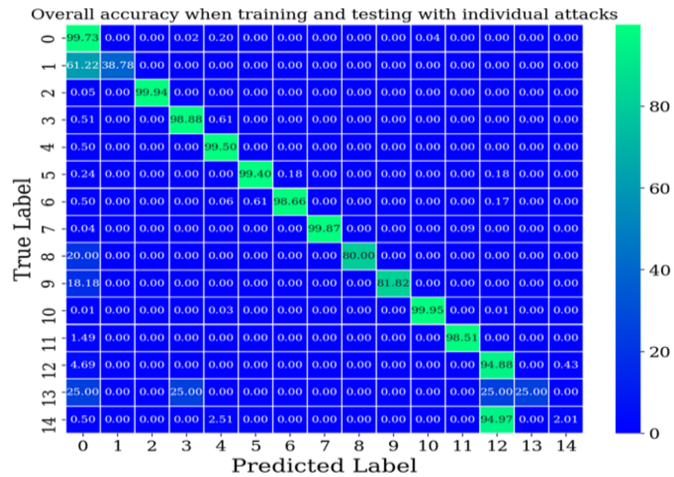


Fig. 3. Results of CNN Architecture on Multi-class Classification Problem.

(Infiltration), that have 11 and 36 data packets in the dataset up to 80% and 81.82% accuracy. Thus, this model can learn the important features from a high dimensional feature set. It concludes that this architecture has performed the best and is the best architecture for studying attack transferability. This architecture does not have recurrent layers, speeding up training time while increasing classification accuracy by 1.12%. This is our proposed base architecture.

E. Results of Attack Correlations

*1) **Results of Base Architecture Trained on Original Dataset**:* Fig. 4 represents the overall accuracy of the base architecture. Figure 5 is the confusion matrix for when the base architecture is tested for the two-class classification task. The diagonal elements are placeholders as they indicate the performance of the model when tested on the training attack. From Fig. 3, it has been observed that the model performs with high accuracies when trained and tested on the same attack. Observed Correlations have been shown in Table XIII. It observes transferability in the off-diagonal elements as:

1) Training with DoS:Golden Eye(Attack 3) scales for DDoS attacks(Attack 2), DoS:HULK(Attack 4), DoS:Slowhttptest(Attack 5) and DoS:Slowloris(Attack 6).
2) Training with DDoS Attack (Attack 2) scales for DoS:HULK(Attack 4) and vice versa.
3) Training with DoS:Slowhttptest(Attack 5) scales for DoS:GoldenEye(Attack 3) and DoS:Slowloris(Attack 6).
4) Training with DoS:Slowloris(Attack 6) scales for DoS:GoldenEye(Attack 3), DoS:HULK(At- tack 4) and DoS:Slowhttptest(Attack5).

TABLE XIII. Observed Correlations

Training Attack	Correlated Attacks
DDoS(2)	DoS:HULK(4)
DoS:GoldenEye (3)	DDoS(2),DoS:Hulk(4),DoS:Slowhttptest(5),DoS:Slowloris(6)
DoS:Hulk(4)	DDoS(2)
DoS:Slowhttptest(5)	DoS:GoldenEye(3),DoS:Slowloris(6)
DoS:Slowloris(6)	DoS:GoldenEye(3),DoS:HULK(4),DoS:Slowhttptest(5)

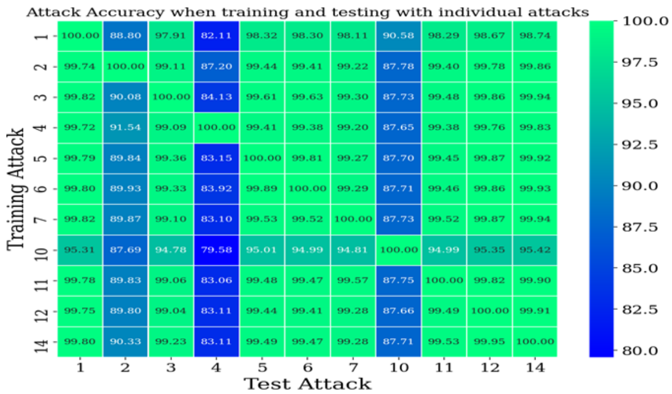


Fig. 4. Overall Accuracy for the Base Architecture for the Two-class Classification Problem.

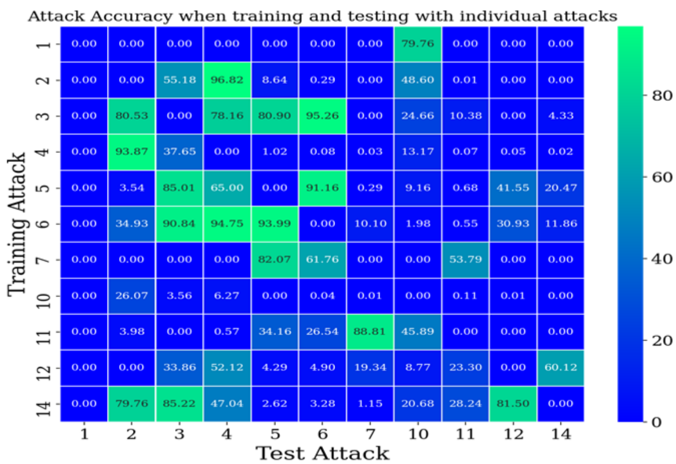


Fig. 5. Attack Accuracy for the Base Architecture for the Two-class Classification Problem.

*2) **Results of Base Architecture Trained on Synthetic and Bootstrapped Dataset**:* Fig. 6, Fig. 7 and Fig. 8 are the confusion matrices of the base architecture when trained on the SMOTE dataset and bootstrapped dataset and tested on SMOTE and real attack data. From Fig. 5, it can observe that most of the attack pairs that showed a correlation when the model was trained with the original training data also showed correlations when the model was being trained with synthetic and bootstrapped data. Like the previous section, the diagonal elements in the confusion matrices are placeholders as they represent the accuracies when training and testing on the same attack that already observed scales well with accuracies above 90% from Fig. 3.

*3) **Results of Attack Boosting Algorithm**:* As observed from Table XIV, the algorithm chooses Attack 11 (Brute Force: SSH Patator) as the selected training attack when tested on Attack 6 (Slowloris). This produces a very low PR. In Fig. 5, it can observe that training the model with Attack 3 (DoS: GoldenEye) alone yields a true positive accuracy of 95.26% and Attack 5 (DoS: Slowhttptest) alone yields a true positive accuracy of 91.16% when tested on Attack 6 (DoS: Slowloris). Table XV can observe a Positive Rate of 0.877 when the model was trained on Attack 11 data and tested on Attack 6. However,

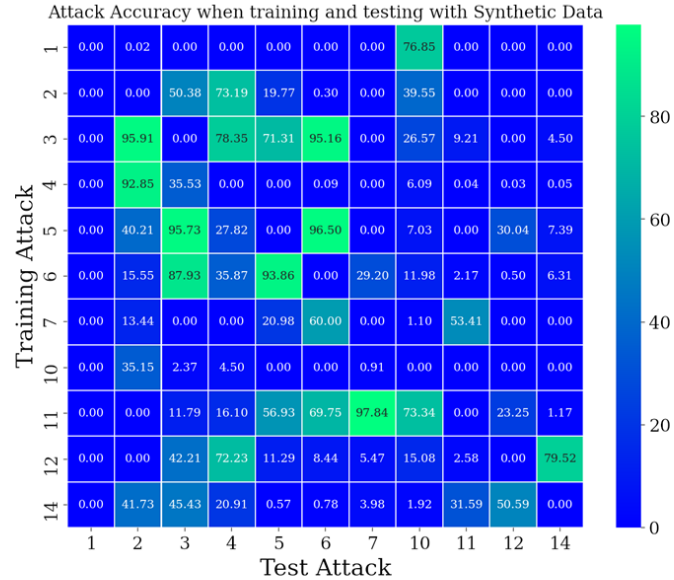


Fig. 6. Confusion Matrix for Base Architecture Trained on SMOTE Generated Attack Data and Tested on SMOTE Generated Benign Data and Real Attack Data.

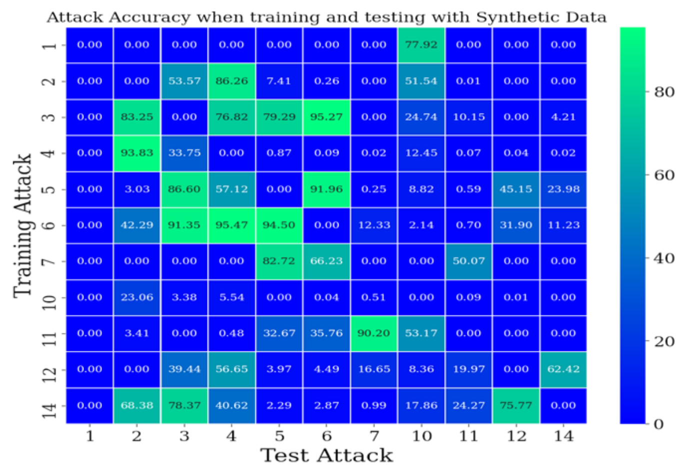


Fig. 7. Confusion Matrix for Base Architecture Trained on SMOTE Generated Attack Data and Tested on Real Benign Data and Real Attack Dataa.

this does not outperform the model's performance, which was trained with a single attack. From Table XIV can also be observed that the selected attacks for Attack 1 cannot correctly classify Attack 1. The selected attack for testing attack Attack 2 also achieves a low PR. From Figure 5, it can observe that training the model with attack data corresponding to only Attack 3 or Attack 4 yields true positive accuracies of 80.53% and 93.87% when tested on Attack 2. The selected attacks for Attack 3 cannot detect attack data packets; however, training the model with Attack 5 or Attack 6 yields a true positive accuracy of 85.01% and 90.84% when tested on attack 3 (as seen in Fig. 5).

Comparing the positive rates from Tables XIV and XV with Fig. 5, it concludes that training with one kind of attack scales better than training with a subset of attacks. Therefore, this algorithm has not been employed to further study attack

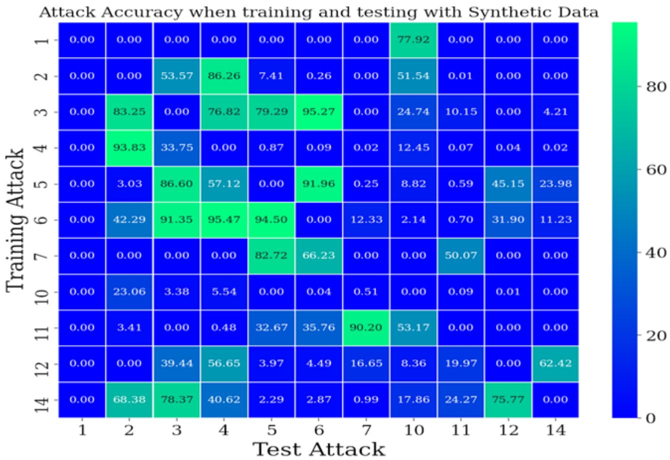


Fig. 8. Confusion Matrix for Base Architecture Trained on Bootstrapped Attack Data and Tested on Real Benign Data and Real Attack Data.

TABLE XIV. ATTACK BOOSTING ALGORITHM ON SELECTED ATTACKS

Test Attack	Selected Attacks	Accuracy	PR	NR	Weighted Acc.
1	10,7	92.75%	0.0	0.99	49.88%
2	10	85.3%	0.392	0.906	64.91%
3	10,7	91.94%	0.0	0.988	49.66%
6	11	99.38%	0.002	0.995	49.82%

TABLE XV. ATTACK BOOSTING ALGORITHM ON ATTACK 6(DOS:SLOWLORIS) WHEN TRAINED ON SMOTE GENERATED BENIGN AND ATTACK DATA

Selected Training Attacks	Accuracy	PR	NR	Weighted Acc.
11	99.54%	0.877	0.995	93.6%

correlations.

This study of attack correlations has concluded that the correlations observed in Fig. 5 are the only strong and consistent correlations that exist. The model exhibits the same correlations even when trained with a larger number of samples from the minority attack data classes using synthetic and bootstrapped datasets. Fig. 6, Fig. 7 and Fig. 8 indicate the same correlations as Fig. 5 when the model is trained with the augmented datasets. The next section discusses these observed correlations with the proposed hypothesis.

F. Results of Hypothesis for Attack Correlations

Table XVI shows the number of features selected for select attack pairs. For attack pairs like (3,2), (3,6) and (3,4) that are correlated, as seen in Fig. 5, Fig. 6, Fig. 7 and Fig. 8, the algorithm selects only 4, 6 and 6 features out of a total of 78 features in the dataset. For uncorrelated attacks like (3,7), the number of features selected by the algorithm is 24, which is considerably higher than correlated attacks. Similarly, for (3,1), which are also uncorrelated, 21 features are selected. This forms the base for the hypotheses. Some selected features from correlated attacks have also been observed in Table XVII. In order to strengthen the hypothesis, Some of the features listed in Table XVII have been used.

Hypothesis 1: Fig. 5 shows that training the base architecture with Dos: GoldenEye(Attack 3) attack causes it to

TABLE XVI. NUMBER OF FEATURES SELECTED FOR SELECT ATTACK PAIRS BY RFE ALGORITHM

Attack(s)	N Features Selected
(3,2)	6
(3,6)	9
(3,4)	6
(4,6)	11
(3,7)	24
(3,1)	21

TABLE XVII. FEATURES SELECTED BY RFE FOR SELECT TRAINING-TESTING ATTACK PAIRS

Select Attack Pairs	Selected Features
(3,2)	Destination Port,Total Length of Fwd Packets, Subflow Fwd Packets, Sub- flow Fwd Bytes, Subflow Bwd Packets, Init Win bytes forward
(3,4)	Destination Port, Bwd Packet Length Std, Flow IAT Mean, Packet Length Mean, Init Win bytes forward, Idle Min
(6,4)	Destination Port, Total Length of Bwd Packets, Bwd Packet Length Std, Flow IAT Mean, Flow IAT Std, Bwd IAT Mean, Bwd IAT Std, Packet Length Mean, Bwd Avg Bulk Rate, Subflow Fwd Packets, Init Win bytes forward, Init Win bytes backward, Active Mean, Active Std

perform well on Slow HTTP Attacks like Attack 5 and Attack 6 that, are Slow HTTP Attacks performed with Slowloris and Slowhttptest attack tools and vice versa. We hypothesize that this is because Attacks 3, 5, and 6 are very similar in the way they are performed.

Training with DoS: slowHTTPattack simulated with Slowhttptest scales well for Slowloris and vice versa as they are essentially the same attack but simulated using different tools. In Slow HTTP attacks, the attacker opens multiple HTTP connections with the target machine and keeps them open by sending packet headers or bodies slowly, thereby keeping all available connections occupied, making the target unavailable to legitimate requests. Since Attack 5 and Attack 6 are both the same kind of attack, it is possible for the model to learn attacking trends.

DoS: A goldenEye is an attack tool that explores vulnerabilities in the target; it tests how susceptible a target is to a DoS attack. DoS: GoldenEye opens multiple HTTP connections with the target server and keeps those connections open with keep-alive headers and caching control options. The connections cannot be closed, making the server unavailable to other HTTP requests. Slow HTTP attacks also open multiple HTTP connections with the target and keep open connections by sending large volumes of data at a slow rate, thereby making the target unavailable to other users. Whether it's for the keep-alive header or data being received slowly, the target is unable to close these connections. The same applies to Slowloris/Slowhttptest testing tools. This implies that although DoS: GoldenEye is not a Slow HTTP attack, it operates similar to Slow HTTP attacks. This could explain why training with DoS: GoldenEye Scales for both the Slow HTTP attacks.

Hypothesis 2: It has been observed from Figure 5 that training the model with attack data from DoS: HULK (Attack 4) causes it to scale for DDoS attacks (Attack 2) and vice versa. Training with Distributed Denial of Service attacks scales for DoS: HULK attack because the DDoS training attack was generated with Low Orbit Ion Cannon that simulated UDP, TCP or HTTP flooding attacks by sending enormous traffic volume to the target machine. DoS: HTTP Unbearable Load King is a flooding attack where the target is flooded with HTTP data packets, sending GET requests and hitting its resource pool. Both these attacks are flooding attacks that may cause them to scale for each other.

Hypothesis 3: Fig. 5 shows that training the model with DoS: GoldenEye (Attack 3) causes the model to scale for DDoS (Attack 2) and DoS: HULK (Attack 4); however, this observed correlation does not apply the other way around. Similarly, training the model with DoS: Slowloris (Attack 6) causes it to scale for DoS: HULK (Attack4) but not vice versa. The focus was on these asymmetric comparisons in this hypothesis. To explain this, it has been hypothesized that correlated attacks will have few dominant features selected by the algorithm, which may explain the correlations between the attacks. In contrast, the uncorrelated attacks will have many features selected by the RFE algorithm, indicating that there are not many dominant features that are important for learning which could explain the low correlations. This can be strengthened by the results in Table XVI for Attack pairs (3,2), (3,4), and (3,6) that have 6, 6 and 9 features selected by the algorithm. Each of these attack pairs is correlated. However, for uncorrelated attacks like (3,7) and (3,1), the RFE algorithm selects 24 and 21 features which are considerably more.

In addition to these hypotheses, some features observed from the selected feature list from Table XVII provide validation for our hypotheses. For example, in addition to DoS: GoldenEye and DDoS attacks having only a few features selected in their reduced feature set, the features selected as indicated in Table XVII are Subflow Fwd packets, Subflow Fwd bytes, Subflow Bwd packets. Subflow Fwd packets are the average number of packets in a sub-flow in the forward direction. Subflow Fwd bytes is the average number of bytes in a subflow in the forward direction, Subflow Bwd packet is the average number of bytes in a subflow in the backward direction. Subflows in the forward and backward direction are usually associated with distributed systems which point to these two attacks being distributed attacks. DoS: The GoldenEye attack tool can be used as a distributed attack as well by specifying the number of workers while simulating the attack. This provides further explanation for generalization.

Similarly, to study the correlation between DoS: GoldenEye, DoS: Slowhttptest, DoS: Slowloris and DoS: HULK, the RFE algorithm chooses features Inter-Arrival Time, Bytes Sent in the Forward Direction, Bytes sent in the Backward Directions in the Initial Window which is coherent with the hypotheses for symmetric comparisons.

V. Conclusion

The deep learning approach shows that it generalizes well for this dataset in comparison to explicit statistical modelling-based methods. After evaluating the performance of several candidate DNN architectures like CLDNN, CNN with BGRU, OCNN and CNN, a CNN architecture has been proposed with the highest accuracy in the multi-class classification problem (Fig. 3). Further, this architecture has been used to study attack transferability and observe its performance in Fig. 5. Training attacks have been identified, for which the model generalizes well on certain testing attacks. The problem of underrepresented attacks in the dataset has been addressed using synthetic and bootstrapped dataset based training methods. Fig. 6, Fig. 7 and Fig. 8 show that it can observe the same attack correlations as in Fig. 5 for the performance of the model trained on the SMOTE dataset and bootstrapped datasets. This confirms our identification of training-testing pairs of attacks that exhibit attack transferability.

Furthermore, this study proposes hypotheses for the observed correlations. A Reduced Feature Elimination Algorithm has been used to strengthen the hypotheses. The relationship between the number of features chosen and the training-testing attack pairs can be observed in Table XVI. Table XVII lists feature selected by the algorithm for selected attacks coherent with the proposed hypotheses.

This work validates the hypotheses with features from the RFE algorithm. Future work could validate these hypotheses using transferability in learning for different datasets, for example, DARPA 2009 IDD, KDDCup99, NSL-KDD, UNSW-NB15, and WSN- DS. Future work could also analyze specific or selected features trends in studying attack correlations. It is interesting to observe from Fig. 5 that it only observed correlations in attacks that are different types of DoS attacks. It did not observe correlations from other kinds of attacks, for example, Attack 1 (Botnets), Attack 7 (Brute Force: FTP-Patator), Attack 8 (Heartbleed), Attack 9 (Infiltration) and Attack 13 (Web Attack: SQL Injection). Future research can focus on identifying attributes that could help the model scale for these attacks and testing for transferability of learning for these attacks. Future work could also focus on attack detection when there is more than one attack being performed simultaneously.

Acknowledgment

This work was supported by the Research Center of Computer Science and Engineering, PES Institute of Technology & Management, Shivamogga. The research centre is recognized by Visvesvaraya Technological University (VTU). The authors are grateful to all of those with whom we have had the pleasure to work during this work. Each of my Research Committee members has provided me with extensive personal and professional guidance and taught me a great deal about scientific research and life in general.

References

[1] A. Lopez, A. Mohan, and S. Nair, "Network traffic behavioral analytics for detection of ddos attacks", SMU Data Science Review, vol. 2, no. 1, 2019.

[2] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani. Toward generating a new intrusion detection dataset and intrusion traffic characterization. ICISSp, 1:108-116, 2018.

[3] M. Alkasassbeh, G. Al-Naymat, A. Hassanat, and M. Almseidin, "Detecting distributed denial of service attacks using data mining techniques", International Journal of Ad- vanced Computer Science and Applications, vol. 7, Jan. 2016. DOI: 10.14569/IJACSA. 2016.070159.

[4] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, "Deep learning approach for intelligent intrusion detection system", IEEE Access, vol. 7, pp. 41 525–41 550, 2019. DOI: 10.1109/ACCESS.2019.2895334.

[5] I. Sharafaldin, A. H. Lashkari, Ghorbani, and A. A, "Toward generating a new intrusion detection dataset and intrusion traffic characterization.", in ICISSp, 2018, pp. 108–116.

[6] M. A. Ferrag, L. Maglaras, S. Moschoyiannis, and H. Janicke, "Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study", Journal of Information Security and Applications, vol. 50, Dec. 2019. DOI: 10.1016/j.jisa.2019. 102419.

[7] J. Zhao, S. Shetty, and J. W. Pan. Feature-based transfer learning for network security. In MILCOM 2017-2017 IEEE Military Communications Conference (MILCOM), pages 17-22. IEEE, 2017.

[8] Zhao, J., Shetty, S., Pan, J. et al. Transfer learning for detecting unknown network attacks. EURASIP J. on Info. Security 2019, 1 (2019). https://doi.org/10.1186/s13635-019-0084-4

[9] Z. Taghiyarrenani, A. Fanian, E. Mahdavi, A. Mirzaei, and H. Farsi. Transfer learning based intrusion detection. In 2018 8th International Conference on Computer and Knowledge Engineering (ICCKE), pages 92-97, 2018.

[10] P. Wu, H. Guo, and R. Buckland. A transfer learning approach for network intrusion detection. In 2019 IEEE 4th International Conference on Big Data Analytics (ICBDA), pages pages 281-285, 2019.

[11] A. Singla, E. Bertino, and D. Verma. Overcoming the lack of labeled data:Training intrusion detection models using transfer learning. In 2019 IEEE International Conference on Smart Computing (SMARTCOMP), pages 69-74. IEEE, 2019.

[12] M. Masum and H. Shahriar. Tl-nid: Deep neural network with transfer learning for network intrusion detection. In 2020 15th International Conference for Internet Technology and Secured Transactions (IC-ITST), pages 1-7. IEEE, 2020.

[13] K. Simonyan and A. Zisserman. Very deep convolutional networks for largescale image ecognition. arXiv preprint arXiv:1409.1556, page arXiv:1409.1556, Sept. 2014.

[14] H. Dhillon and A. Haque. Towards network traffic monitoring using deep transfer learning. In 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), pages 1089-1096.IEEE, 2020.

[15] R. Vinayakumar, K. Soman, and P. Poornachandran. Applying convolutional neural network for network intrusion detection. In 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pages 1222-1228. IEEE, 2017.

[16] R. Chalapathy, A. K. Menon, and S. Chawla, Anomaly detection using one-class neural networks, 2019. arXiv: 1802.06360 [cs.LG].

[17] Abhijit Das, Pramod and Sunitha B S, "Anomaly-based Network Intrusion Detection using Ensemble Machine Learning Approach" International Journal of Advanced Computer Science and Applications(IJACSA), 13(2), 2022. http://dx.doi.org/10.14569/IJACSA.2022.0130275

[18] Abhijit Das, Pramod and Sunitha B S, "An Efficient Feature Selection Approach for Intrusion Detection System using Decision Tree" International Journal of Advanced Computer Science and Applications(IJACSA), 13(2), 2022. http://dx.doi.org/10.14569/IJACSA.2022.0130276

[19] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani. Toward generating a new intrusion detection dataset and intrusion traffic characterization. ICISSp, 1:108-116, 2018.

Design and Development of an Efficient Network Intrusion Detection System using Ensemble Machine Learning Techniques for Wifi Environments

Abhijit Das
Research Scholoar, VTU-Belagavi, CSE, PESITM, Shivamogga
Asst. Prof., Dept. of CSE, BNMIT,
Bangalore, India

Pramod
Associate Professor, Dept. of ISE
PES Institute of Technology & Management
Affilated to VTU, Shivamogga, India

***Abstract*—Intrusion Detection Systems(IDS) are vital for computer networks as they protect against attacks that lead to privacy breaches and data leaks. Over the years, researchers have formulated IDS using machine learning (ML) and/or deep learning(DL) to detect network anomalies and identify attacks. Network Intrusion Detection Systems (NIDS) within corporate networks is a form of security that detects and generates an alarm for any cyberattacks. In both academia and industry, the concept of deploying a NIDS has been studied and adopted. The majority of NIDS research, on the other hand, has focused on detecting threats that emerge from outside of a wired connection. In addition, the NIDSs recognize Wi-Fi and wired networks alike. The Wi-Fi network's accessible connectivity distinguishes this from the wired network. A wired connection is highly resistant to many insider threats that could occur on a Wi-Fi router. A conventional view to developing NIDSs may miss malicious activities. This paper aims to design a multi-level NIDS for Wi-Fi predominant networks to identify both organizational Wi-Fi networks malicious activity and standard network malicious activity. Wi-Fi devices are common on campuses and businesses, and they are incorporated into the fixed wired network at the gateway. Wi-Fi networks are the primary target for this implementation; however, they are also designed to function in wired environments. For the Multi-Level NIDS, the proposed model used an ensemble learning method that pools the strengths of multiple weak learners into a single strong learner.**

Keywords*—Machine learning; ensemble learning; intrusion detection system; Wi-Fi security

I. Introduction

When discussing modern technology, the dominating subject of conversations is the Internet and its continual progress. More specifically, how easy it has become to obtain access to it. This is why the number of Internet users has spiked up. Moreover, Internet applications in various industries have increased, providing a wider range of services. As a result, a considerable amount of information is to the users' disposition. Nevertheless, this information is available as well to the skilled network adversaries.

The 2022 Global Digital suite has reported that Internet users have reached 4.6 billion, as shown in Figure 1 [1]. At the beginning of the COVID-19 pandemic, one of the most talked-about stories was how much the world relied on the internet, especially as countries went into lockdown. Although the fluctuations in restrictions on movement over the last two years, the most recent data show that people are actually spending more time than ever before using connected technology. 5.3 billion people will be online by 2023, according to a report cited by the source. Over the period of 2018 to 2023, compound annual growth is 6%. A growth rate of 7.7 % from 2018 is expected in 2019 when there will be an additional 300 million internet users. It is also revealed that that the contribution to this accelerated number of people accessing the Internet can be attributed to the facilities now at the disposition to a large number of customers in relation to more affordable devices and mobile data plans. From all these devices owned by two-thirds of inhabitants of the world, around half of them are considered smart, which only means they have access to content provided by the Internet.

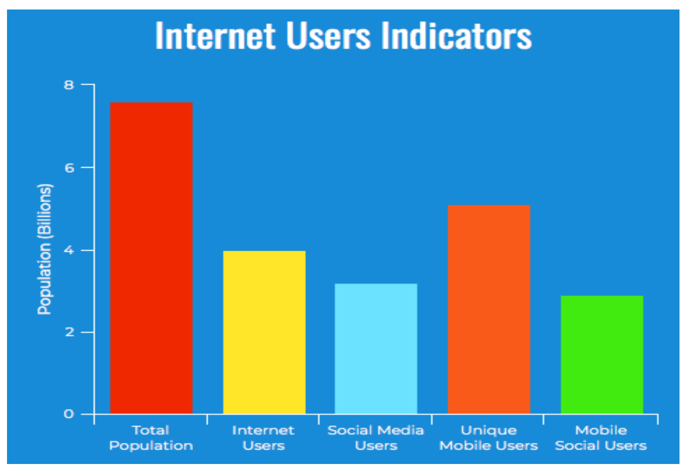


Fig. 1. Internet and Social Media Accessibility Statis [1].

The Internet Stack: The Internet protocol stack needs to be exposed to demonstrate the risks involved in blindly putting the trust in modern data transmission and management, which depicts the layers that describe the phases through which the data traverses between the communicating parties. Each layer fulfils its purpose and integrates with the others to provide the data transmission service. In Fig. 2, each layer could be explored as to its own world; nevertheless, this work aims to expose their main role in exchanging information, their characteristics, and the weaknesses that attackers commonly target at the highest level of the stack, application layer, which interacts directly with the users and programs. Right below appears the Transport layer that offers services to network applications. The network layer makes communication possible by facilitating Network addressing and routing. The Data Link layer is responsible for data transmission between devices belonging to the same Network. Finally, the physical layer refers to the actual hardware employed for the transmission. Knowing these layers' functionality helps us understand that network attacks performed in different layers should be studied in a different fashion. In this initial stage of development of the NIDS, the focus was to set, Network, and Data Link layers. Studying, thoroughly these layers help us become more familiar with scenarios in which cyber-attackers exploit weaknesses. More specifically, it has been determined what classes of attacks involving breaching in the mentioned layers will be part of the study.

Layer #	Layer Name	Protocol	Protocol Data Unit	Addressing
5	Application	HTTP, SMTP, etc...	Messages	n/a
4	Transport	TCP/UDP	Segments/ Datagrams	Port #s
3	Network or Internet	IP	Packets	IP Address
2	Data Link	Ethernet, Wi-Fi	Frames	MAC Address
1	Physical	10 Base T, 802.11	Bits	n/a

Fig. 2. A Depiction of the 5 Layers of the Internet Stack Described by [2].

Classes of Network Attacks: The objective of flooding attacks is usually to generate a denial of service effect on the victim. Flooding attacks overwhelm the communication service with a considerable amount of traffic, which most of the time results in the collapse of an already established communication between client and server. Another network attack taken into account is the impersonation attack. Here, the intruder's ultimate goal is to falsify the identity of a trusted entity to obtain sensitive data. The third and last attack to be considered in the analysis of network attacks in this research is an injection attack. This class of attack attempts to introduce maliciously input to a particular network, machine, or program. This action can result in several consequences ranging from denial of service to data theft [3]. Providing details regarding these attacks, their mechanism, and how they relate to the characteristics of the link layer in wireless networks is key in discovering the effective implementation of an IDS. Enlightening research work regarding Wi-Fi network attacks makes available a public dataset that was created with the special purpose of the analysis of IDSs. This dataset is called the AWID dataset [4]. The background research made as part of the generation of this dataset provided basic network attack information and specifications for types and classes of attacks, which are very popular among network attackers.

Machine Learning in Network Security: After an overview of how the network attacks were performed and the data was collected, it presented the approach taken to develop the IDS, which is ML. Nowadays, ML has become a very popular technology used in several fields, including information technology. Modern research has leaned toward this technology and has demonstrated that it is one of the most effective approaches to develop a model that can be trained to scan network traffic and detect if a network attack is being held [5]. Even though this technology appears to be emerging, research continues to evolve rapidly and makes available more efficient techniques. In recent years, it has become very popular to design learning models in an ensemble fashion as the combination of several ML models causes every model to help each other in correcting weaknesses. Therefore, the ensemble model will present an improved performance compared to the performance of each model considered alone. The outcome of including ensemble learning in network security is a great benefit since the ability to more accurately distinguish between normal and malicious traffic can prevent the compromise of sensitive data by its unauthorized use, misuse, or abuse [6] in any instance where the situation may happen.

It is critical as well to highlight that the technology of ML provides a precise way in which the effectiveness of a model can be measured. Most ML packages normally provide accuracy, which is considered in classifying if the network traffic is in a normal state or an attack is present, which becomes of great importance. Nevertheless, several other metrics can present deeper information about how effectively the model detects intruders. For example, logarithmic cost, F1 score, mean absolute error or mean squared error, among many others. These metrics are as important as accuracy and, in some cases, probably even more, especially when the samples of one class are imbalanced compared to the other class samples [7]. This is exactly the case of data describing a network attack. In a realistic network, it is very unlikely that an attack is being held, so if a network is monitored for a certain amount of time, the time frame of the network operating in a normal state is extremely larger than the time frame of the network while it was the victim of a network attack. Therefore, if an IDS classifies all network traffic as normal traffic, it will obtain an accuracy very high because the very few instances where it classified the network behaviour incorrectly were attacks. This result is clearly misleading and strongly suggests that in the case of IDS development, there needs to be included as many precision metrics as possible.

During last year, considering up to the 10 most disrupting events regarding cyberattacks, they compromised 5 million records with sensitive credit information in the mildest of these events up to breaching 1 billion records in the worst case [8]. It can be stated that organizational data breaches generate a greater concern about the safety of personal data. The study of a wireless network that simulates more the personal network setup does not restrict the findings to benefit these types of networks. Even though security levels may vary from a personal wireless network to an organizational one, the principles followed to compromise a network are the same. However, network communication is not just one type in an organization as they are traditionally in a personal network. Therefore, it is a more significant benefit to study the development of an IDS capable of scanning the behaviour of a network with a structure that combines wired and wireless networks. Even though the data, which was previously exposed, is comprehensive and provides a considerable amount of data to develop an effective model, it lacks the combination of different types of networks that, as mentioned before, are more common in the organization. An experimental setup that includes these requirements is necessary to simulate the performance of an IDS against network threats presented in this network architecture. The result of the combination of these two different types of networks is the motivation for the development of an IDS capable of scanning and analyzing not only wireless traffic but also Ethernet traffic. The inclusion of wired networks allows this research to include attacks in this type of network. Wireless packets provide very useful information in their headers pertaining to the link layer, which is one of the layers said to be focused on. Moreover, there is also a demonstration of how much information, probably more than it is needed, is present in packets at the link layer in the AWID dataset previously exposed. This dataset contains 155 different characteristics, called attributes, that can be displayed for each Wi-Fi frame.

So far, it has been referring to the unit of measure for network traffic as a frame for wireless connection. A different way of analyzing the traffic for the layers above the Link layer is employed. A unit flow is used, a condensed structure that describes network traffic characteristics in aggregation. Attacks against a wired network in the link layer are not as common as attacks to the higher layers, where several varieties of techniques are designed to create breaches. Therefore, an expected characteristic of an IDS that analyzes wired traffic is that it should detect only high-layer layer attacks. Also, it is necessary to take into account that wireless networks can be victims of the same type of attacks for the higher layers; nevertheless, wireless traffic is very well known to be the target of link-layer attacks as this network is open to all devices within range. For this reason, another component in the section of intrusion detection that should exist in the IDS is the capability of analyzing Wi-Fi frames and performing the respective data processing to detect either in wired network attacks or wireless attacks. Moreover, another component should focus its analysis on link-layer attacks and perform its analysis and attack detection on wireless frames.

Network attacks that are performed either in the wired or wireless environment are commonly studied from a perspective that is leaned towards the type of attack, whether it is flooding, impersonation, or injection type. Or more focused on the network's structure, whether it is a home network set up or enterprise network set up; or it can stress more the Internet layer that the intruder aims at, whether the network layer or link-layer as discussed in the preceding paragraph. All these perspectives are critical for a comprehensive study of cybersecurity, but this work proposes the consideration of all the mentioned factors from the standpoint of subjects that already are members of the network and exploit this privilege to trespass security barriers. This particular scenario is referred to as internal network attacks, and it is considered the most preferred scenario for network intruders to put into practice their malicious techniques. A study has demonstrated that 74% of security incidents resulted from the extended enterprise level. Of this 74%, 42% represent actual inadvertent employees. A total of 74% is reached considering, besides employees, customers, and suppliers, entities that are known to the company. The remaining 26% is attributed to parties unknown to the organization. Insider attacks are not a current issue. Another study refers to a survey made to United States security personnel, where insider incidents were cited 59% of the time [9].

II. Related Work

This section briefly reviews the literature for the current work, where researchers present the design and implementation of NIDS.

A. Wi-Fi Focused Research

Kolias et al. [4] extensively studied the attacks against Wi-Fi networks and categorized them. The contribution that proves to be the highlight of this research is the introduction of the Aegean Wi-Fi Intrusion Detection (AWID) dataset. In addition, this work also included the processing and analysis of the dataset. The dataset was fed into several ML algorithms. There are several characteristics that made this work stand out from similar research made in the field. The authors provided a wide overview of attacks being performed in the 802.11 standards in general. They constructed the dataset with well-supported assumptions, which is what makes this work recognized. This work also excelled at exposing a complete analysis of normal 802.11 network traffic containing normal network traffic behaviour and behaviour that describes a network attack. For the study of Wi-Fi IDSs, this dataset became a pillar, providing a format that is easy to distribute and very high-quality content as it is composed of real traces. The authors achieved the best accuracy using the J48 algorithm. They reported a 96.20% detection rate with all the features in the dataset and 96.26% with 20 features. Alotaibi et al. [10] attempted to improve the accuracy by applying the majority voting technique in which several ML algorithms were used with the AWID dataset, and then voting was performed on their results for the final prediction. Another singularity proposed in this work is the use of a technique based on the ensemble method of Extra Trees, which improves performance and is used for feature selection. The proposed solution was based on the combination of several machine-learning algorithms to learn patterns for different network behaviours. The initial procedure where patterns are constructed is called the offline stage. It was then followed by the online stage, where the classification of network attacks actually occurs. Before the intrusion detection, a filtering process was placed where all unnecessary features were disposed to leave only those that contribute to the classification. This work used the similar ensemble algorithms that have been used in this work, the combination of machine-learning algorithms, with a different approach and features. The participating algorithms used were Bagging, Random Forest(RF), and Extra Trees. Although these ML algorithms were used for the classification task, the Majority Voting, the voting technique employed, is what finally determined the classification output. The authors clearly specified that the voting technique was used particularly to improve the accuracy. As desired by the authors, it outperformed the result from Kolias et al. and reported an accuracy of 96.32%.

The literature review also found studies that make use of the DL approach. In [11], the author used DL for feature extraction, leaving the classification task to a Stacked-Auto encoder (SAE) classifier. The author mentioned that it is the first work considering this approach for IEEE 802.11 networks. The author presented the implementation of the neural network structure used for this classification problem. The neural network was composed of several layers. The first 3 layers were placed in the neural network to learn what features were the most useful to determine patterns. These patterns aided in classifying network traffic behaviour as an attack. The author used an emerging option to be employed as the activation function, which is the Rectified Linear Unit (ReLU) function as an alternative to the traditionally used Sigmoid function for the DL models. This process was performed, as stated, to provide a self-learning characteristic to this classification model only to determine or "learn" the most effective features to be considered for the actual classification. Later on, the author came to expose the type of classifier used for the actual anomaly detection, which is the Softmax Regression, a classifier capable of handling multi-class classification. An accuracy of 98.66% was obtained after the mentioned techniques were applied. However, information in relation to data preparation was very limited. More than half of the features in the AWID dataset contain missing values. It's also worth noting that there are a variety of features in the dataset, including hexadecimal characters and numerical values. Because of this, the dataset pre-processing is essential before using it in a model. The lack of attention to data preparation has been noted in the literature. On the other hand, this study shows that the data preparation for this work was meticulous.

B. Research Focused on Ensemble Models and Multi-level Intrusion Detection

Zaman et al. [12] proposed a more detailed perspective of network traffic analysis for intrusion detection. In this work, the authors focused on the different layers of the Internet stack. They proposed four different solutions that correspond to attacks analyzed from the perspective of the 4 upper layers in the Internet Stack. The IDSs are categorized as follows: Application Layer IDS, Transport Layer IDS, Network Layer IDS, and Link Layer IDS. The authors claimed that the results of this approach demonstrate improvement in system performance and scalability. For the task of feature selection, the authors employed Fuzzy Enhanced Support Vector Decision Function. As a result, the highest accuracy reported was 99.84% for the Transport layer IDS, which used Neural Networks for classification. Considering all layers, the best accuracy was 99.41% using a Neural Network classifier.

Zainal et al. [13] proposed an approach where classifiers with different learning paradigms are combined into a single Ensemble model. The paradigms employed are Linear Genetic Programming, Adaptive Neural Fuzzy Inference System, and RF. The authors' objective is to enhance the precision and lower the false alarm rate in IDS. Two principal steps were exposed in this work. First, select relevant features. Second, developing an ensemble model composed of classifiers with different learning paradigms. They demonstrated that the performance of the ensemble model surpassed the performance of the three models used separately for all types of attacks.

Li et al. [14] presented the use of Rough Set Theory and Quantum Genetic Algorithm for attribute reduction and as a method of classification. The author mentioned that attributes from network packets were reduced using the Quantum Genetic Algorithm. They used rough Set Theory to implement a rough meta-learning classification strategy. This strategy combined multiple rough learning methods. After the experiment, the author demonstrated that the detection rate was noticeably improved when Ensemble-Rough Classifiers were used versus the use of Single-Rough classifiers. Particularly, the detection rate increased from 76.86% to 86.25% for DoS attacks.

Wang et al. [15] propose the use of ensemble learning by using a Bayesian Network and Random Tree as base classifiers. These algorithms were combined with meta-learning algorithms using "Random Committee". Then, voting was performed for the classification task. In this work, the authors mentioned that the KDDcup99 dataset was used. One of the main objectives of this work was tackling the unbalanced nature of this dataset using ensemble learning. The model is evaluated using receiver operating characteristic (ROC) curves. The authors computed the area under the ROC curves (AUC) for more specific results. In the results, it was found that the ensemble model outperforms the single based models.

In [16], Nenekazi et al. uncover the issue of researchers not being able to determine the performance of an ensemble-based NIDS until after it is implemented. This work is based on the study of average information gain, which determines performance. This average information gain is associated with the features. Adaboost, the weak ensemble classifier, is used to obtain the average information gain. The NSL KDD dataset was used, and accuracy was the metric considered to measure performance in this work. The author demonstrates that average information gain lies in the range of 0.045651 and 0.25615 when accuracy will reach as much as 90%. A gradient boosted machine was used by Tama and Rhee [17] to increase the detection accuracy of anomaly-based IDS. Gradient boosted machine's best results are achieved by a grid search of input parameters. Rezvy et al. [18]used the AWID dataset that they used in [19]. They combined autoencoder frameworks with feed-forward neural networks. 99% of the time, their model can classify everything correctly. Faik Kerem Ors et al. [20] have developed a ML-based Wi-Fi IDS to protect IoT devices better. A single multi-class classifier operating on encrypted data from the Wireless Data Link Layer demonstrates that the benign traffic and six types of IoT attacks can be identified with an overall accuracy of 96%.

III. Methodology

The proposed installation of NIDS for the Wi-Fi network is shown in Fig. 3. The NIDS for the Wi-Fi network will receive Wi-Fi frames from the NIDS sensors for intrusion detection. As the NIDS is implemented using a ML approach, the model needs to train and validate it before deployment using the existing Wi-Fi frames dataset; it is the AWID dataset in this case. An overview of the various ensemble algorithms employed in the model development process is also offered here. Next, the algorithms will be applied to training data to identify the best one, and their evaluation of test data has been discussed in detail. Fig. 3 depicts the NIDS implementation process.

A. Datasets

The AWID Wi-Fi intrusion data was gathered in a typical Wi-Fi network environment. All of the devices connected to the Wi-Fi network were connected to a single access point for the Internet connection. A single intruder machine launched the attacks. Wi-Fi frames were captured using a computer in monitor mode. It was equipped with high processing capabilities to be able to capture a large number of frames at a

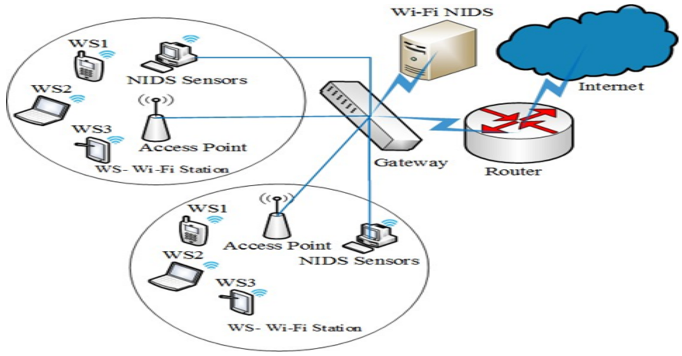


Fig. 3. An IDS Integrated with Infrastructure Wi-Fi Network.

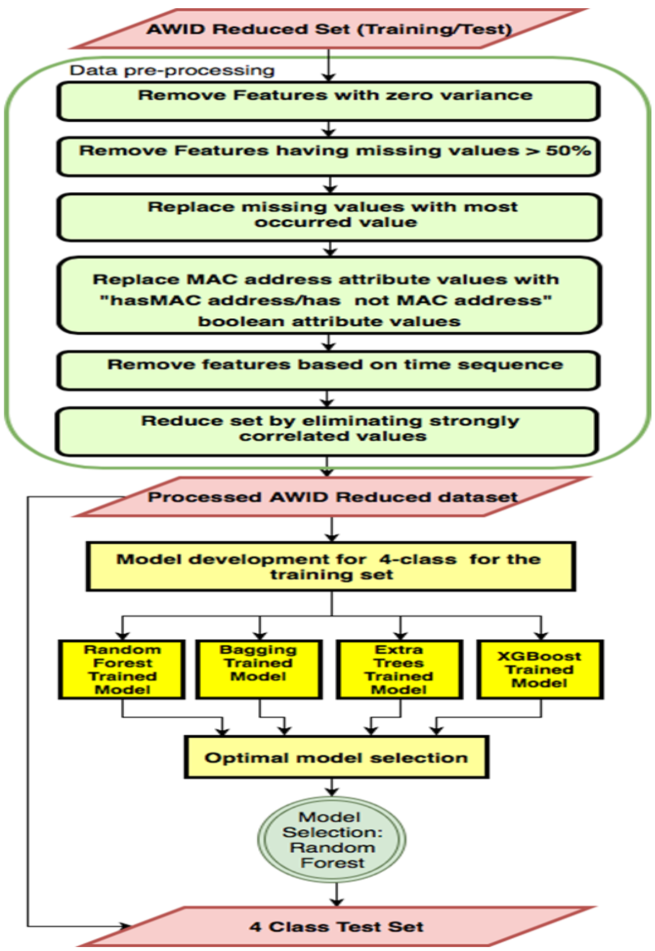


Fig. 4. Proposed IDS Model for Wif-Fi.

very high speed. All these equipment for the experimental set-up had a variety of hardware, operating systems, and other characteristics. Furthermore, to maintain the data capturing as realistic as possible, mobile devices, such as smartphones and tablets, were kept in constant motion, laptops experienced sporadic movement, and the desktop and the smart TV were kept in fixed locations during the experiment. The Wi-Fi frames were collected for around five days to release different datasets version, discussed below, for the intrusion detection. Both threat-class and threat-specific versions of the AWID dataset are available. Every record in the dataset is classified as either flooding, impersonation, injection attacks or normal. The identical information is labelled with threat specific class for 17 different Wi-Fi threats or benign. Many devices can use this available enormous or small version of this data format. In total, there are four AWID datasets available. The threat-class version of the dataset is used in this work. Table I shows the distribution of the records in the dataset. The AWID dataset has an imbalanced allocation, with more than 90% of the records in train and test sets being normal. As it can be seen, there is a significant difference in features and labels between large and small sets. The smaller dataset was collected separately from the larger dataset using various methods and techniques. As a result, there is no connection between them other than the sheer number of features and labels. Each record in the AWID dataset is a Wi-Fi frame with 155 features. These features represent various header fields values in each captured frame. They can define some traffic patterns helpful to detect intrusions. Nevertheless, some features may represent noise due to the raw state of the dataset. Control, management, and data frames are also included in Wi-Fi frames. Therefore, not every feature will apply to every frame. Therefore, some records may have missing values for those features.

TABLE I. Traffic Records Distribution for the Training and Test Sets in the Reduced AWID Dataset [10]

	Training %	Testing %
Normal	1633190 (91%)	530785 (92.2%)
Flooding	48483 (2.6%)	8096 (1.3%)
Injection	65377 (3.5%)	16684 (2.8%)
Impersonation	48521(2.6%)	20078 (3.1%)
Total	1795565(100%)	575643 (100%)

B. Preparation of Data

In the AWID dataset description, some values of the features in each record may represent noise for the intrusion detection task or maybe missing values. As a result, pre-processing the dataset is absolutely necessary. The dataset contains 154 features for each record and one target class (a total of 155 features). Features with zero variance, i.e., features whose value was the same across all records, have been eliminated. The AWID dataset description contained 27 features, some of which may represent noise or missing values for the intrusion detection task; because of this, preparing the dataset is a must. Each record in the dataset has 154 features and a single target class (a total of 155 features). A feature that had the same value in all data was eliminated because it had zero variance. Statistically, 27 characteristics have been found that had no statistical variation. Attributes with less than half of their values were removed. Following these steps, there are now 36 features in Table II. It was necessary to replace an attribute's incomplete data with its most common value. Access points and receivers have MAC addresses of 29, 30, 31, 32, and 33 for Wi-Fi adapters. They constantly adapt to the changing needs of the machine. The values of all of these features were changed to indicate whether or not the device had an address. Features 0, 1, 2, 3, and 13 were omitted because this deployment does not make use of time series. In order to achieve success, prediction is handled on an independent record basis. Considering that this feature was a combination of Features 20 and 21, it was omitted from the final product.

Feature 35 was also left out due to the fact that it indicates the sequence identifier.

TABLE II. AFTER PRE-PROCESSING FEATURES REMAINED IN THE AWID DATASETS

No.	Features name	Definitions
0	frame.time epoch	Time of epoch
1	frame.time delta	Time delta from previous captured frame
2	frame.time delta displayed	Time delta from previous displayed frame
3	frame.time relative	Time since reference or first frame
4	*frame.len*	*Frame length*
5	frame.cap len	Frame length stored into the capture file
6	*radiotap.length*	*Header length*
7	radiotap.present.tsft	TSFT
8	radiotap.present.flags	Flags
9	radiotap.present.channel	Channel
10	radiotap.present.dbm antsignal	dB Antenna Signal
11	radiotap.present.antenna	Antenna
12	radiotap.present.rxflags	RX flags
13	radiotap.mactime	MAC timestamp
14	*radiotap.datarate*	*Data rate (Mb/s)*
15	*radiotap.channel.freq*	*Channel frequency*
16	*radiotap.channel.type.cck*	*Complementary Code Keying*
17	radiotap.channel.type.ofdm	Orthogonal Frequency-Division Multiplexing
18	*radiotap.dbm antsignal*	*Antenna signal*
19	wlan.fc.type subtype	Type/Sub-type
20	*wlan.fc.type*	*Type*
21	*wlan.fc.subtype*	*Sub-type*
22	*wlan.fc.ds*	*DS status*
23	*wlan.fc.frag*	*More Fragments*
24	*wlan.fc.retry*	*Retry*
25	*wlan.fc.pwrmgt*	*Power Management*
26	*wlan.fc.moredata*	*More Data*
27	*wlan.fc.protected*	*Protected flag*
28	*wlan.duration*	*Duration*
29	*wlan.ra*	*Receiver address*
30	*wlan.da*	*Destination address*
31	wlan.seq	Sequence id
32	wlan.sa	Source address
33	wlan.bssid	BSS Id
34	*wlan.frag*	*Fragment number*
35	wlan.ta	Transmitter address

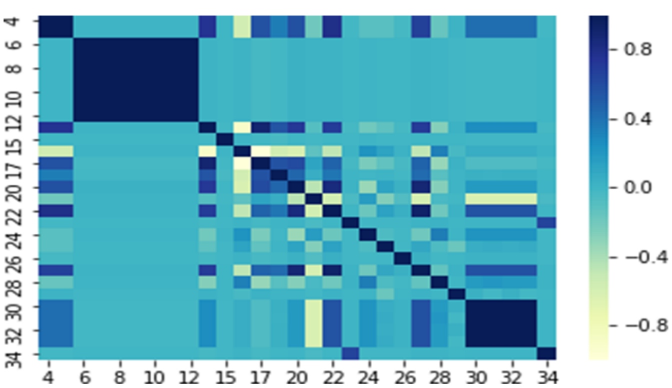


Fig. 5. Features Correlation Heat Map.

Finally, the correlation between the attributes has been calculated after following these steps.. It has been discovered that a few of the characteristics are strongly linked to each other. Fig. 5 depicts the feature correlation heat map. A strong positive correlation can be seen in the figure between the features numbered 4, 5, 6, 7, 8, 9, 10, 11, and 12, as well as 30, 31, 32, and 33, where the corresponding correlation value is 1. A strong negative correlation exists between features 16 and 17, with a correlation value of -1. One feature has been chosen from that group and discarded the rest for each group. There were four features that were chosen from the groups: Features 4, 6, 30, and 16. Thus the model came up with a new list of 18 features after going through all of these steps. Table II displays these features in italic type.

C. Ensemble Algorithms

An ensemble model is a technique that utilizes various algorithms or combines them to enhance the strength and efficiency of any of the components' algorithms. Syarif et al. (2012) [21] proposed that the benefit of ensemble methods is that they can be adjusted more appropriately than single models in any modifications in the controlled data stream.

1) Bagging: Unlike Boosting, the Bagging algorithm can be used for both classification and regression. In 1996, Leo Breiman developed the Bagging algorithm, which is also known as Bootstrap aggregating (Breiman, 1996) [22]. It was developed to improve the algorithm accuracy by producing diverse models and then combining them to produce the final model. Breiman's experiments indicated that Bagging could improve the efficiency of inconsistent learners but reduce the efficiency of a steady one when it was first established (Breiman, 1996) [22]. Changing the training data in the unstable learner can significantly impact the hypothesis it generates (Dietterichl, 2002) [23]. The experimental results reveal that Bagging works better than Boosting and Randomization in the presence of noise (Dietterich, 2000b) [24].

2) Random Forest: It is a decision tree-based variation of bagging. A decision tree's learning ability allows RF to extract precise information considering all aspects. This results in a precise performance. But it can also lead to excessive variation and over-fitting. Together with the bagging algorithm's principles, this strategy could fix the issue. Bagging in RF creates random subsets of data. RF validates that samples from one subset have less correlation than samples from other subsets. Each subgroup has a decision tree. A decision tree's training subset is used to classify data. Finally, the most common occurrence from each decision tree is picked as the overall classification output. The fact that RF randomly chooses a portion of features from each decision tree's feature set increases performance over bagging.

3) Extra Trees: The RF algorithm has a lot in common with this technique. Decision trees are at the heart of both approaches. When comparing it to RF, the main distinction is that it adds more randomness. Random features are scoured for each decision tree, and the locally best feature/split combination is computed. A random number is chosen for the split in ET, unlike RF. Over-fitting is less likely with this approach, which broadens the application of the algorithm.

4) XGBoost: The Gradient Boosting (GB) technique is effectively implemented in XGBoost. Multiple weak learners are sequentially used in the GB. Each learner focuses on those samples that the previous learner misclassified in the sequence during the training process. The GB consists in minimizing a cost function. The cost function describes the difference between an actual value and the approximation corresponding to the actual value. The minimization problem is tacked with derivatives, and the objective is to find the fastest descent in the difference between the actual value and the approximation. One of the main concerns in the use of

GB is time. Even though this algorithm is well-known for its ability to learn, the performance comes at the expense of considerable computational resources. These problems led to the creation of a set of tools called XGBoost that can help with the computational load of GB.

D. Implementation of Multi-Level NIDS

Fig. 6 shows a Wi-Fi dominant campus network prototype for the proposed ML-NIDS implementation with a connected Gateway switch. The Gateway switch has several Ethernet ports. Wi-Fi access points (AP) and other switches for local sub-networks can be connected using these ports. A monitoring node (installed with IDS in the figure) captures Wi-Fi frames for the traffic outgoing from or incoming to the wireless stations within a wireless network. The NIDS receives the packets for wired stations as well. Besides data, a network packet or frame consists of several headers corresponding to the different layers implemented in the network software stack of a wired or wireless station. These layers include application, transport, network, and data-link layers arranged in a top-down manner in the TCP/IP stack.

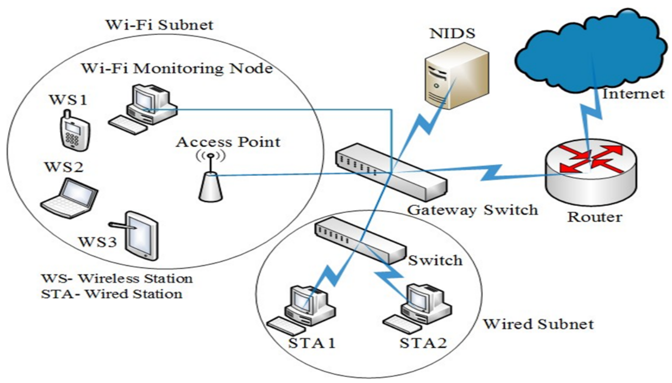


Fig. 6. Proposed IDS Model for Wif-Fi.

The NIDS processes data-link layer header fields in the case of a Wi-Fi frame and extract features to detect any internal attacks specific to Wi-Fi. The NIDS then processes header fields of network, transport, and application layers of Wi-Fi frame or wired packet and extracts features to detect generic network attacks. The functionality of the ML-NIDS is shown as a flow chart depicted in Fig. 7.

E. Obtaining Attack Traffic for ML-NIDS Implementation

A publicly available dataset has been selected, which contains captures of network traffic through an Ethernet connection for attacks and normal network behaviour. CTU University in the Czech Republic provides this dataset. Overall, it consists of 13 captures of normal traffic networks mixed with attacks. It is also available as separate captures, normal and attacks traffic behaviour. One particular CTU dataset has been selected that captured network traffic during a malware attack launched using an Ethernet connection. The model also used another dataset from [25]for normal traffic and DoS attacks. From the experiment mentioned, network traffic has been captured using tcpdump during the time when the attacks were launched. Since the length of the time captured was not too long, it has

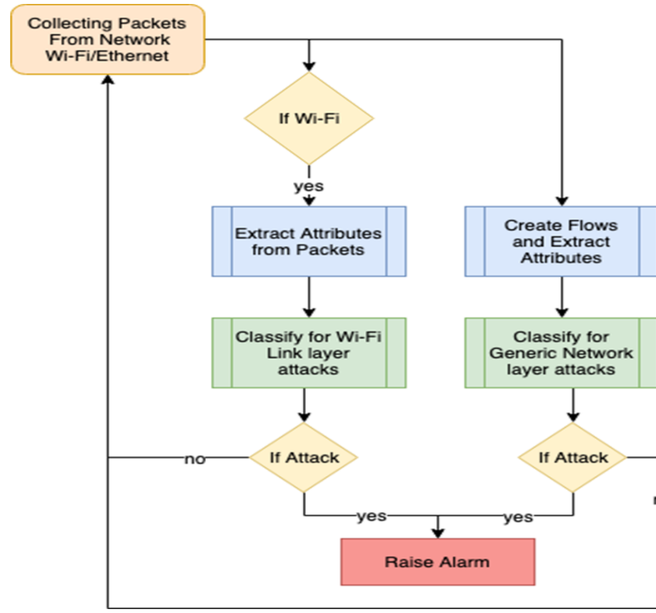


Fig. 7. Flowchart Demonstrating the Functioning of Multi-Level NIDS.

been decided that these captures were not useful for model training purposes. Therefore, these captures have been kept for testing. In this manner, it is possible to determine how the model would perform if an unrelated dataset is used for testing. Data Preparation and Features Extraction

In order to use the data for the ML-NIDS development, the raw data obtained is needed to prepare. All raw data manipulated in this study is in the PCAP format. To extract information for these files, "scapy" tool has been used. Scapy is a library that permits the manipulation of packets to extract their information in the python programming language. The manner in which information is extracted from the captured file to develop an intrusion detection model varies from the way proposed for a wireless network in this same study. This section will refer to the concept of flow to present a procedure that focuses more on finding meaningful information from the dataset used. A flow is a sequence of packets that share similar characteristics in their headers. Flows are commonly used to resolve performance issues. Because of the great amount of information they contain, using flows is the primary manner in which traffic network is represented for analysis. The features were derived by expressing statistical values describing characteristics such as direction (incoming/outgoing), size, flags, among many others. After that scapy was used to extract the features, it continued to process the data further to see if there was an opportunity to reduce the number of features. To determine what redundant features were redundant, their correlation and removed those highly correlated features. The heat map shown in Fig. 8 depicts the correlation status of all 68 features.

A threshold of 95% has been set, so from a group of features 95% correlated, only one feature has been kept. This process reduced the number of features from 68 to 37. A similar correlation heat map shown in Fig. 9 displays the correlation status of the 37 remaining features.

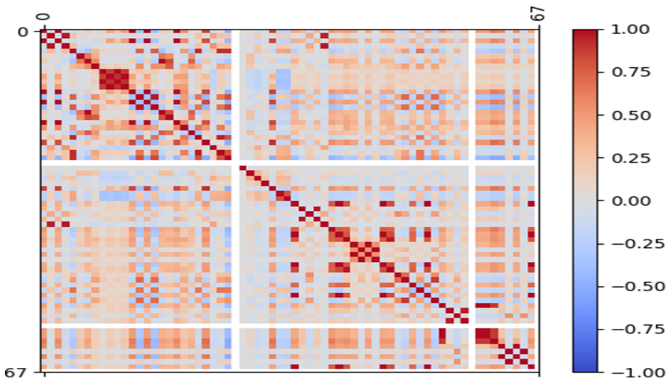


Fig. 8. Heat Map of Correlation between Original Features.

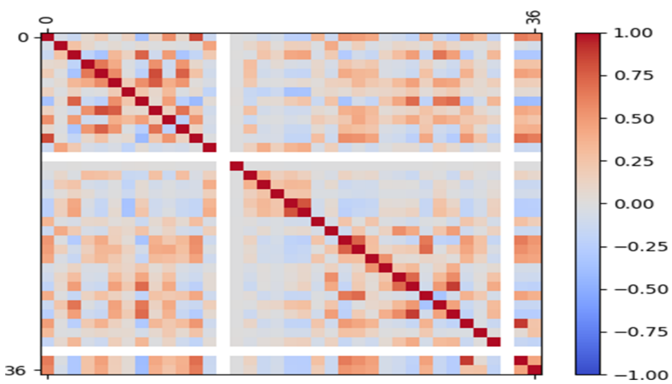


Fig. 9. Heat Map of Correlation between Features after Reduction of Features.

F. Ensemble Machine Learning Model Development

After the datasets were processed and converted into a format that can be used to train the NIDS model, the ML model was implemented. RF, Bagging, Extra Trees, and XGBoost were used as ensemble ML classification algorithms [27]. Training the models has been started with the dataset containing all 68 original features. Afterwards, the model was trained with the reduced 37 feature dataset. In the training process has been performed with 10 cross-validations and calculated the metrics: Accuracy, F1-score, Precision, and Recall.

It has been decided to include these metrics due to the nature of network behaviour. In a typical network, attacks are not something that happens very often. Therefore, it captures where a network attack was present; the number of packets involved in the attack will be considerably less than the number of packets involved in normal network activity.

For this reason, datasets containing network attacks are very unbalanced. In these cases, accuracy can be a deceiving metric to measure the performance of a NIDS. The other considered metrics can help us better understand the performance of a model trained and tested with very unbalanced datasets. This implementation corresponds to the second module of the Multi-Level NIDS. Therefore, the functionality of this system is organized in the flow chart shown in Fig. 7. The monitor node collects the data, and if the packets come from a Wi-Fi connection, they are first analyzed to determine if there is a Link Layer attack present. Regardless of the type of connection, Wi-Fi or Ethernet, all packets are analyzed to determine if an attack at the network layer or above is present.

IV. Result and Discussion

Models for the AWID training dataset were developed using the ensemble method depicted in Fig. 4. Models are built using sci-kit-learn, a Python-based ML tool. The models for each algorithm were developed using a 10-fold cross-validation process. The best model was chosen by averaging the accuracy of all of the algorithms, each using a distinct set of random states. Each algorithm's classification accuracy is shown in Table III. The most accurate model was the RF one. This NIDS model for the Wi-Fi network has been used as the last model for testing the dataset.

TABLE III. Accuracy of Ensemble Algorithms

Ensemble Technique	**Accu%**
Bagging	98.967
ExtraTrees	99.018
RF	99.097
XGBoost	98.944

This work evaluated the prediction results of different experiments using the test scores produced from the confusion matrix [26]. The confusion matrix is a two-dimensional matrix that represents the correlation of true conditions and predictive results shown in Table IV. TP describes the number of abnormal samples being accurately classified. TN defines the number of normal samples being accurately classified. FP specifies the number of normal samples being falsely classified as abnormal samples. FN specifies the number of abnormal samples being falsely classified as normal samples. Various test scores were calculated using the confusion matrix in this work: Accuracy, Precision, Recall, and F1 Score.

TABLE IV. Confusion Matrix

	Positive Prediction	Negative Prediction
Positive Condition	True Positive (TP)	False Negative (FN)
Negative Condition	False Positive (FP)	True Negative (TN)

$$Accuracy(Accu) = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1Score = \frac{2 * Recall * Precision}{Recall + Precision} \tag{4}$$

The RF model was 95.873% accurate in classifying the records. Fig. 10 depicts the model's confusion matrix. Each of the traffic classes has its own set of average values for precision, recall, and f1-score has shown in Fig. 11. Due to its f1-score of 0.13, the RF model's impersonation attack performance was poor. Fig. 10 shows that most impersonation threats are classified as injection threats. The model has an f1-score of 0.95 and intermediate accuracy, recall, and F1-score of

0.96. The majority of the impersonation assault records were classed as ”normal” even though the accuracy was slightly lower than that of [4]and [10]. In this regard, the proposed RF model outperforms theirs because it classifies the impersonation attack as a different type of attack. In addition, [4] reported a lower f1-score than the proposed model in this work. The model has an f1-score of 0.95 and intermediate accuracy, recall, and F1-score of 0.96. If flooding, impersonation, and injection threats are considered the sole attacks, it can map the confusion matrix of Fig. 10 to Fig. 12. When it comes to determining if the information is either malicious or benign, the model obtained 99.996% accuracy. To put it another way, in this context, [4]and [10] were both 96.28% and 96.28% accurate when it came to this question.

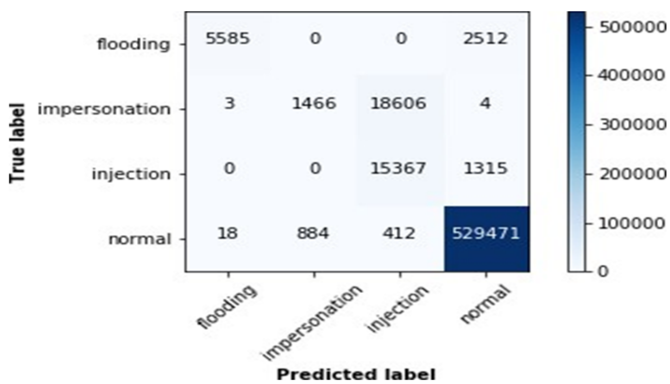


Fig. 10. Confusion Matrix for the 4-class Classification Model.

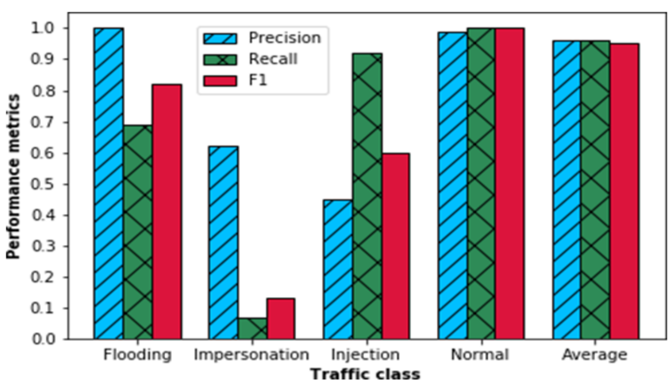


Fig. 11. Performance Metrics for Each Class.

The performance metrics were recorded after training the four mentioned ensemble models with the datasets before reducing the number of features. These are exposed in Table V. As shown, RF leads all 4 algorithms in all 4 metrics with an accuracy of 99.96%, F1-Score of 99.90%, the precision of 99.92, and recall of 99.89%.

TABLE V. Classification Performance on the Training Dataset

Algorithm	Accu	F1 Score	Precision	Recall
Random Forest	99.96	99.90	99.92	99.89
Bagging	99.94	99.87	99.89	99.87
Extra Trees	99.94	99.87	99.90	99.85
XGBoost	99.91	99.85	99.90	99.80

From the efficiency point of view, the training times RF is

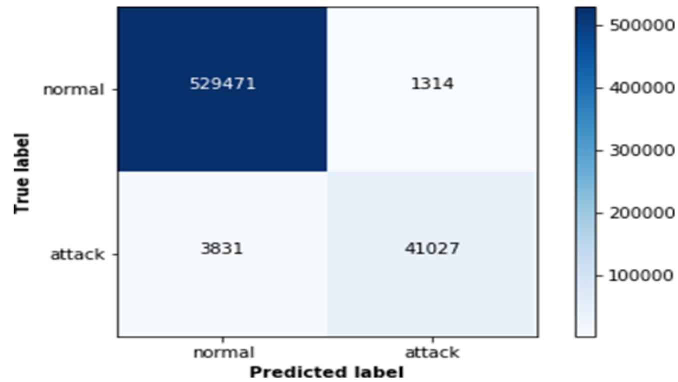


Fig. 12. 4-Class to 2-Class Confusion Matrix Conversion.

the second fastest model with 24.86 seconds after Extra Trees with a training time of 14.39 seconds. Bagging took longer to train with 59.12 seconds, and the second slowest was XGBoost with 31.83 seconds.

After decreasing the number of features, the same models were trained and the metrics recorded are shown in Table VI. Here it can be seen that Bagging and XGBoost performed very similarly. Both with an accuracy of 99.96%, F1 Score of 99.90%, and precision of 99.85%. Bagging lightly surpasses XGBoost in recall with a 99.85% versus 99.83%. The rest of the models score similar but slightly lower metrics.

TABLE VI. Classification Accuracy on the Training Dataset after Feature Reduction

Algorithm	Accuracy	F1 Score	Precision	Recall
Random Forest	99.95	99.89	99.98	99.85
Bagging	99.96	99.90	99.98	99.85
Extra Trees	99.95	99.89	99.97	99.83
XGBoost	99.96	99.90	99.98	99.83

In training time, the fastest model once again was Extra Trees with 13.03 seconds of training time, followed by XGBoost with 18.89 seconds. RF took 20.44 seconds for training, and the Bagging model trained in 36.33 seconds. The performance metrics improved after the feature reduction and model efficiency increased. The Bagging model improved significantly in performance and efficiency. Therefore, it has been decided to select it to perform the classification task for the NIDS. This model has been tested against three different test sets. The first one was generated with the training set but separated before the training process. The second one was obtained from CTU University. This set is not related in any form to the CTU dataset used to create the training set. The third is network attack capture obtained from our experiment performed in the isolated network previously introduced. The proposed model started with the test set separated from the training set before the training process. After running the bagging model for classification, it was found that for all four metrics, this model obtained a perfect performance score. This exceptional result could have been caused due to the fact that this test set was generated together with the set used to train the model. Even though they were separated before the training process, they still share very similar characteristics. The confusion matrix shown in Fig. 13 exposes the classification output of the model when tested with the mentioned set.

Then the model was tested with the additional malware capture obtained from CTU. After testing the model with this dataset, the results were as follows: 99.79% for accuracy, f1-score of 99.89%, perfect precision score, and 99.79% for recall. The entire dataset was a malware attack, so it can be observed that very few records were misclassified. In the confusion matrix in Fig. 14, the classification report has been presented for this set. In the same confusion matrix, it can also see that almost all records were classified as attacks. This outcome was expected since the entire test set was a capture of a malware attack. Since the miss-classified records were very few, it is hard to identify them in the confusion matrix. At last, the model was tested against a dataset generated during the experiment mentioned previously. The duration of the attacks launched was not very long and did not generate as many records as desired. Therefore, this network attack capture has been combined with normal traffic and balanced it to be composed of 10% network attack traffic and 90% normal network traffic.

The results reflected perfect performance for this small test set for all metrics considered. Performance with this test is displayed in the confusion matrix shown in Fig. 15

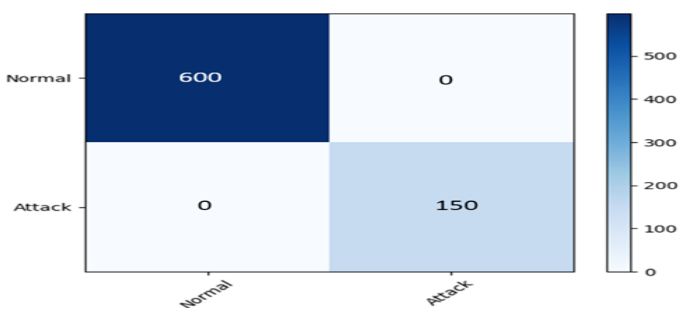


Fig. 13. Confusion Matrix for 1st Test Set.

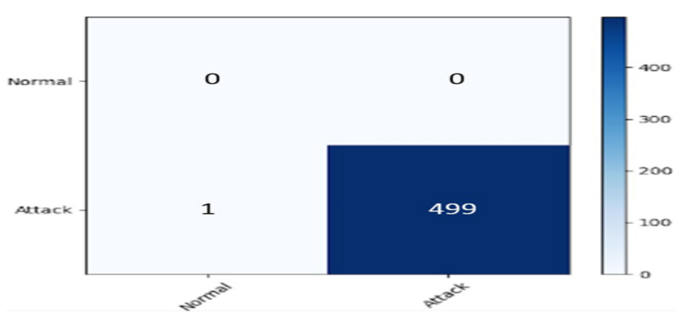


Fig. 14. Confusion Matrix for 2nd Test Set.

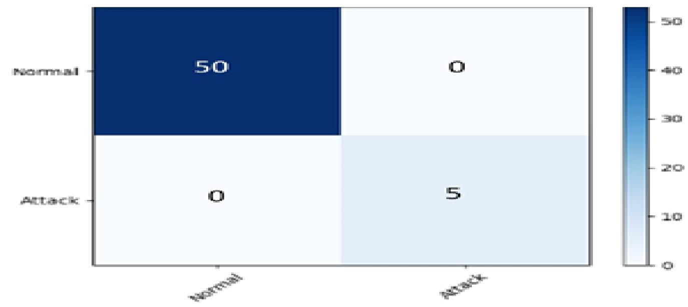


Fig. 15. Confusion Matrix for Test Set Obtained from Experiment.

V. CONCLUSION

A NIDS has been implemented to detect generic network attacks and Wi-Fi specific attacks. The most useful features selection techniques have been applied to select the features from the datasets to achieve precise and efficient performances. Ensemble ML models have been applied to classify network traffic as either normal traffic or malicious network traffic. The Ensemble ML models have been implemented separately to analyse Wi-Fi specific attacks and generic network attacks. RF for Link Layer Attacks and Bagging for Network Layer attacks were the best performing models. It is worth highlighting that the models performed well accuracy, performance measure, F1 score, precision, and recall. As mentioned, these additional metrics gave a more trustworthy judgment of the performance of the models against several imbalanced test sets of different nature. For Link Layer attacks, the accuracy obtained for two-class detection was 99.106%. A perfect score has been obtained for Network Layer attacks for two test sets. It performed well for the test separated from the training set before the training process with an accuracy of 99.79%, F1 score of 99.89%, perfect precision score, and 99.79% for recall. This work demonstrated that careful attribute selection improves efficiency and speeds up the training and classification process. Attribute selection also helps the model consider useful information that directly affects how records are classified. Thus, ignoring as much noise as possible, the model performs more precisely. The objective of this work is not just to operate a wireless network but to emulate an organization network environment where Ethernet connections are considered an important component. It is important to mention that this work planned to analyze generic network attacks for Ethernet connections. The packets must collect for two specific reasons; first, to gather information for the NIDS implementation. Second, for the existing system to capture the packets in real-time.

The proposed work has many branches through which it can be expanded and improved in future. The following are a few considerations for upgrading this work to enhance its functionality:

- One of the most significant challenges faced in this study was generating a larger original dataset for training. Even though it was possible to set up the desired infrastructure and launch several network attacks, the model was not able to collect massive network captures of a wider variety of attacks. Therefore, the model considers obtaining more penetration tools to launch a wider variety of network attacks as future work. These tools can help generate a more trustworthy dataset.
- Another consideration is to provide this ensemble-based multi-level intrusion detection service from the cloud. It has become popular in delivering cloud-based services that can be considered this feature as one of the key upgrades. This can be viewed as a key upgrade because computational resources are expensive, and for data processing tasks, hardware

requirements are high. By providing this cloud-based service, it can become more accessible for anyone without the hardware resources needed for efficiently processing network data.

- An additional tool that should be considered for future work is using Big Data tools for implementation. Network data is considerably large and could be viewed as the perfect scenario for using Big Data tools such as Apache Spark or Hadoop to provide a distributed approach for tasks such as reading, filtering, and so forth. These tools also offer Machine-Learning libraries that can be used to implement ensemble models.

REFERENCES

[1] Simon Kemp, "Dgital 2022: Global Overview Report," Jan 2022. [Online]. Available: https://datareportal.com/reports/a-decade-in-digital?utm_source=Global_Digital_Reports&utm_medium=Partner_Article&utm_campaign=Digital_2022

[2] ©2021 Microchip Technology, Inc. "Developer help." [Online]. Available: http://microchipdeveloper.com/tcpip:tcp-ip-five-layer-model

[3] I. Muscat, "What are injection attacks? - dzone security," Apr 2017. [Online]. Available: https://dzone.com/articles/what-are-injection-attacks

[4] C. Kolias, G. Kambourakis, A. Stavrou, and S. Gritzalis, "Intrusion detection in 802.11 networks: Empirical evaluation of threats and a public dataset," pp. 184–208, Firstquarter 2016.

[5] Abhijit Das, Pramod and Sunitha B S, "Anomaly-based Network Intrusion Detection using Ensemble Machine Learning Approach" International Journal of Advanced Computer Science and Applications(IJACSA), 13(2), 2022. http://dx.doi.org/10.14569/IJACSA.2022.0130275

[6] B. Mukherjee, L. T. Heberlein, and K. N. Levitt, "Network intrusion detection," pp. 26–41, May 1994.

[7] A. Mishra and A. Mishra, "Metrics to evaluate your machine learning algorithm," Feb 2018. [Online]. Available: https://towardsdatascience.com/metrics-to- evaluate-your-machine-learning-algorithm-f10ba6e38234

[8] D. Bisson, "The 10 biggest data breaches of 2018... so far." [Online]. Available: https://blog.barkly.com/biggest-data-breaches-2018-so-far

[9] M. B. Salem, S. Hershkop, and S. J. Stolfo, "A survey of insider attack detection research," Jan 1970. [Online]. Available: https://link.springer.com/chapter/10.1007/978-0-387-77322-3 5

[10] B. Alotaibi and K. Elleithy, "A majority voting technique for wireless intru- sion detection systems," in 2016 IEEE Long Island Systems, Applications and Technology Conference (LISAT), April 2016, pp. 1–6.

[11] V. L. L. Thing, "IEEE 802.11 Network Anomaly Detection and Attack Classi- fication: A Deep Learning Approach," in 2017 IEEE Wireless Communications and Networking Conference (WCNC), March 2017, pp. 1–6.

[12] S. Zaman and F. Karray, "Tcp/ip model and intrusion detection systems," in 2009 International Conference on Advanced Information Networking and Appli- cations Workshops, May 2009, pp. 90–96.

[13] A. Zainal, M. A. Maarof, S. M. Shamsuddin, and A. Abraham, "Ensemble of one-class classifiers for network intrusion detection system," 2008 The Fourth International Conference on Information Assurance and Security, 2008.

[14] L. Shen and L. Feng, "An efficient architecture for network intrusion detection based on ensemble rough classifiers," 2013 8th International Conference on Com- puter Science Education, 2013.

[15] Y. Wang, Y. Shen, and G. Zhang, "Research on intrusion detection model us- ing ensemble learning methods," 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), 2016.

[16] N. N. P. Mkuzangwe, F. Nelwamondo, N. N. P. Mkuzangwe, and F. Nelwamondo, "Ensemble of classifiers based network intrusion detection system performance bound," 2017 4th International Conference on Systems and Informatics (ICSAI), 2017.

[17] B.A. Tama, K.-H. Rhee, An in-depth experimental study of anomaly detection using gradient boosted machine, Neural Comput. Appl. 31 (4) (2019) 955–965.

[18] S. Rezvy, Y. Luo, M. Petridis, A. Lasebae, and T. Zebin. An efficient deep learning model for intrusion classification and prediction in 5g and iot networks. In 2019 53rd Annual Conference on Information Sciences and Systems (CISS), pages 1–6, March 2019

[19] C. Kolias, G. Kambourakis, A. Stavrou, and S. Gritzalis. Intrusion detection in 802.11 networks: Empirical evaluation of threats and a public dataset. IEEE Communications Surveys Tutorials, 18(1):184–208, Firstquarter 2016.

[20] F. K. Örs, M. Aydın, A. Boğatarkan and A. Levi, "Scalable Wi-Fi Intrusion Detection for IoT Systems," 2021 11th IFIP International Conference on New Technologies, Mobility and Security (NTMS), 2021, pp. 1-6, doi: 10.1109/NTMS49979.2021.9432662.

[21] I. Syarif, E. Zaluska, A. Prugel-Bennett, G. Wills, Application of bagging, boosting and stacking to intrusion detection, in: Machine Learning and Data Mining in Pattern Recognition, Springer, 2012, pp. 593–602.

[22] Breiman, L. (1996). Bagging predictors. Machine learning, 24(2), 123-140.

[23] Dietterichl, T. G. (2002). Ensemble learning. The handbook of brain theory and neural networks, 405-408.

[24] Dietterich, T. G. (2000b). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. Machine learning, 40(2), 139-157.

[25] Q. Niyaz, W. Sun, and A. Y. Javaid, "A deep learning based ddos detection system in software-defined networking (SDN)," CoRR, vol. abs/1611.07400, 2016. [Online]. Available: http://arxiv.org/abs/1611.07400

[26] Abhijit Das, Pramod and Sunitha B S, "An Efficient Feature Selection Approach for Intrusion Detection System using Decision Tree" International Journal of Advanced Computer Science and Applications(IJACSA), 13(2), 2022. http://dx.doi.org/10.14569/IJACSA.2022.0130276

[27] Abhijit Das, Pramod and Sunitha B S, "Anomaly-based Network Intrusion Detection using Ensemble Machine Learning Approach" International Journal of Advanced Computer Science and Applications(IJACSA), 13(2), 2022. http://dx.doi.org/10.14569/IJACSA.2022.0130275

Deep Reinforcement Learning based Ensemble Model for Intrusion Detection System

Abhijit Das
Research Scholoar, VTU-Belagavi, CSE, PESITM, Shivamogga
Asst. Prof., Dept. of CSE, BNMIT, Bangalore, India

Pramod
Associate Professor, Dept. of ISE
PES Institute of Technology & Management
Affiliated to VTU, Shivamogga, India

***Abstract*—Powered by advancements in information and Internet technologies, there has been a rapid development in network-based applications. Meanwhile, it is recognized that more attention needs to be paid to the issue of cybersecurity. The security of the network environment plays a vital role in the stable functioning of society. Cybersecurity research has become more active lately. Researchers have proposed several approaches to protect the network. Among them, a broadly practised approach is the intrusion detection system (IDS). This work suggested the potential value of reinforcement learning in building intrusion detection systems at the packet-level. A novel embedding approach has been proposed, namely image embedding, to encode the network traffics. Utilizing image encoding and raw network traffic, which are difficult to tackle by machine learning models, can be converted to images. Thus, the experiments applied convolutional neural networks. In addition, packets embedded in images are arranged in time order. In this way, it can integrate flow statistics with packet information and convert intrusion detection tasks to image-associated tasks. In this experiment, a Deep Q-Learning algorithm was selected for the ensemble with 1D-CNN and CNN and designed a training module and an interaction module. Experiments results indicate that the proposed RL-image-based approach can attain high performance on raw DDoS traffic provided by DDoS2019 and outperforms other traditional deep learning approaches.**

Keywords*—*Reinforcement Learning; Intrusion Detection System; Packet level IDS; Deep Neural Networks (DNNs); Deep Q Learning (DQN)

I. Introduction

Network is the substantial underpinning for developing a modernized economy. The progression of all professions and trades, including business, scientific research, entertainment as well as education, requires the assistance of a reliable and secure network system. However, numerous network security problems have also emerged, causing tremendous damage to individuals or the collective. For example, in April 2020, World Health Organization (WHO) announced that it had encountered an increased number of cyber-attacks, with about 450 email addresses and passwords of WHO and thousands of related staff exposed. If the network infrastructure is damaged and destroyed, not only will personal information be leaked, but the whole society will be thrown into chaos. Therefore, cybersecurity research is crucial for the successful and stable development of society.

Seeing the rapid development and the success of artificial intelligence (AI) attained in research areas such as computer vision, natural language processing and robotics, an increasing number of network engineers and researchers have started exploring the applications of artificial intelligence techniques in cybersecurity. As mentioned above, conventional intrusion detection systems rely on user-defined filtering rules under which the matched network traffic will be filtered. However, this filtering method has severe defects. Firstly, it is difficult to set up faultless filtering rules, even if experts specify the rules. In the face of tremendous network traffic and a huge variety of attacks, limited artificial rules cannot cover the characteristics of these different types of network traffic. Secondly, from hackers' perspective, rules are formulated by people, so hackers can easily figure out the pattern of those rules by certain tests. Hence, it is straightforward for them to launch attacks that can bypass those rules. Thirdly, the update of rules is inefficient. Because malicious network attacks are evolving quickly, intrusion detection systems also require frequent updates and upgrades to detect these attacks. However, it is time-consuming and laborious to determine what kind of outdated rules should be discarded and to lay down pivotal regulations.

Employing artificial intelligence approaches in intrusion detection systems can mitigate the problems mentioned earlier [1][2][3][4]. The filter rules are rendered automatically when using a machine learning-based model as the IDS to detect malicious network attacks. Machine learning models, such as deep neural networks (DNNs) [5], can efficiently learn representative features of different network attacks and then implement classification or detection; hence, manual rules are not required anymore. Because neural networks are a type of black box to everyone, hackers cannot acquire what kind of patterns and relationships are captured by the neural networks. Therefore, it is harder for hackers to launch novel attacks that bypass detection.

Reinforcement learning research [6] has been booming in recent years. Incorporating deep neural networks (DNNs), reinforcement learning algorithms outperform in many fields in comparison to traditional deep learning approaches, such as games, recommendation systems and many scientific fields. Some well-known reinforcement learning algorithms, including Deep Q-Learning (DQN) [7] and Proximal Policy Optimization (PPO) [8], have already been widely used in many areas. AlphaGo is one of the most representative achievements of RL in games. In addition, many enterprises, such as YouTube, Alibaba [9], etc. are exploring novel RL based recommendation techniques. Also, RL has found many great applications and research value in various engineering fields. Studies [10] use deep reinforcement learning algorithms to solve the low-level control problem of autonomous underwater vehicles (AUVs). Many researchers [11] apply reinforcement learning algorithms, especially PPO, to control unmanned aerial vehicles (UAVs). These successful applications demonstrate that reinforcement learning is at the stage of rapid growth. Motivated by the importance of the cybersecurity issue and the promising results of RL in many applications, a study has been conducted in this paper using reinforcement learning. Policy gradient methods, such as Asynchronous Advantage Actor-Critic (A3C), Trust Region Optimization (TRPO), and Proximal Policy Optimization (PPO), are widely applied in RL in devising game agents, including both discrete-action and continuous-action games. Among these methods, PPO stands out by achieving high scores in many testbeds, such as Atari games and MuJoCo [12]. The core innovation of PPO is the clipped surrogate objective. However, studies [13][14][15][16] indicate that the high achievements of PPO result from some additional code-level optimizations, instead of its key idea in a clipped surrogate objective. Thus, further understanding of these code-level optimizations is essential for applying PPO to tackle more problems, such as intrusion detection.

Reinforcement learning and intrusion detection are remarkably compatible in many aspects. Reinforcement learning algorithms are used for solving Markov problems [17]. Essentially, network flow is a special dynamic process type that a Markov process can model. Besides, when deciding on the current state, the RL agent can consider foresighted states by introducing discount factors when calculating the cumulative reward. This is also conducive to intrusion detection. The network flow consists of a sequence of different packets. When the current packet is examined, those packets behind it in the same flow can also provide useful information. Treat a packet as a state by using the cumulative reward with the discount factor, and reinforcement learning algorithms can be employed to take the features of future packets into account.

Furthermore, intrusion detection can be considered a special game. Hence, it is feasible to formulate and solve the intrusion detection problem in the reinforcement learning framework. Motivated by these pertinent aspects, this work attempt to study how to improve the performance of intrusion detection by using Deep Reinforcement Learning based Ensemble methods.

II. Background and Related Work

As an important method of machine learning, reinforcement learning (RL) is based on agents learning strategies to maximize returns or achieve specific goals while exploring and interacting with the environment. The basic framework of reinforcement learning is shown in Fig. 1. In general, a reinforcement learning framework consists of an agent and the interaction environment. The agent sends actions to the environment, and then the environment moves to a new state and generates rewards, which are sent back to the agent for the next step. The state represents the observation of the environment, and the rewards evaluate the outcome of the current action.

Similar to the controller in a control system, the agent is responsible for taking actions and interacting with the environment. An agent is controlled by its policy. As shown in Table I, the policy function takes the states/observations as the input and generates the action for the agent at the current step. In the actual implementation, deep neural networks (DNNs) are commonly employed to realize the policy function.

TABLE I. The Explanation of Policy

Input	Output	Transformation	Function
State	Action	Action = Function (State)	Deep Neural Networks

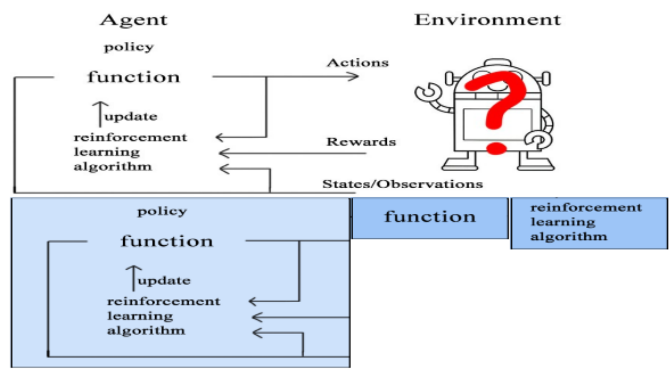


Fig. 1. Reinforcement Learning Framework.

In many cases, it is required to define the reward feedback rules manually. The rewards are used for updating the policy. After the policy is updated, it takes the current state as the input to issue new actions. In short, the policy can be considered a function that maps states to actions and is subject to changes according to rewards, as shown in Table I. The objective of RL is to learn the optimal policy for the agent to take the best action at each step.

RL Algorithms: In Fig. 2 some frequently used reinforcement learning algorithms are shown. A policy can be optimized by directly learning an actor or indirectly learning a critic instead. According to different learning targets, reinforcement learning algorithms can be divided into three main categories: policy-based algorithms, such as Proximal Policy Optimization (PPO) [8] and Trust Region Policy Optimization (TRPO); value-based algorithms, such as Q-Learning [18]; policy and value based algorithms, such as Actor-Critic (A2C) [19] and Asynchronous Advantage Actor-Critic (A3C) [20]. Policy-based algorithms train an actor who makes decisions directly while value-based algorithms train a critic who can evaluate current decisions. Policy and value-based algorithms utilize both the critic and the actor to train the agent.

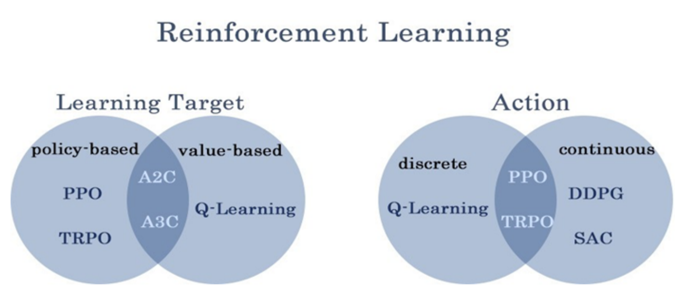


Fig. 2. Classification of Reinforcement Learning Algorithms.

Another important classification basis is the type of actions taken by the agent. In real implementation, actions can be either discrete or continuous. For example, the actions taken in Atari games are transferred to discrete values, but in MuJoCo games [12] actions are transferred to continuous values or vectors. In addition, Q-Learning is employed in the discrete domain. PPO and TRPO are suitable for both discrete domain and continuous domain. Deep Deterministic Policy Gradient (DDPG) [21] and Soft Actor-Critic (SAC) [22] are specifically devised for the continuous domain.

Q-Learning: Q-Learning [18] is a widely used model free value-based reinforcement learning algorithm, and it always works in the discrete domain. In this algorithm, the Q function Q(s, a) works as the critic, which is devised to evaluate the quality of the action taken in the current state. Its value, also called the Q-value, is updated by the equation (1):

$$Q^{new}(s_i,a_i) \leftarrow \underbrace{Q(s_i,a_i)}_{CurrentValue} + \underbrace{\alpha}_{LearningStep} \underbrace{(r_i + \gamma \underset{a}{max} Q(s_{i+1},a_{i+1}) - Q(s_i,a_i))}_{TD\ Learning} \tag{1}$$

The equation (1) demonstrates the idea of the temporal difference (TD) learning, which relies on the one-step boot-strapping technique [23] to update the values. Once up- dated, Q values are stored in a tabular form. The Q value is being updated until it stops changing or only changes slightly. After learning the Q values of all existing state and action pairs, the optimal policy can be generated under the following rule using equation (2):

$$a = argmaxQ(s,a) \tag{2}$$

Deep Q-Learning (DQN) [7] is an advanced version of Q-Learning, where deep neural networks (DNNs) are introduced to estimate Q values. Although Q-Learning is powerful and widely adopted, it can only estimate the Q values of existing state and action pairs that appeared during training. It is almost impossible for a problem with a large state space to store each value into a tabular form. Incorporating Q-Learning with DNNs can solve the problem. The deep neural network can estimate the Q values by using a state-action pair as the input. Even if the state-action pair has never appeared in the training stage, the deep neural network can still output an appropriate value. Compared with traditional tabular Q-Learning method, the generality of DQN is superior. DQN has several improved versions, such as Double DQN [24] and Dueling DQN [25]. DQN and its improved versions have been successfully applied in the game control. Researchers [7] have employed a variant of DQN to play seven Atari 2600 from the Arcade Learning Environment. The state is the game screen in the image form and thus convolutional neural networks (CNNs) are adopted to estimate the Q values. Experience replay is proposed to reduce the correlation among the training data. Results show that DQN outperforms almost all previous approaches when playing Atari games, including Sarsa [6] and human experts. However, in many cases, Q values tend to be overestimated by DQN. To handle this problem, Double DQN [24] is proposed. Studies [24] compare the DQN with the Double DQN on Atari games and find that the Double DQN can generate much more appropriate Q values while DQN always overestimates them, and this is the reason why Double DQN outperforms DQN in most Atari games.

Policy Gradient: Policy gradient method is a policy-based reinforcement learning technique that can be applied in both discrete and continuous domains. Policy gradient methods can directly optimize the parameterized policies $\pi\theta^2$ by maximizing the long-term cumulative reward $E_\pi[G_i]$ based on the gradient ascent approach. The policy gradient method has many advantages over other reinforcement learning algorithms. It can make the polices' convergence moregreedy over time autonomously. Furthermore, compared to Q-Learning and its variants, the policy gradient method can work effectively in continuous control problems. The policy gradient estimator can be expressed in the following form:

$$\hat{g} = \hat{E}\pi_\theta[\nabla \lg \pi_\theta(a_i|s_i)\hat{\psi}_i] \tag{3}$$

The equation (3) where π_θ is a parameterized policy with respect to parameter θ, $\hat{\psi}_i$ is a function related to rewards and $E\hat{\pi}_\theta$ is the mean over a finite batch of samples collected under π_θ. As shown in Table II, there are several expressions of ψ_i adopted by different policy based algorithms. Specially, the advantage function and the TD-error are widely applied. Compared to the average reward, the average function measures how much reward can be gained by taking the action a_i in the current state s_i, compared to the average reward. As shown in Table III, the TD-error calculates the difference between the estimated accumulative reward to be received starting at given time step i and the actual accumulative reward received starting at step i, and the algorithm acts to reduce such error.

TABLE II. MULTIPLE EXPRESSIONS OF $\hat{\psi}$

Expression	Explanation
$\sum_{i=0}^{\infty} r_i$	total reward after taking *ai*
$\sum_{i^i=i}^{\infty} r_i i$	add a baseline (constant)
$\sum_{i^i=i}^{\infty} r_i i - b$	state-action value function
$Q^\pi(s_i,a_i)$	advantage function
$Q^\pi(s_i,a_i) - V^\pi(s_{i+1}) - V^\pi_{s^i}$	TD error

TABLE III. TD-ERROR

Expression	Explanation
r_i	instant reward at step *i*
$V^\pi(s_{i+1})$	estimated accumulative reward starting at step *i* + 1
$r_i + V^\pi(s_{i+1})$	actual accumulative reward starting at step i
$V^\pi(s_i)$	estimated accumulative reward starting at step *i*

Several RL algorithms are developed from the policy gradient method, including the well-known Trust Region Policy Optimization (TRPO) and Proximal Policy Optimization (PPO), which have been successfully implemented in many areas. PPO has been extensively applied to game AI research and solving control problems. PPO is an off-policy reinforcement learning algorithm applicable to continuous and discrete environments. The core idea of PPO is the clipped surrogate objective. PPO not only attains high scores in test-beds like Atari and MuJoCo [12], but also solves complex problems such as DOTA [25][26] and Glory of Kings [27]. The well-known StarCraft agent AlphaStar [28] and Mahjong agent Suphx [29] are also trained with PPO. Additionally PPO has also be exploited to solve control problems, such as unmanned aerial vehicles (UAVs) control [11] and an autonomous underwater vehicles (AUVs) control [10].

Packet-level and Flow-level Intrusion Detection: Packet-level and flow-level analyses are two fundamental methods for designing intrusion detection systems. In the packet-level intrusion detection, messages and headers transported by network packets are primarily extracted for detecting malignant traffics [1, 2]. In flow-level intrusion detection, the traffic flow characteristics, which usually contains numerous packets, are extracted for detecting attacks [3, 4].

Packet-level intrusion detection systems analyze and extract features from the network packet data through transmissions for intrusion detection. As shown in Table IV, a network packet travels across five layers, including an application layer, a transportation layer, a network layer, an Ethernet layer and a physical layer (not considered in the paper). A packet structure consists of messages generated in the application layer and three headers generated in the remaining three layers. The content of messages generated in the application layer is different with respect to different protocols. For example, HTTP generates special HTTP request messages. Different headers carry different knowledge. In the transportation layer, TCP and UDP are two most common protocols with TCP and UDP headers as the most common ones in this layer. IP is the most common protocol in the network layer with the IP header. Similarly, in the Ethernet layer, an Ethernet header is generated. The knowledge carried inside headers is called field.

TABLE IV. NETWORK LAYERS AND PACKET STRUCTURE

Layer	Main Protocol	Packet Structure
Application Layer	FTP HTTP	Messages
Transportation Layer	TCP UDP	TCP/UDP Header+Messages
Network Layer	IP	IP Header+TCP/UDP Header +Messages
Ethernet Layer (Data Link Layer)	Ethernet	Ethernet Header + IP Header + TCP/UDP Header + Messages
Physical Layer	Not Considered	Not Considered

Packet-level and flow-level techniques approach the intrusion detection problem from two different directions. Packet-level research studies the data knowledge carried inside packets but ignores the relationship among different packets. On the contrary, flow-level research focuses on flow knowledge and attempts to capture the relationship among packets, but ignores data transported inside a packet. Both approaches have their advantages and disadvantages [30], so this study attempts to combine the packet-level approach with the flow-level one and build a more robust intrusion detection system.

ML-Based Intrusion Detection: In recent years, machine learning algorithms are broadly adopted in cybersecurity studies [1–4]. Some well-known machine learning algorithms, such as Random Forest (RF) [31], support vector machine (SVM) [32], and deep neural networks (DNNs), have shown promising results in the area of cybersecurity. In the work, the focus was on supervised learning-based approaches, such as deep neural networks. Table V shows the fundamental supervised learning-based intrusion detection framework. The entire intrusion detection procedure can be divided into two stages: training stage and detection stage.

In step 1, feature engineering is first conducted on the network traffics for the training stage. It is important to design appropriate features of traffics to improve the performances of intrusion detection systems. In flow-based research, certain standard flow-level features have already been given for most datasets available online, so the feature engineering step is not mandatory. However, for the packet-based research, it is often necessary to conduct feature engineering and extract features from raw network traffics. In this case, the quality of features is the key to the successful application of the intrusion detection system.

TABLE V. INTRUSION DETECTION FRAMEWORK

Step	Training	Detection
0	Collect Network for Training	Collect Network for Detection
1	Feature Engineering	Feature Extraction
2	Data Preprocessing	Data Preprocessing
3	Train AI Models	Monitored by IDS
4	Apply AI Model as IDS	Evaluate Detection Results

In step 2, data preprocessing is necessary for both flow-level and packet-level re- search. The features always contain three different value types: continuous value, categorical value, and discrete value. In order for the features to be processed by machine learning models, we need to transform the categorical and discrete values into certain suitable formats.

In step 3, an appropriate machine learning model need to be select. This step is of great importance to the final performance of the IDS. The most suitable model has been selected according to different rules with respect to different tasks. For example, if the input data is image, convolutional neural networks (CNNs) [33] are often the optimal choice. If it is time series data, the 1D-CNN [34] and the long-Short Term Memory (LSTM) network can be chosen. Afterward the training of the model and tuning of hyperparameters has been performed. Finally in step 4, the optimal model as the intrusion detection system has been applied.

For the detection stage, it was able to extract features engineered at the training stage from raw network traffics and conduct data preprocessing on these features. Subsequently, the pre-processing data was fed into the trained intrusion detection system. In order to evaluate the quality of the IDS, proper metrics for the machine learning model, such as the accuracy, precision, recall and F1 measures, etc. need to adopt.

TCP/IP Model and Packet Switches: This section introduces how a packet traverses from the source host to the destination host. Furthermore, important knowledge carried by the packets through the entire transmission trip has been introduced in details. Fig. 3 [35] shows a classical TCP/IP network model and the path of a packet traveling from the source host to the destination host. The TCP/IP model consists of an application layer, a transportation layer, a network layer, a data link layer (Ethernet layer) and a physical layer. The physical layer is not considered in this work. The packet transmission procedure [35] from a client side such as a browser has been explained.

First of all, the client launches an HTTP request in the application layer. The client creates a socket and sends the HTTP request message to the transportation layer through the socket. Both TCP and UDP protocols operate in the transportation layer, and HTTP can launch the TCP connection for transportation. After receiving HTTP messages from the application layer, the transportation layer creates a logic TCP connection with the transportation layer of the server through

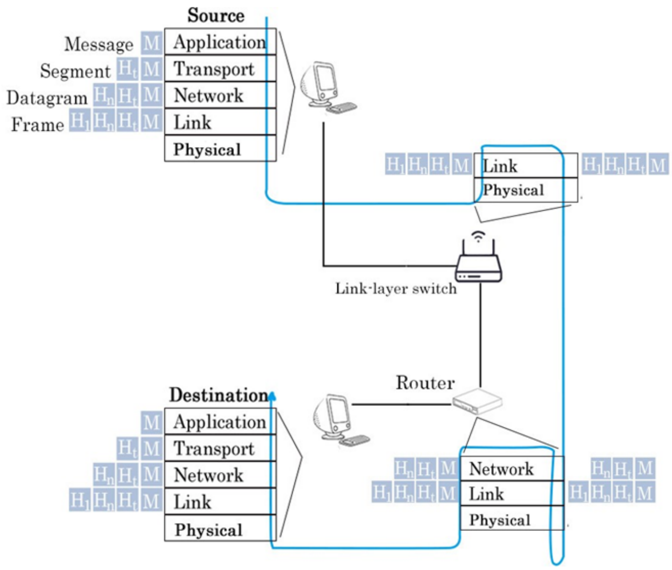


Fig. 3. TCP/IP Model and Packet Switches. M Represents the Application Messages Generated in the Application Layer. H_t, H_n, H_l Represent Transportation Layer Header, Network Layer Header and Data Link Layer Header, respectively.

three-way handshaking. Then, the transportation layer generates a TCP header (H_t), and adds it to the message, composing a TCP segment. The most important knowledge carried in the TCP header is the source port and destination port. The port identifies the application process, and different ports are bound with different application protocols. In this case, HTTP is bound with the port 80. Subsequently, the TCP segment is sent to the network layer for routing and forwarding. After receiving the TCP segment from the transportation layer, the network layer generates and adds an IP header (H_n) to the TCP segment, composing the IP datagram. The most important knowledge transported in the IP header is the source and destination IP addresses. IP address indicates the packet trip's start and end network address. Then, the IP datagram is sent to the data link layer. In addition to IP protocol, ICMP and ARP protocols also operate in the network layer. In most cases, Ethernet protocol is used in the data link layer, so the data link layer is also named the Ethernet layer. After receiving the IP datagram, the data link layer generates and adds an Ethernet header (H_l) to the IP segment, composing the Ethernet frame. Finally, the frame is sent to the server-side. Similarly, receiving a packet on the server-side is the reverse of sending a packet as described in the above. In general, this work has taken advantage of the quintuple, i.e. source IP address, destination IP address, source port, destination port, and the transportation protocol, to distinguish between multiple packet transmissions. In the following experiments, it has been attempted to adopt these properties to split network traffics.

Packet Headers and Fields: The knowledge transported by packet headers will be extracted for packet-level intrusion detection. The application layer generates application messages, and the remaining three layers generate special headers. A packet header consists of different type of fields to store the most representative knowledge of each layer. Table VI lists some common headers and fields. Taking the TCP header as an example, a TCP header consists of various fields, including source and destination port, sequence number, acknowledgment number, review window, header length, options and flag. The successful operation of TCP protocol requires the support of these filed knowledge [35]. Since fields store the most essential knowledge of headers, it is natural to adopt fields knowledge to establish a stable intrusion detection system. Fields are stored in the binary form in bytes, and in the following study, we use this property to transfer fields into pixels.

TABLE VI. HEADERS AND FIELDS

Header	Field
IP Header	Version Number, Header Length, Type of Service, Datagram Length, Identifier, Flags, Fragmentation Offset, Time-to-Live, Protocol, Header Checksum, Source and Destination IP, Options
TCP Header	Source and Destination Port, Sequence Number, Acknowledgement Number, Receive Window, Header Length, Options, Flag
UDP Header	Source and Destination Port, Length, Checksum
Ethernet Header	Source and Destination Address, Type, CRC, Preamble

Packet headers and application messages are recorded in 'pcap' (packet capture) files, which can be acquired by 'wireshark' [36], a commonly used network analysis tool. In the experiment, after selecting the network card of the network device on 'wireshark', it can automatically capture all the packets (Ethernet frames) traveling through the network card in chronological order and record these packets in 'pcap' files in a hexadecimal format. 'Pyshark' can be used to read packets from 'pcap' files. The DDoS2019 package, which contains raw 'pcap' files for the packet-level IDS research, has been adopted in this case study.

Convolutional Neural Networks: Convolutional neural networks (CNNs) [33] are feed-forward neural networks with a deep structure facilitating convolutional computations. It is one of the most representative algorithms of deep learning. CNN is widely applied in computer vision, such as image classification [37][38]. The name of CNN indicates that the network utilizes the convolution operation extensively. A basic CNN consists of convolutional layers, pooling layers and fully connected layers. CNN gains in popularity because of its affine invariance property, which is achieved by receptive field, shared weights and pooling.

A convolutional layer conducts feature extraction on images with convolution kernels, through which the receptive field and shared weights can be employed. The receptive field signifies that a neuron in the CNN is required to sense information of a part of the image, which can considerably reduce the number of parameters connected to a neuron. This is beneficial considering that the local correlation of an image is relatively stronger. By shared weights, different neurons can share the same convolution kernels, further reducing the number of parameters required for training. As shown in Fig. 4(b), a kernel (feature detector) performs the image's convolution operation to extract image features. Different kernels can extract different type of features. The output of the convolutional layer is called the feature map. Notably, one kernel generates one feature map.

A pooling layer can reduce the dimensionality of the data. As a general rule, the pooling layer is sandwiched between continuous convolutional layers, and it can compress the amount of data/parameters and prevent over-fitting. This work adopted two pooling methods, the maximum pooling and the average pooling, to perform down-sampling on each feature map.

The other commonly used CNN is 1D-CNN [34]. 1D-CNN is especially efficient when extracting essential features from shorter and fixed-length segments of the overall data set, and when the feature within the segment is not of high relevance. Accordingly, 1D-CNN is good for analyzing sensor time-series data [39]. There are two main differences between the 1D-CNN and the CNN. First, the input dimensions are different, and second, the convolution kernel traverses the input data in different ways, as shown in Fig. 4. As displayed in Fig. 4(a), assuming that the input is a sentence, each word can be transferred into a vector by word embedding approaches. The convolution kernel can be treated as the feature detector. The feature detector in 1D-CNN always covers the whole word, and scans from the start position to the final position. For example, in Fig. 4(a), the height of the detector is 2, and it moves from the top to the bottom with the step of 1. As shown in Fig. 4(b), the input is a batch of 2D images with respect to CNN. Using a square window, the convolution kernel slides horizontally and vertically across the image. As shown in Fig. 4(b), the feature detector operates in a 2x2 window.

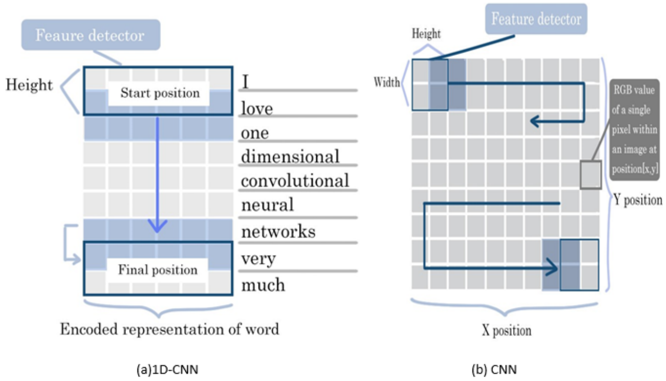


Fig. 4. Convolutional Operation Comparison. (a): 1D-CNN Convolutional Operation Example. (b): CNN Convolutional Operation Example.

Packet-level approaches, combined with machine learning algorithms and deep neural networks, have already been broadly applied in various areas of cybersecurity, including intrusion detection, traffic classification, etc.

In studies [1], an LSTM-based IDS is designed for detecting malicious traffics in raw network traffics at packet-level. The fields carried by packet headers are extracted and processed for experiments. They treat a field, such as IP version in the IP header, and TCP source port in the TCP header, as a word. A packet represents a sentence which consists of these fields (words). Researchers in this field have given a novel word embedding approach. They have produced a dictionary to map these words into integers. In this manner, one can transfer fields knowledge into 64-dimension vectors in the integer format. After implementing data extraction and word embedding, a 3-layer LSTM model with dropout layers is established for detection. Studies [2] also employ packet-level knowledge for detecting malicious traffics in IoT environment. Similarly, they conduct feature extraction on raw traffic flow. Field information is extracted from separated packets and each field represents a feature. The authors concentrate on header fields, including frame, IP and TCP/UDP associated information and exploit one-hot encoding to encode those categorical features. After implementing feature extraction and data preprocessing, they design a three- layer fully connected neural network with the sparse cross-entropy loss function for multi-class classification.

III. Methodology

A. Methods and Procedures

The novel intrusion detection framework proposed at packet-level is shown in Table VII. The framework consists of two major modules and several sub-modules. The preprocessing module is devised for data transformation and feature engineering. The reinforcement learning module is devised for training the intrusion detection system with RL approaches. This module further comprises a training module and an interaction module. An additional anomaly detection module is deployed to detect those attacks which are blind to the training module. Each module has been introduce in the following sections in details.

TABLE VII. Intrusion Detection Framework at Packet-level

Module	Sub-Module	Function
Data Preprocessing	None	Data Transformation and Feature Engineering
Reinforcement Learning	Interaction	Collect batch data and store in replay buffer[1]
	Training	Train the RL agent with batch data
Anomaly Detection	None	Detect attacks blind to the training module

B. The Data Preprocessing Module

Motivated by the extensively used embedding approaches such as word2vec, a novel embedding method has been proposed, namely the image embedding for the RL based IDS developed in this work. Several designated transformations are performed in this method to convert the network packets into images. Table VIII show the main steps of the data preprocessing module. In the following, these steps have been explained in details.

TABLE VIII. The Data Preprocessing Module at Packet-level

Step	Process (Input $\rightarrow$ Method $\rightarrow$ Output)
1	Raw Network Traffic $\rightarrow$ Session-based Rule $\rightarrow$ Separated Sessions
2	Separated Sessions $\rightarrow$ Image Embedding $\rightarrow$ Session Images
3	Session Images $\rightarrow$ Labeling with Log Files $\rightarrow$ Images and Label
4	Images and Label $\rightarrow$ Normalization $\rightarrow$ Applicable Dataset

In step 1, the vast network traffic recorded in the 'pcap' files has been split into separated and small traffic files according to specific rules, such as the session-based and flow-based partition rules. The difference between a session and a flow is shown in Table IX. The packet transmission directions in the same session can be opposite, but the directions must be the same in the same flow. The common rule of session-based and flow-based partition is that if two packets share the same 5-tuple knowledge (source IP, source port, destination IP, destination port, transportation protocol), these packets will be categorized in the same way session or flow. As explained previously, such 5-tuple knowledge determines the travel route of a packet, thus, this rule is used for flow or session categorization. A session or a flow always contains numerous packets. In this experiments, the session-based rule for partition has been used. After partition, several separated sessions stored in 'pcap' files can be obtained; each session contains various packets recorded in the order of capture time.

TABLE IX. Performances of Different Discount Values on Four Agents

Source	Flow I	Destination
IP 1	Packet 4, Packet 3, Packet 2, Packet 1	IP 2
Port 1	←(one direction)	Port 2
	TCP or UDP Connection	
Destination	**Flow II**	**Source**
IP 1	Packet 4, Packet 3, Packet 2, Packet 1	IP 2
Port 1	→ (one direction)	Port 2
	TCP or UDP Connection	
Source/Des	**Session I**	**Des./Source**
IP 1	Packet 4, Packet 3, Packet 2, Packet 1	IP 2
Port 1	⇔(bi-direction)	Port 2
	TCP or UDP Connection	

In step 2, image embedding is conducted on these separated sessions. The structure of a network packet can be found in Fig. 5(a). Generally, a network packet consists of an Ethernet header, a TCP or UDP header, an IP header and application messages. Only fields stored in packet headers has been extracted in this work.

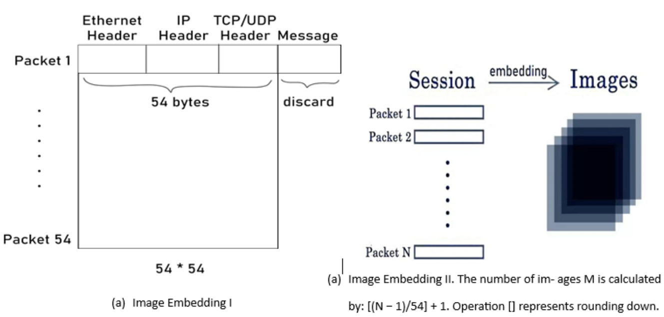


Fig. 5. Procedure of Image Embedding. (a): Transfer Packets to an Image. (b): Transfer a Session to a Batch of Images.

Application messages have been discarded because the length of a packet header is fixed, while the length of the application messages is not. The fixed embedding length is conducive to image embedding since the input image's size can be fixed. The whole procedure of image embedding is shown in Fig. 5. The fields transported by the packets are stored in bytes, so they can be converted to base 10 form with a range of 0 to 255, and one byte represents one pixel. Hence one packet represents a line of an image. With simple calculations, it has been found that the total length of a packet header is 54 bytes. Thus, the standard image format with a fixed size of 54×54 has been chosen, meaning that an image consists of only 54 packets. Some sessions may contain more than 54 packets, in this case, more images can be used to embed the extra packets. If the remaining packets are less than 54, the missing parts have been filled in with zeros. Packets are embedded in the order in which they are captured. This way, session knowledge into the packet-level experiments can be attached.

In step 3, after conducting image embedding, each session has been labeled by matching the time stamp given in the log file provided by the raw dataset. In step 4, the normalization has been performed on these images by dividing 255. All pixels of images are then normalized into [0, 1] from [0, 255].

C. The Reinforcement Learning Module

Before developing a reinforcement learning module, it is necessary to first trans fer the intrusion detection problem into an RL-based problem. For this purpose, the intrusion detection problem has been compared to the Atari games studied in [12], considering intrusion detection as a special game. The comparison is shown in Table X.

TABLE X. Comparison Between Atari Game and the Intrusion Detection Game

RL	Atari	IDS	IDS Space
State	Image(s)	Image	Images in dataset
Action	Game Operation	Prediction	Label Space {0, 1, 2...}
Reward	Game Feedback	Reward Mechanism	{+1, -1}
Episode	Game Round	Session	-
Agent	Game User	Classifier	-

Sessions identified in an intrusion detection game correspond to game rounds in an Atari Game. When the agent is playing the game, it takes the following trajectory:

$$s_0, a_1, r_1, s_1, a_2, r_2 ... s_t$$

where s represents the game screen, a represents the game operation and r represents the game reward. The game agent repeats until it reaches the terminated state st (game over) of an episode. Similarly, the intrusion detection agent can also take the above trajectory until it reaches the session's end (last image). The state is the game screen (in image format) in the Atari games, and in intrusion detection, the state is also an image after conducting image embedding on the sessions. The action space of Atari games contains game operations, which can be expressed as discrete numbers, e.g. 0 (one step left), 1 (one step right), 2 (one step up), etc. The action space of the intrusion detection game contains the types of the traffic class, which can also be expressed as discrete numbers like 0 (normal), 1 (attack 1), 2 (attack 2), etc. However, the reward system is different for the above two problems. Game developers have devised the reward mechanism of Atari games. For the intrusion detection problem, the reward system has to design carefully. After conducting the research it has found that the reward scale has a significant impact on the performance of reinforcement learning algorithms. Thus, inspired by reward clip for the code-level optimization implemented in the Atari game agent, the following reward feedback rule has been designed: if the prediction made by the agent is correct, the reward is 1; otherwise, the reward is -1. PPO is the main algorithm used for solving the discrete problem. Deep Q-Learning algorithm has been applied, rather than PPO. Deep Q- Learning is also an excellent choice for solving discrete problems, and it is much easier to implement compared to PPO. A comparison case study has been shown that the Deep Q-Learning outperforms policy gradient methods in the intrusion detection task.

After identifying these essential reinforcement learning ingredients, designing a reinforcement learning module has been started. Convolutional neural networks (CNNs) are chosen as the network structure of the intrusion detection agent since the input states are images. It should be noted that, contrasting to Atari games and other computer vision tasks, 1D-CNN is also appropriate in this intrusion detection system. One reason is that session images have been treated as a type of time-series data since packets embedded in the image are arranged in chronological order according to the capture time. Most importantly, as shown in Table XI, the intrusion detection problem is similar to the language model, where 1D-CNN is widely applied. Generally, sentences consist of different words and have different lengths. Similarly, a packet can be treated as a word, and a session can be treated as a sentence. When dealing with language tasks, advantage of word2vec approaches has been taken to conduct data preprocessing, while this IDS used the proposed image embedding to conduct data preprocessing for packets. Due to the above, this work is motivated to adopt the 1D-CNN in this experiment for the proposed IDS. The main structure of the two RL agents, i.e. DQN-CNN and DQN-1D-CNN, is shown in Fig. 6. The RL agents' input

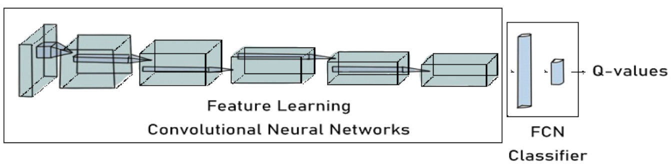


Fig. 6. Main Structure of Two Reinforcement Learning Agents DQN-CNN and DQN- 1D-CNN. For DQN-CNN, the Structure of the Feature Learning Part is CNN. For DQN-1D-CNN, the Structure of the Feature Learning Part is 1D-CNN.

states (s) are a finite batch of images. CNN and 1D-CNN are applied for feature extraction on DQN-CNN and DQN-1D-CNN, respectively. An additional fully connected layer is deployed for final classification and detection. The output is Q-values of the current states, and the prediction/action (a) is decided based on Eqn. (2) with the Q-values. As discussed above, the RL module in the IDS has two sub-modules, the interaction module and the training module. The complete procedure of the interaction module is shown in Algorithm 1. In the beginning, a session from the dataset has been sampled randomly. A session contains numerous images, and each image represents a states. If the current image is not the last image in the session, it store s, r, s_ and a into the replay buffer and continue. If the current image is the last one, the model stores s, r, and a into the replay buffer and randomly samples another dataset session. The complete procedure of the training module is shown in Fig. 7. As mentioned above, Deep Q-Learning algorithm is employed in this experiments. Two networks (the Update and the Target) operate together to achieve the approximate regression. The functionality of the target network is to improve the training stability by fixing the regression target in N steps. The structure of the update network is the same as the agent introduced in Fig. 6. The target network copies the structure from the update network and is initialized with the parameters of the update network. The update network is updated through back propagation, and the target network is updated by copying the parameters from the update network every N times.

The training module and the interaction module work alternately. For example, once the replay buffer is full, training

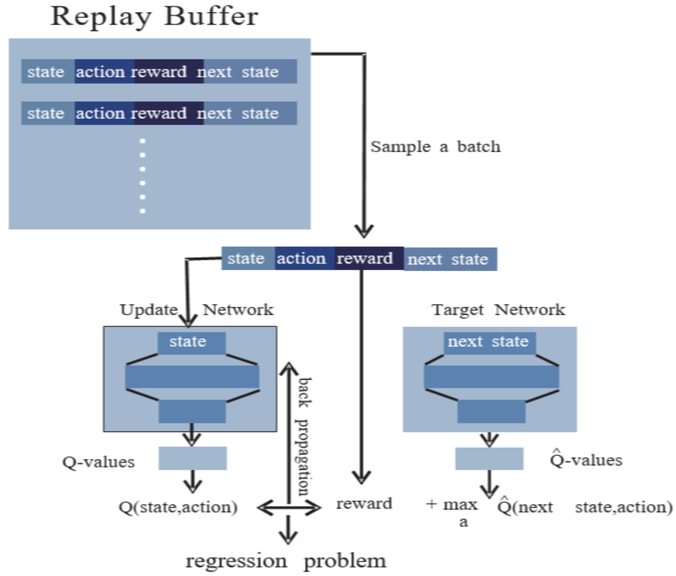


Fig. 7. Deep Q-Learning Training Module. Update Network is Updated through Back Propagation. Target Network is Initialized with the Parameters of Update Network, and Updates are made every N Times through Copying the Parameters from Update Network. The Mean Squared Error (MSE) is used as the Objective Function for the Approximate Regression Problem.

is started for several iterations. After training, the new agent can be used to interact with the environment and store the new data into the replay buffer and remove the old data. The complete pack-level IDS algorithm including both the interaction module and training module is shown in Algorithm 2.

Algorithm 1 – Interaction Module

```
Require: Start Interacting
for interaction process do
    Randomly sample a session and get its label
    Take first image in the session as current state s
    Feed s into agent and obtain action (prediction)
    Feed a and label into reward mechanism, get reward r
    for each episode do
        if Last image in the session then
            Store (s, r, a, None) into replay buffer
            break
        else
            Take next image in the session as next state s
            Store (s, r, a, s ) into replay buffer
            Set s = s_
        end if
    end for
end for
```

D. The Additional Anomaly Detection Module

An anomaly detection model has been employed to detect attacks blind to the training set, by considering them as an anomaly class. This is important to a robust intrusion detection system because it is impossible to include all types of attacks in the training set. As shown in Fig. 8, a confidence score

Algorithm 2 – Deep Q-Learning Framework for Detection at Packet-level

Require: Extract sessions from raw traffic file
Split sessions based on session-based rule
Conduct image embedding on each session
Initialize Q-function Q for agent, target Q-function $\hat{Q} = Q$
for each episode: **do**
  Randomly choose a session from the dataset.
  for each image within session,$t \in [0, N]$ *do*
    Given image s_t, take action at based on Q
    Compare at with true label, obtain reward r_t and $\bar{r}_t$
    Derive next state s_{t+1} :the image behind current state
    Store s_t, a_t, r_t, s_{t+1} into agent replay buffer
  end for
  Sample a batch s_t, a_t, r_t, s_{t+1} from agent replay buffer
  Target $y = r_t + \gamma\ max_a \hat{Q}(s_{t+1}, a)$
  Update Q through back propagation to make $Q(s_t, a_t)$ close to y
  Every C steps reset $\hat{Q} = Q$
end for

TABLE XI. Comparison between the Intrusion Detection System and the Language Model

	IDS	Language Model
Input	Session	Sentence
Element	Packet	Word
Embedding	Image Embedding	Word2vec
Input Format	[batch, len, dim]	[batch, Len, Dim]

has been outputted and set a threshold γ manually. In the experiment, a Softmax layer has been added at the end of the agent and view the output of the Softmax layer (Q-values) as the confidence score of each class. As shown in Eqn. (4), if all confidence scores are smaller than γ, the input will be determined as the 'anomaly' attack. Conversely, the class that belongs to the max confidence score is the expected detection result.

$$f(s) = \{ \begin{matrix} arg\ \underset{a}{max}\ Q(s,a)\ if\ \underset{\alpha \epsilon x}{max} Q(s,a) \leq \lambda \\ anomaly\ if\ \underset{\alpha \epsilon x}{max} Q(s,a) \leq \lambda \end{matrix} \quad ... \qquad (4)$$

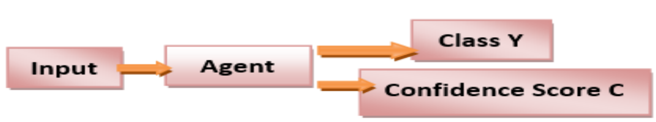


Fig. 8. The Work Flow of Anomaly Detection Model.

E. Dataset

DDoS2019 dataset has been selected as the major dataset for packet-level experiments. DDoS2019 is a relatively new dataset [40] which is collected for DDoS and packet-level research. The datset detals have been shown in Table XII The distributed denial of service (DDoS) attack denotes that multiple attackers in different locations simultaneously attack one or more targets. One attacker takes control of multiple machines in different locations and uses them to attack victims simultaneously. The originating points of the attack are distributed in different places in DDoS attacks, and there can be multiple attackers. DDoS attacks can cause great damages to the society. Various types of DDoS attacks pose huge challenges for intrusion detection systems. The images after image embedding has shown in Fig. 9. In such circumstances, DDoS2019 was collected to facilitate the DDoS research. Eight traffic types are collected in the training set: Normal, PortMap, Net- BIOS, LDAP, MSSQL, UDP, UDP-Lag, SYN. These are multiple types of DDoS attacks which have come up frequently in recent years. DDoS2019 is a balanced and sufficient dataset. Each attack contains at least 10000 samples. The test set contains 12 types of traffic. In addition to the eight types collected in the training set, other four types have been included: NTP, DNS, WebDDoS, TFTP. These additional attacks put forward higher requirements to the generality of the intrusion detection system because they are blind to our intrusion detection system at the training stage. For this problem, an anomaly detection model has been deployed to detect these four additional attacks, by considering them as an anomaly class.

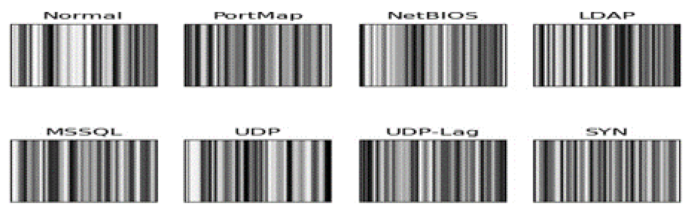


Fig. 9. Images after Image Embedding for each Class.

TABLE XII. Dataset Details

Traffic Class	Session	Max Session Len	Average Session Len	Image
Normal (0)	20000	786	71	31256
PortMap (1)	20000	687	54	25258
NetBIOS (2)	20000	556	69	29564
LDAP (3)	20000	456	34	23102
MSSQL (4)	20000	632	56	27695
UDP (5)	20000	346	53	25210
UDP-Lag (6)	20000	452	43	24750
SYN (7)	20000	563	60	27610

IV. Result and Discussion

A. Evaluation Metrics

Four metrics, including the accuracy, precision, recall and F1 score has been adapted to evaluate the intrusion detection system.

The definition of these four metrics [41][42] is shown in Table XIII and Table XIV. These evaluation metrics are regularly used in two-class classification problems. When calculating TP, FP, FN and TN with respect to each traffic class, the current class as the positive one, and the remaining classes as the negative one use to consider. With the assistance of these machine learning metrics, this intrusion detection system can be evaluated comprehensively.

B. Experimental Results

Deep Q-Learning algorithm combined with CNN and 1D-CNN for this IDS experiments has been used. In this IDS, two agents based on DQN-CNN and DQN-1D-CNN has been

TABLE XIII. CONFUSION MATRIX

Predicted	Actual	
86.86%	**Positive**	**Negative**
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

TABLE XIV. EVALUATION METRICS

Metric	Calculation
Accuracy	(TP+TN)/(TP+FP+FN+TN)
Precision	TP/(TP+FP)
Recall	TP/(TP+FN)
F1	(2*Recall*Precision)/(Recall+Precision)

devise. As a comparison, the model devises another two agents PG-CNN and PG-1D-CNN, which are trained with policy gradient methods. The sessions are extracted from raw network traffic data in 'pcap' files and split according to the session-based rule. The dataset is split in a ratio of 4:1 (16000 sessions for training and 4000 sessions for validation). During the partition, it has been ensured that the session length distribution of the training set and validation set is consistent, thus the image ratio can also reach approximately 4:1. After partition, the statistics of the dataset is shown in Table XV.

TABLE XV. TRAINING AND VALIDATION SET DETAILS

Traffic Class	Training Images	Validation Images
Normal	25012	6256
PortMap	20200	5058
NetBIOS	23652	5912
LDAP	18019	5083
MSSQL	22986	4709
UDP	19159	6051
UDP-Lag	19552	5198
SYN	22640	4970

Experiments are conducted on the computer configured as RTX2070, i5-9600k and 32GB Memory. Three discount values γ, including 0.1, 0.5 and 0.9 on four agents have been compared. The value of accuracy averaged used over each class on validation set to evaluate the performance and select the optimal discount value. The experiment results are shown in Table XVI. It has been noticed that when γ equals to 0.1, all of the four agents attain the highest performance. With regard to reinforcement learning algorithms, DQN-1D-CNN slightly outperforms PG-1D-CNN over all discount factors, and DQN-CNN also marginally outperforms PG-CNN over all discount factors. Meanwhile, it is observed that PG-based agents are more sensitive to discount factors. Overall results indicate that Deep Q-Learning is a relatively better algorithm than policy gradient methods in our experiments. Hence in the following case studies, it has been assumed that the discount value is fixed as 0.1, and mainly adopt Deep Q-Learning algorithms for the agent.

Detailed experiment results of DQN-1D-CNN with discount value γ equal to 0.1 are shown in Table XVII. The table shows that the proposed detection system can reach high accuracy at 98.78%. 100% of normal traffic can be detected with respect to each class. The detection rate (Recall) of MSSQL is the least among all eight types, but still reaching 97.30%. Table XVIII lists the validation results of DQN-CNN. It is seen that the performance of DQN-CNN is slightly worse

TABLE XVI. PERFORMANCES OF DIFFERENT DISCOUNT VALUES ON FOUR AGENTS

Discount Factor	Val Acc (DQN-1D-CNN)	Val Acc (DQN-CNN)
0.1*	**98.78%**	**96.07%**
0.5	95.21%	93.22%
0.9	87.12%	86.24%
Discount Factor	**Val Acc (PG-1D-CNN)**	**Val Acc (PG-CNN)**
0.1	96.17%	96.03%
0.5	90.15%	87.12%
0.9	79.68%	70.69%

than that of DQN-1D-CNN, but still reach a high accuracy at 96.07%. The experimental results prove that both DQN-1D-CNN and DQN-CNN can perform well when dealing with image tasks. In addition, in this experiment, the image can also be readily treated as the time series data, and DQN-1D-CNN can perform much better. The test set details has been shown in Table XIX. Afterwards the proposed DQN-1D-CNN agent has been evaluated on the test set. In this stage, it is required to set another important hyperparameter: λ. λ controls the detection performance of (unknown) anomaly traffic. The known eight types of attacks have been treated as the negative class, and the anomaly traffic as the positive class. We use precision and recall to evaluate the performances and select the optimal λ. Results are shown in Table XX. As λ is increasing, the precision decreases, which means that an increasing volume of network traffic is classified as the anomaly type. Conversely, as λ is increasing, the recall is also increasing, meaning that an increasing amount of anomaly traffic is detected. The value of λ can be determined based on actual requirements of the intrusion detection system. In this experiment, the λ is fixed as 0.7, which is a trade-off between 0.5 and 0.9. Finally, the intrusion detection system is evaluated on the test set with γ = 0.1 and λ = 0.7. Results are shown in Table XXI. Due to the existence of anomaly traffic, the general accuracy has declined greatly to 84.27% compared with the validation accuracy at 96.07%. With respect to all classes, the DQN-1D-CNN agent can attain relatively high performances on those known types. With regard to anomaly types, although these anomaly types are completely blind to this system, 71.05% of them can still be detected.

TABLE XVII. VALIDATION RESULTS OF DQN-1D-CNN WITH DISCOUNT VALUE γ = 0.1. THE COMPUTATION TIME FOR DETECTION IS 0.112S

Traffic Class	Accuracy[1]	Precision	Recall	F1 Score
Normal	98.78%	99.32%	100%	99.66%
PortMap	98.78%	98.04%	98.95%	98.49%
NetBIOS	98.78%	98.87%	99.26%	99.06%
LDAP	98.78%	98.85%	98.37%	98.61%
MSSQL	98.78%	98.47%	97.30%	97.86%
UDP	98.78%	98.95%	99.62%	99.28%
UDP-Lag	98.78%	98.02%	98.31%	98.61%
SYN	98.78%	98.64%	97.83%	98.23%
Average	**98.78%**	**98.65%**	**98.71%**	**98.73%**

[1]According to the calculation formula displayed in Table XIV, accuracy is the general metric of a intrusion detection system, so the accuracy value of each category is the same, and it represents the overall accuracy of the system.

C. Comparison with Deep Learning Approaches

The performance of the proposed approach has been compared with traditional 1D-CNN and CNN approaches (without

TABLE XVIII. Validation Results of DQN-CNN with Discount Value γ = 0.1. The Computation Time for Detection is 0.071s

Traffic Class	Accuracy	Precision	Recall	F1 Score
Normal	96.07%	96.91%	99.68%	98.27%
PortMap	96.07%	95.67%	93.53%	94.59%
NetBIOS	96.07%	97.35%	95.75%	96.55%
LDAP	96.07%	95.76%	96.95%	96.35%
MSSQL	96.07%	95.06%	96.03%	95.54%
UDP	96.07%	95.32%	96.00%	95.66%
UDP-Lag	96.07%	95.05%	94.92%	94.99%
SYN	96.07%	97.14%	94.87%	95.99%
Average	**96.07%**	**96.03%**	**95.97%**	**95.99%**

TABLE XIX. Test Set Details

Traffic Class	Session	Max Session Len	Average Session Len	Image
Normal	5000	691	65	9514
Portmap	5000	578	59	8768
NetBIOS	5000	612	54	6978
LADP	5000	531	41	6013
MSSQL	5000	598	61	9143
UDP	5000	408	46	6784
UDP-Lag	5000	425	39	6102
SYN	5000	601	55	7035
Anomaly	5000*4	752	57	29260

TABLE XX. Anomaly Detection λ Selection

λ_1	Precision	Recall	λ_2	Precision	Recall	λ_3	Precision	Recall
0.5	93.94%	19.17%	**0.7***	85.13%	71.05%	0.9	43.21%	86.86%

TABLE XXI. Test Results of DQN-1D-CNN with Discount Value γ = 0.1 and λ = 0.7. The Computation Time for Detection is 0.143s

Traffic Class	Accuracy	Precision	Recall	F1 Score
Normal	84.27%	88.81%	93.64%	91.16%
PortMap	84.27%	86.83%	91.38%	89.05%
NetBIOS	84.27%	84.62%	90.69%	87.55%
LDAP	84.27%	82.36%	88.31%	85.23%
MSSQL	84.27%	88.24%	92.14%	90.14%
UDP	84.27%	83.55%	89.40%	86.38%
UDP-Lag	84.27%	82.07%	87.92%	84.90%
SYN	84.27%	84.67%	89.57%	87.05%
Anomaly	84.27%	81.12%	71.05%	75.75%
Average	**84.27%**	**84.70%**	**88.23%**	**86.36%**

the reinforcement learning framework). The performances using two metrics. The first one is the detection rate (Recall), which measures how many attacks are correctly detected. The other is the false alarm rate, which measures how many normal traffics are classified as malignant types. Anomaly attacks are excluded in this experiment. Results are shown in Table XXII. It can be seen that traditional deep learning methods can also achieve relatively high performances on the test set, which supports that image embedding is a useful data preprocessing approach for network traffic analysis. Most importantly, the proposed RL-based approach still outperforms these traditional approaches in detecting and false alarm rates without increasing the computation time.

TABLE XXII. Comparison with Deep Learning Approaches (Anomaly Types in the Test Set are Excluded)

Approach	Detection Rate	False Alarm	Detection Time
1D-CNN	92.12%	5.12%	0.096s
CNN	90.68%	6.44%	0.084s
DQN-1D-CNN*	**97.69%**	**0.12%**	0.101s
DQN-CNN	95.14%	3.50%	**0.079s**

V. Conclusion

First of all, a novel image embedding approach has been designed. Using image embedding, raw traffic files can be transferred to images, which are generally difficult to process by artificial intelligence techniques. This is significant because many artificial intelligence techniques can achieve high performance on image tasks. In addition, those packets classified in the same session are arranged in the chronic order of the capture time. This way the image-based task can be extended to the time-series based task, where supplementary AI techniques such as 1D-CNN and LSTM, can also be applied. Secondly, Markov process has been used to model the dynamic process of network session. Therefore, a reinforcement learning framework can be devised to train the system. Introducing discount value γ, it can make the agent consider more farsighted information. This property is also important to the intrusion detection problem. Assuming that the model encounter a malignant flow, if agent is forced to consider more farsighted, the agent can detect the intrusion earlier. Thirdly, an additional anomaly detection system has been designed to detect those unknown attacks, and propose the idea of the confidence score. This anomaly detection module makes it possible to detect some unknown and novel attacks. It is crucial because the real network environment is complicated and ever changed, and hackers will continue to launch novel attacks which are blind to intrusion detection systems. The experimental results show that, with the assistance of CNN, 1D-CNN and reinforcement learning algorithms, the intrusion detection system can attain high detection rate and maintain low false alarm rate.

More research can be done in the future based on this reinforcement learning framework. Firstly, GAN can be introduced into the reinforcement learning framework. Since the traffic data into images already have been transformed, GAN can be used to generate some novel flows or sessions. Furthermore, GAN can use to simulate a dynamic network environment for interaction. Secondly, a more robust and high-accuracy anomaly detection system can be devised. In this experiment, the Q-value has been treated as the confidence score. Nevertheless, it is difficult to adjust the value of λ. One possibility is to train the agent to learn the confidence score by itself.

References

[1] R.-H. Hwang, M.-C. Peng, V.-L. Nguyen, and Y.-L. Chang, "An lstm-based deep learning approach for classifying malicious traffic at the packet level," Ap- plied Sciences, vol. 9, no. 16, 2019, ISSN: 2076-3417. DOI: 10.3390/app9163414. [Online]. Available: https://www.mdpi.com/2076-3417/9/16/3414.

[2] M. Ge, X. Fu, N. Syed, Z. Baig, G. Teo, and A. Robles-Kelly, "Deep learning- based intrusion detection for iot networks," in 2019 IEEE 24th Pacific Rim In- ternational Symposium on Dependable Computing (PRDC), IEEE, 2019, pp. 256– 25 609.

[3] R. Doshi, N. Apthorpe, and N. Feamster, "Machine learning ddos detection for consumer internet of things devices," in 2018 IEEE Security and Privacy Workshops (SPW), IEEE, 2018, pp. 29–35.

[4] B. J. Radford, L. M. Apolonio, A. J. Trias, and J. A. Simpson, "Network traffic anomaly detection using recurrent neural networks," arXiv e-prints, arXiv– 1803, 2018.

[5] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, Deep learning. MIT Press, 2016, vol. 1.

[6] R. S. Sutton and A. G. Barto, Reinforcement Learning: An introduction. MIT press, 2018.

[7] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," arXiv preprint arXiv:1312.5602, 2013.

[8] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," arXiv preprint arXiv:1707.06347, 2017.

[9] Y. Hu, Q. Da, A. Zeng, Y. Yu, and Y. Xu, "Reinforcement learning to rank in e-commerce search engine: Formalization, analysis, and application," in Pro- ceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 368–377.

[10] I. Carlucho, M. De Paula, S. Wang, Y. Petillot, and G. G. Acosta, "Adaptive low-level control of autonomous underwater vehicles using deep reinforcement learning," Robotics and Autonomous Systems, vol. 107, pp. 71–86, 2018.

[11] E. Bøhn, E. M. Coates, S. Moe, and T. A. Johansen, "Deep reinforcement learn- ing attitude control of fixed-wing uavs using proximal policy optimization," in 2019 International Conference on Unmanned Aircraft Systems (ICUAS), IEEE, 2019, pp. 523–533.

[12] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, "OpenAI gym," arXiv:1606.01540, 2016.

[13] P. Henderson, R. Islam, P. Bachman, J. Pineau, D. Precup, and D. Meger, "Deep reinforcement learning that matters," in Proceedings of the AAAI conference on artificial intelligence, vol. 32, 2018.

[14] L. Engstrom, A. Ilyas, S. Santurkar, D. Tsipras, F. Janoos, L. Rudolph, and A. Madry, "Implementation matters in deep policy gradients: A case study on ppo and trpo," in International Conference on Learning Representations, 2020.

[15] P. Henderson, J. Romoff, and J. Pineau, "Where did my optimum go?: An empirical analysis of gradient descent optimization in policy gradient methods," ArXiv, vol. abs/1810.02525, 2018.

[16] M. Andrychowicz, A. Raichuk, P. Stan´czyk, M. Orsini, S. Girgin, R. Marinier, L. Hussenot, M. Geist, O. Pietquin, M. Michalski, et al., "What matters in on-policy reinforcement learning? a large-scale empirical study," in ICLR 2021- Ninth International Conference on Learning Representations, 2021.

[17] M. J. Beal, Z. Ghahramani, and C. E. Rasmussen, "The infinite hidden markov model," Advances in Neural Information Processing Systems, vol. 1, pp. 577– 584, 2002.

[18] C. J. Watkins and P. Dayan, "Q-learning," Machine learning, vol. 8, no. 3-4, pp. 279–292, 1992.

[19] V. R. Konda and J. N. Tsitsiklis, "Actor-critic algorithms," in Advances in Neural Information Processing Systems, Citeseer, 2000, pp. 1008–1014.

[20] M. Babaeizadeh, I. Frosio, S. Tyree, J. Clemons, and J. Kautz, "Reinforcement learning through asynchronous advantage actor-critic on a gpu," arXiv preprint arXiv:1611.06256, 2016.

[21] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," arXiv preprint arXiv:1509.02971, 2015.

[22] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in In- ternational Conference on Machine Learning, PMLR, 2018, pp. 1861–1870.

[23] R. Wehrens, H. Putter, and L. M. Buydens, "The bootstrap: A tutorial," Chemometrics and intelligent laboratory systems, vol. 54, no. 1, pp. 35–52, 2000.

[24] H. Van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double q-learning," in Proceedings of the AAAI Conference on Artificial Intel- ligence, vol. 30, 2016.

[25] Y. Huang, G. Wei, and Y. Wang, "Vd d3qn: The variant of double deep q- learning network with dueling architecture," in 2018 37th Chinese Control Con- ference (CCC), IEEE, 2018, pp. 9130–9135.

[26] C. Berner, G. Brockman, B. Chan, V. Cheung, P. Debiak, C. Dennison, D. Farhi, Q. Fischer, S. Hashme, C. Hesse, et al., "Dota 2 with large scale deep reinforcement learning," arXiv preprint arXiv:1912.06680, 2019.

[27] D. Ye, Z. Liu, M. Sun, B. Shi, P. Zhao, H. Wu, H. Yu, S. Yang, X. Wu, Q. Guo, Q. Chen, Y. Yin, H. Zhang, T. Shi, L. Wang, Q. Fu, W. Yang, and L. Huang, "Mastering complex control in moba games with deep reinforcement learning," in AAAI, 2020.

[28] O. Vinyals, I. Babuschkin, W. M. Czarnecki, M. Mathieu, A. Dudzik, J. Chung, D. H. Choi, R. Powell, T. Ewalds, P. Georgiev, et al., "Grandmaster level in starcraft ii using multi-agent reinforcement learning," Nature, vol. 575, no. 7782, pp. 350–354, 2019.

[29] J. jie Li, S. Koyamada, Q. Ye, G. Liu, C. Wang, R. Yang, L. Zhao, T. Qin, T. Liu, and H. Hon, "Suphx: Mastering mahjong with deep reinforcement learning," ArXiv, vol. abs/2003.13590, 2020.

[30] S. A. Salloum, M. Alshurideh, A. Elnagar, and K. Shaalan, "Machine learning and deep learning techniques for cybersecurity: A review.," in AICV, 2020, pp. 50–57.

[31] G. Biau, "Analysis of a random forests model," The Journal of Machine Learn- ing Research, vol. 13, pp. 1063–1095, 2012.

[32] W. S. Noble, "What is a support vector machine?" Nature Biotechnology, vol. 24, no. 12, pp. 1565–1567, 2006.

[33] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, "Object recognition with gradient-based learning," in Shape, Contour and Grouping in Computer Vision, Springer, 1999, pp. 319–345.

[34] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1d convolutional neural networks and applications: A survey," Mechanical systems and signal processing, vol. 151, p. 107 398, 2021.

[35] J. F. Kurose and K. W. Ross, Computer Networking: A Top-Down Approach (6th Edition), 6th. Pearson, 2012, ISBN: 0132856204.

[36] U. Lamping and E. Warnicke, "Wireshark user's guide," Interface, vol. 4, no. 6, p. 1, 2004.

[37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recog- nition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[38] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[39] W. Tang, G. Long, L. Liu, T. Zhou, J. Jiang, and M. Blumenstein, "Rethink- ing 1d-cnn for time series classification: A stronger baseline," arXiv preprint arXiv:2002.10061, 2020.

[40] I. Sharafaldin, A. H. Lashkari, S. Hakak, and A. A. Ghorbani, "Developing realistic distributed denial of service (ddos) attack dataset and taxonomy," in 2019 International Carnahan Conference on Security Technology (ICCST), IEEE, 2019, pp. 1–8.

[41] Abhijit Das, Pramod and Sunitha B S, "An Efficient Feature Selection Approach for Intrusion Detection System using Decision Tree" International Journal of Advanced Computer Science and Applications(IJACSA), 13(2), 2022. http://dx.doi.org/10.14569/IJACSA.2022.0130276

[42] Abhijit Das, Pramod and Sunitha B S, "Anomaly-based Network Intrusion Detection using Ensemble Machine Learning Approach" International Journal of Advanced Computer Science and Applications(IJACSA), 13(2), 2022. http://dx.doi.org/10.14569/IJACSA.2022.0130275

An Enhanced Predictive Approach for Students' Performance

Mohamed Farouk Yacoub[1], Huda Amin Maghawry[2],
Nivin A Helal[3], Tarek F. Gharib[5]
Faculty of Computer and Information Sciences
Ain Shams University
Cairo, Egypt

Sebastián Ventura[4]
Department of Computer
Sciences and Numerical Analysis
University of Cordoba
Spain

***Abstract*—Applying data mining for improving the outcomes of the educational process has become one of the most significant areas of research. The most important corner stone in the educational process is students' performance. Therefore, early prediction of students' performance aims to assist at-risk students by providing appropriate and early support and intervention. The objective of this paper is to propose an enhanced predictive model for students' performance prediction. Selecting the most important features is a crucial indicator for the academic institutions to make an appropriate intervention to help students with poor performance and the top influencing features were selected in feature selection step besides the dimensionality reduction and build an efficient predictive model. DB-Scan clustering technique is applied to enhance the proposed predictive model performance in the preprocessing step. Various classification techniques are used such as Decision Tree, Logistic regression, Naive Bayes, Random Forest, and Multilayer Perceptron. Moreover ensemble method is used to solve the trade-off between the bias and the variance and there are two proposed ensemble methods through the experiments to be compared. The proposed model is an ensemble classifier of Multilayer Perceptron, Decision Tree, and Random Forest classifiers. The proposed model achieves an accuracy of 83.16%.**

Keywords*—Educational data mining; students' performance; classification; feature selection; machine learning

I. Introduction

Educational systems contain massive data about students' behavior, enrollment of students, results, and attendance that could be analyzed to improve outcomes of the educational process [1]. Therefore, Educational Data Mining (EDM) has become a necessity to discover knowledge that helps decision makers to improve the educational process [2].EDM can reduce the drop-out rate by providing the academic institutions with knowledge to be able to develop appropriate strategies. It can also provide a timely decision to help students who are vulnerable to failure. Data mining deals with the educational field to identify and evaluate several important learning indicators from the data [3].

Students' performance prediction is one of the most challenging and interesting topics of EDM research [4]. It helps instructors to track their students' performance to identify those at-risk [12]. Research on students' performance prediction is useful to identify the features, behaviors and hidden relations that affect the students' performance [6][11].Many organizations, such as Accreditation Council for Business Schools and Programs (ACBSP) and Accreditation Board for Engineering and Technology (ABET) evaluate the educational programs quality based on the learning outcomes [7]. There are two research perspectives, the first one is to find the features that affect the performance of students. The second one is to find an effective methodology to predict students' performance [4]. It is very crucial to apply the feature selection to discover high influence features that need to be improved the dropout rate and enhance student performance [5]. Machine learning is applied to discover hidden patterns and the relations between the features in addition to the prediction of at-risk students [20].

The massive growth of the educational data gives the education institutes the opportunity to apply data mining techniques to extract useful and hidden information for predicting students' academic performance effectively [22]. Therefore, the objective of this paper is to propose an enhanced predictive model for accurate prediction of students' performance. Various machine learning techniques were experimented. For enhancing the predictive model, we applied DB-Scan clustering technique and feature selection approach. Experimental results proved the effectiveness of the proposed model. The following sections of the paper are organized as follows: Section II presents the related work in page 1 and 2, Section III introduces the methodology in page 2, experiments and results are discussed in Section IV in page 2, 3, and 4, and finally conclusion is stated in Section V in page 5.

II. Related Work

Predicting learning outcomes and especially students' performance became very important in the overall learning and educational process. Francis et al. achieved an accuracy of 75.47% using a hybrid data mining technique to predict weak students [8]. (Pojon, 2017) compared various machine learning techniques such as Decision Tree, Linear Regression, and Naïve Bayes classifiers [13]. The experiments were conducted on two datasets. Experiments have shown that models with clustering and classification techniques achieve better results in predicting students' performance. The accuracy achieved using classification and regression trees (CART) algorithm were of 93% and 78%.

Lau et al. built a prediction model for students' performance using neural networks technique. The model contains two hidden-layers and the output layer with 11 features as

input. The prediction model achieved precision of 84.8% [10]. Xing et al. conducted a prediction model using deep learning approach based on the perspective of predicting temporal dropout for improving the online learning. The proposed model achieved an accuracy of 90.8% and 96.1%, after the 1st week and the 7th week, respectively [21]. Divyabharathi et al. constructed a predictive model using Naïve Bayes classification technique to detect and prevent performance academic risk based on the students' data, and they achieved an accuracy of 94% [22]. Raihana et al. applied classification based on the academic performance as well as the quality of life that were psychological health, physical health, social relationship, environment, and the overall life quality. They used support vector machine (SVM) algorithm and achieved an accuracy of 73.33%. The experiments revealed that students, who achieved good academic performance, were psychologically healthy, physically healthy, have good social relationship, were in a good environment and have good overall life quality [23]. Uzel et al. experimented various machine learning predictors such as Decision Tree, Naïve Bayes, Random Forest, Multilayer Perceptron, and Ensemble method. They also applied Apriori technique to discover the hidden patterns from the data. They achieved the accuracy of 80.6% for the classification by the voting method classifier[19].
Sana et al. showed that there are many features and factors that affect the final students' performance significantly such as the number of student's absence days and the involvement of parents with students in the learning process. The accuracy of 78.1% was achieved by their approach using Artificial Neural Network (ANN) algorithm with highly ranked features [18].

Comparative results of some related works are shown in Table I.

TABLE I. COMPARATIVE RESULTS OF RELATED WORK

Ref.	Algorithm	Accuracy	Dataset
[8]	Hybrid Data Mining Approach	75.47%	480 records and 16 features
[10]	Neural Networks	84.8%	1000 records
[13]	CART	78%	480 records and 16 features
[18]	ANN	78.1%	480 records and 16 features
[19]	Voting Classifier	80.6%	480 records and 16 features
[21]	Deep Learning	96.1%	3617 records and 13 features
[22]	Naïve Bayes	94%	500 records and 8 features
[23]	SVM	73.33%	60 records and 26 features

There are many studies that focus either on finding hidden patterns, discovering relation between features, or enhancing the accuracy. However there is a lack in making a comprehensive work that combines enhancing the preprocessing step, finding the top influencing features, and proposing a predictive model to enhance the performance based on different measures to check the model's stability from different perspectives and show the variations of results to get insights from the experiments. Therefore the performance of prediction models in the previous studies needs to be enhanced to help at-risk students to improve their performance and help the academic institutions to make an appropriate intervention to assist the students with poor performance. Therefore, in this study we propose a predictive model to face these limitations and improve students' performance prediction.

III. METHODOLOGY

The purpose of this paper is to propose a predictive model for students' performance prediction. This is achieved by exploring various classification techniques, besides the ensemble method that solves the trade-off between the bias and variance, to investigate which one would achieve the best performance. Moreover, DB-Scan was used in preprocessing for outlier detection and features selection was used to enhance the predictive model of students' performance. The proposed model is shown in Fig. 1. The following subsections describe the proposed model.

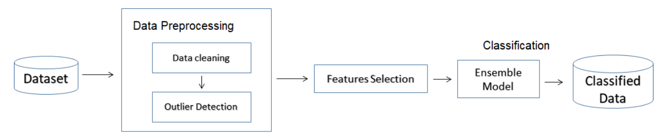


Fig. 1. The Proposed Model.

A. Preprocessing

Pre-processing is an essential process for any data set. It includes data cleaning and transformation. We pre-processed the dataset through three steps. Firstly, data was converted from nominal to numerical values. Secondly, some features were reshaped to be within a certain range using standardization method. Finally, DB-Scan clustering methodology was applied for outlier detection, as it had a great efficiency in [16].

B. Feature Selection

Feature selection aims to select the most important and influencing features in the dataset. Also, it is very important for dimensions' reduction before implementing the prediction and classification methods. It works by selecting the best features that contribute most to the target variable based on univariate statistical tests.

We used the SelectKBest technique [15]. SelectKBest technique selects the first k features with the highest score values based on the Chi-Square test, for comparing the actual and predicted results, as a score function [14] using equation (1).

$$X^2 = \sum (Oi – Ei)^2 / Ei \qquad (1)$$

where Oi and Ei are the actual and expected values, respectively.

C. Classification

For classification, we designed two different Ensemble models. One consists of Multilayer Perceptron (MLP), Random Forest (RF) and Decision Tree (DT), the other consists of RF, Logistic regression (LR) and DT.

IV. EXPERIMENTS AND RESULTS

To evaluate how the proposed models will perform in predicting the students' performance, several evaluation measures were used on a Learning Management System (LMS) dataset through several experiments as follows in page 3:

A. Dataset

The dataset is collected from a Learning Management System (LMS) [24]. It contains 480 records of students' data in various educational levels with 16 features. The features were categorized as follows:

• Academic features: Section id, Semester, Educational stages, Viewing announcements, Grade levels, Topic, Discussion groups, Visited resources, Raised hand, and Student absence days.

• Personal features: Gender, Parent responsible for student, Parent answering survey, and Parent school satisfaction.

• Demographic features: Nationality and Place of birth.

We pre-processed the dataset through two steps. Firstly, we converted the data from nominal to numerical values for the features: Section Id, Semester, Educational stages, Grade levels, Topic, Discussion groups, Gender, Parent responsible for student, Parent answering survey, Parent school satisfaction, Nationality, Place of birth, and Student absence days.

Secondly, the following features: Viewing announcements, Discussion, Visited resources, and Raised hand, were reshaped to be within a certain range using standardization method. Fig. 2 shows a sample of the dataset's features and instances.

	gender	NationalITy	PlaceofBirth	StageID	GradeID	SectionID	Topic	Semester	Relation	raisedhands	VisITedResources	AnnouncementsView	Discussion	ParentAnsweringSurvey	ParentschoolSatisfaction	StudentAbsenceDays
470	0	11	5	1	5	0	8	0	0	0.81	0.868687	0.877551	0.408163	0	1	0
471	0	11	5	1	5	0	8	1	0	0.78	0.828283	0.795918	0.530612	0	1	0
472	0	11	11	1	5	0	11	0	0	0.80	0.878788	0.755102	0.683673	0	1	0
473	0	11	11	1	5	0	11	1	0	0.85	0.888889	0.806122	0.704082	0	1	0
474	1	5	5	1	5	0	10	0	0	0.02	0.070707	0.040816	0.071429	1	0	1
475	1	5	5	1	5	0	10	1	0	0.05	0.040404	0.051020	0.071429	1	0	1
476	1	5	5	1	5	0	11	0	0	0.50	0.777778	0.142857	0.275510	1	0	0
477	1	5	5	1	5	0	11	1	0	0.55	0.747475	0.255102	0.285714	1	0	0
478	1	5	5	1	5	0	8	0	0	0.30	0.171717	0.142857	0.571429	1	0	1
479	1	5	5	1	5	0	8	1	0	0.35	0.141414	0.234694	0.622449	1	0	1

Fig. 2. Sample of the Dataset Features and Instances.

B. Experiment Environment

Anaconda Navigator [17] was used to simplify the packages management and deployment. The implementation language was Python using pandas, numpy, and sklearn libraries. The experiments were performed on a machine with 2.60 GHz processor, 16 GB memory, and Windows 10 64-bit operating system. All the experiments were performed using the 10-fold cross validation [9].

C. Evaluation Measures

The accuracy measure is used for the evaluation of each model using equation (2).

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \tag{2}$$

It indicates a proportion of correctly predicted observation to the total observations, where: TP = True positive, FP = False positive, TN = True negative and FN = False negative.

Moreover, Precision, Recall, and F1-Score measures are used for the evaluation of each model using equations (3), (4), and (5) respectively.

$$\text{Precision} = (TP) / (TP + FP) \tag{3}$$

$$\text{Recall} = (TP) / (TP + FN) \tag{4}$$

$$\text{F1-Score} = 2 \text{ X } (\text{Precision X Recall}) / (\text{Precision} + \text{Recall}) \tag{5}$$

D. Experiments

According to dataset features, the performance indicator was the success levels that were classified to three categories:

• Low-Level contains 127 of the data instances.
• Middle-Level contains 211 of the data instances.
• High-Level contains 142 of the data instances.

To evaluate our ensemble methods, several experiments were conducted. First, we studied the effect of using the DB-Scan as a preprocessing step as shown in the methodology section.

A comparison of our first and second ensemble classifiers' performance results before and after Applying DB-Scan in terms of accuracy, Precision, Recall, and F1-Score measures is shown in Table II.

TABLE II. PERFORMANCE OF OUR FIRST AND SECOND ENSEMBLE METHODS BEFORE AND AFTER DB-SCAN

Method	Accuracy	Precision	Recall	F1-Score
Our First Ensemble (MLP, RF, DT) (Before DB-Scan)	79.17%	79.2%	81.3%	80.1%
Our First Ensemble (MLP, RF, DT) (After DB-Scan)	83.16%	84.1%	83.3%	83.6%
Our Second Ensemble (RF, DT, LR) (Before DB-Scan)	78.125%	78.2%	80.6%	79.2%
Our Second Ensemble (RF, DT, LR) (After DB-Scan)	80.0%	80.7%	80.8%	80.7%

A comparison of classifiers' performance showing the evaluation results in terms of accuracy, Precision, Recall, and F1-Score measures after applying the DB-Scan algorithm as well as the evaluation of these classifiers using all the data without applying the DB-Scan clustering technique, and theses comparison's results are shown in Table II.

Second experiment as shown in Table III is designed to evaluate our two ensemble models comparable to the Naïve Bayes, DT and MLP models. Evaluation is done using the accuracy, precision, recall and F1-Score.

In addition to assess the accuracy of our proposed models, we compared our work with the results of [18]. We compared our two ensemble models using the given evaluation measures.

The experiments in table IV show that our first ensemble model achieved better performance in terms of different evaluation measures than [18] and our second ensemble model.

TABLE III. PERFORMANCE OF OUR FIRST AND SECOND ENSEMBLE METHODS AS WELL AS NAÏVE BAYES, DECISION TREE, AND MLP CLASSIFIERS

Method	Accuracy	Precision	Recall	F1-Score
Naive Bayes	81.05%	81.2%	83.1%	81.6%
Decision Tree	77.89%	78.7%	78.0%	78.2%
MLP	80.0%	81.2%	80.0%	80.5%
Our First Ensemble (MLP, RF, DT)	83.16%	84.1%	83.3%	83.6%
Our Second Ensemble (RF, DT, LR)	80.0%	80.7%	80.8%	80.7%

TABLE IV. PERFORMANCE OF OUR FIRST AND SECOND ENSEMBLE METHODS BESIDES [18]

Method	Accuracy	Precision	Recall	F1-Score
[18]	78.1%	78.6%	78.9%	78.8%
Our First Ensemble (MLP, RF, DT)	83.16%	84.1%	83.3%	83.6%
Our Second Ensemble (RF, DT, LR)	80.0%	80.7%	80.8%	80.7%

Finally, to study the feature selection technique, the SelectKBest technique is applied based on chi-square test for feature selection approach for selecting the most influencing features to be involved in the model building phase. It works by selecting the best features that contribute most to the target variable based on univariate statistical tests.

The experiments were implemented using the top 10 and 5 features besides all the features from the dataset to be used in the classification.

Table V shows a comparison of models' performance in terms of accuracy measure using all the features of dataset, top 10, and top 5 features of the dataset correlated to the target.

TABLE V. PERFORMANCE ACCURACY WITH VARIOUS NUMBER OF SELECTED BEST FEATURES

Method	All Features	Top 10 Features	Top 5 Features
Naive Bayes	81.05%	80.0%	64.21%
Decision Tree	77.89%	68.42%	66.32%
MLP	80.0%	81.05%	68.42%
Our First Ensemble (MLP, RF, DT)	83.16%	78.95%	65.26%
Our Second Ensemble (RF, DT, LR)	80.0%	75.79%	65.26%

According to Fig. 3, the accuracy corresponding to each number of top K features is measured and plotted for selecting the best number of features in feature selection stage and the best accuracy can be achieved by using only the top correlated 9 features to the target that is the accuracy when we used all the features.

Table VI shows a comparison of models' building time in milliseconds using all the features of dataset, top 10, and top five features of the dataset correlated to the target. The results showed that the difference between models' building time is very small. Therefore building models using all the features

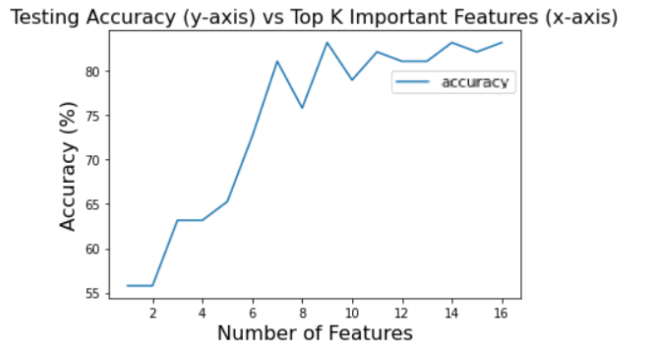


Fig. 3. Features' Selection Graph.

is the best choice. Moreover Fig. 4 shows the bar-chart graph representation of the results of Table VI.

TABLE VI. COMPARISON BETWEEN MODELS' BUILDING TIME IN MILLISECONDS WITH VARIOUS NUMBER OF SELECTED BEST FEATURES

Method	All Features	Top 10 Features	Top 5 Features
Naive Bayes	8.026	7.947	6.952
Decision Tree	10.121	9.091	8.285
MLP	145.065	162.463	155.802
Our First Ensemble (MLP, RF, DT)	217.56	231.273	207.702
Our Second Ensemble (RF, DT, LR)	127.004	142.911	110.764

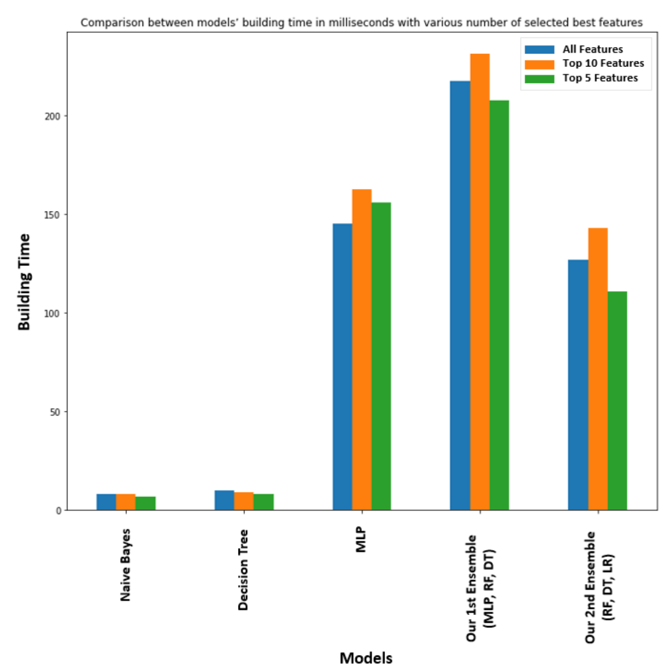


Fig. 4. Models' Building Time Graph.

V. CONCLUSION

One of the essential objectives of educational data mining is to accurately predict students who are vulnerable to dropout for providing them with more support and the suitable intervention. EDM helps the academic institutions to make an appropriate intervention to assist those students enhance their performance. In this paper, we proposed an enhanced predictive model for students' performance to improve the prediction accuracy. We applied various machine learning techniques for predicting the students' performance. Additionally, DB-Scan clustering algorithm and feature selection steps have been exploited, for choosing the significant features. Our first ensemble method has achieved an accuracy of 83.16%, 78.95%, and 65.26% using all the features, the top 10 influencing features, and the top 5 influencing features, respectively. The proposed predictive model outperformed previous work using the same dataset from the learning management system. Applying DB-Scan clustering technique as a preprocessing step has a great effect on enhancing the predictive model performance and the distribution of results as seen in the confusion matrix of each predictive model. For future work, we intend to apply the proposed predictive approach to various datasets, experiment different feature selection techniques, and implement alternatives for DB-Scan clustering technique.

REFERENCES

[1] A. Dutt, M. A. Ismail, and T. Herawan, "A systematic review on educational data mining", IEEE Access, IEEE Xplore, 2017.

[2] A. Narvekar, V. Menezes, O. P. Angle, M. Lotlekar, S. Naik, and A. Purohit, "Students performance prediction using data mining techniques", International Journal for Research in Applied Science and Engineering Technology (IJRASET), 2020.

[3] A. S. Olaniyi, S. Yakub, H. Moshood, S. Ibrahim, and A. Nathaniel, "Student's performance analysis using decision tree algorithms", Anale: Seria Informatică. Computer Science Series, 2017.

[4] A. Khan, S. K. Ghosh, "Student performance analysis and prediction in classroom learning: A review of educational data mining studies", Educ Inf Technol 26, 205–240, Springer, 2020.

[5] A. U. Khasanah, H. Harwati, "A comparative study to predict student's performance using educational data mining techniques", Institute of Physics Conference Series: Materials Science and Engineering, Volume 215, 2017.

[6] A. Namoun, A. Alshanqiti, "Predicting student performance using data mining and learning analytics techniques: a systematic literature review", Applied Sciences, 11(1), 237, 2021.

[7] A. Rajak, A. K. Shrivastava, and D. P. Shrivastava, "Automating outcome based education for the attainment of course and program outcomes", Fifth HCT Information Technology Trends (ITT), pp. 373–376, 2018.

[8] B. K. Francis and S. S. Babu, "Predicting academic performance of students using a hybrid data mining approach", Journal of Medical Systems, 2019.

[9] C. Goutte and J. Larsen, "Optimal cross-validation split ratio: experimental investigation", International Conference on Artificial Neural Networks, pp. 681-686, 1998.

[10] E. T. Lau, L. Sun, Q. Yang, "Modelling, prediction and classification of student academic performance using artificial neural networks", SN Applied Sciences, 2019.

[11] L. M. Zohair, "Prediction of student's performance by modelling small dataset size", International Journal of Educational Technology in Higher Education, 16-27, 2019.

[12] M. Swathi, K. L. Soujanya, R. Suhasini, "Review on predicting student performance", International Conference on Communications and Cyber Physical Engineering (pp.1323-1330), Springer, 2020.

[13] M. Pojon and J. Laurikkala, "Using machine learning to predict student performance (Master's thesis)", University of Tampere, Finland, 2017.

[14] R. Miller and D. Siegmund, "Maximally selected chi square statistics", Biometrics, pp. 1011-1016, 1982.

[15] R. Nair and A. Bhagat, "Feature selection method to improve the accuracy of classification algorithm", International Journal of Innovative Technology and Exploring Engineering (IJITEE), 2019.

[16] R. Rupasi, T. Jwalitha, R. Sudarshan, T. Manasa, and G. Gundabatini, "Analysis of educational data mining using machine learning algorithms", Journal of Xi'an University of Architecture & Technology, 2021.

[17] R. Phelps, M. Krasnicki, R. A. Rutenbar, L. R. Carley, and J. R. Hellums, "Anaconda: robust synthesis of analog circuits via stochastic pattern search", IEEE, 1999.

[18] Sana, I. F. Siddiqui, and Q. A. Arain, "Analyzing students' academic performance through educational data mining", 3C Tecnología. Glosas de innovación aplicadas a la pyme. Edición Especial, pp. 402–421, 2019.

[19] V. N. Uzel, S. S. Turgut, and S. Ayşe, "Prediction of students' academic success using data mining methods", 2018 Innovations in Intelligent Systems and Applications Conference (ASYU), IEEE, 2018.

[20] W. F. Yaacob, N. M. Sobri, S. A. Nasir, N. D. Norshahidi, W. Z. Husin, and K. Shaalan, "Predicting student drop-out in higher institution using data mining techniques Institute of Physics Conference Series: Materials Science and Engineering, 2020.

[21] W. Xing and D. Du, "Dropout prediction in moocs: using deep learning for personalized intervention", Journal of Educational Computing Research, 2019.

[22] Y. Divyabharathi and P. Someswari "A framework for student academic performance using naive bayes classification technique", Journal of Advancement in Engineering and Technology, 2018.

[23] Z. Raihana and A. M. Nabilah, "Classification of students based on quality of life and academic performance by using support vector machine", Journal of Academia UiTM Negeri Sembilan Vol. 6, pp. 45-52, 2018.

[24] Cortez, P. (2008) [Students' Academic Performance Dataset]. https://www.kaggle.com/aljarah/xAPI-Edu-Data

A Comparative Performance of Optimizers and Tuning of Neural Networks for Spoof Detection Framework

Ankita Chadha
School of Computer Science,
Taylors University, Subang Jaya,
Selangor, Malaysia 47500

Azween Abdullah
School of Computer Science,
Taylors University, Subang Jaya,
Selangor, Malaysia 47500

Lorita Angeline
School of Computer Science,
Taylors University, Subang Jaya,
Selangor, Malaysia 47500

***Abstract*—The breakthroughs in securing speaker verification systems have been challenging and yet are explored by many researchers over the past five years. The compromise in security of these systems is due to naturally sounding synthetic speech and handiness of the recording devices. For developing a spoof detection system, the back-end classifier plays an integral role in differentiating spoofed speech from genuine speech. This work conducts the experimental analysis and comparison of up-to-date optimization techniques for a modified form of Convolutional Neural Network (CNN) architecture which is Light CNN (LCNN). The network is standardized by exploring various optimizers such as Adaptive moment estimation, and other adaptive algorithms, Root Mean Square propagation and Stochastic Gradient Descent (SGD) algorithms for spoof detection task. Furthermore, the activation functions and learning rates are also tested to investigate the hyperparameter configuration for faster convergence and improving the training accuracy. The counter measure systems are trained and validated on ASV spoof 2019 dataset with Logical (LA) and Physical Access (PA) attack data. The experimental results show optimizers perform better for LA attack in contrast to PA attack. Additionally, the lowest Equal Error Rate (EER) of 9.07 is obtained for softmax activation with SGD with momentum wrt LA attack and 9.951 for SGD with nestrov wrt PA attack.**

Keywords*—Spoof detection; speech synthesis; voice conversion; convolutional neural networks; optimizers; gradient descent algorithm; spoofed speech; automatic speaker verification

I. Introduction

The uniqueness of voice makes it a popular choice as a biometric for securing smart-phones, telephonic-verification for banking, online-shopping and interestingly, voice-based logins. This also requires the voice biometric based systems to be resilient to unauthentic access in the form of spoofing attacks. The Automatic Speaker Verification (ASV) systems are thriving to make the voice-based applications secured through countermeasures or spoof detection algorithms. These spoof detection systems are embedded in the ASV pipeline as standalone or as a sub-part of the security stream. Apparently, the spoofing attacks may be categorized as synthetic or replay attacks. The synthetic speech (SS) is generated by means of Voice Conversion (VC) or Text-to-Speech (TTS) generators. The replay speech is acquired through careful filtering of the target speech through a recording device. Apart from this category, the spoofing attacks may be direct or Physical Access (PA) attacks and indirect or Logical Access (LA) attacks. The former requires physical space i.e., microphone while the latter is conducted by direct injection of speech, exempting the sensor involvement. Lately, there is rise in spoofing attacks on the ASV systems due to impeccable quality of synthetic speech generators and cheaper high-end recording devices. Yet, these attacks are not preventable, but they can be detected by building countermeasures ensuring the safety of the ASV framework.

II. Related Literature

The robustness of a spoof detection system depends on its internal building blocks which includes the feature extraction and classification. During the training mode, the input speech is processed to reduce redundancy in data and filter out the required information from the speech, that is, naturality and speaker specific content. These unique features are then trained using appropriate machine learning algorithm to get a training rule or trained model. Consequently, in the testing mode the appropriate features are extracted from the test samples and fed to the trained model following which the samples are categorized as genuine or spoofed. For building a model, the number of samples and types of attacks must be considered in a dataset as they will help boast the validation accuracy. The datasets available for PA attacks are Red Dots [1], VoicePA [2] and ASV spoof 2017 [3] while for LA attack, SAS [4] and ASV spoof 2015 are popular . The ASV spoof 2019 [5], [6] has all three kind of attacks samples including synthetic and replay speech. Hence, it is chosen over all other datasets for developing an anti-spoofing algorithm.

With regards to capturing human and synthetic traits from the speech, the glottal excitation and source-filer parameters have been extracted along with prosodic features. The Linear Prediction Co-efficient (LPC) [7], Linear Prediction-Residual parameters [8] and Line Frequency Cepstral Co-efficient (LFCC) [9] based spectral features are found to represent speech quite profoundly. Additionally, the perceptual parameters have also been explored as they have similarity with human perceptual filter bank. The Mel Frequency Cepstral Co-efficient (MFCC), Constant Q-factor Cepstral Co-efficient (CQCC) [10] and Cochlear Filter Cepstral Coefficients with Instantaneous Frequency (CFCCIF) [9] are successfully tested perceptual features for LA attack. The Constant Q transform (CQT) unlike the standard Fourier Transform has irregular frequency bins that allows it to maintain a constant Q-factor throughout the spectrum [11]. This promotes evident discrimination within the spectrum as spoofing related characteristics are revealed distinctly. Hence, considering the efficiency of CQT parameters, we are using these as the feature extraction technique.

In addition, the classification techniques have also caught attention of researchers for building a robust spoof detection system. Usually, the Gaussian Mixture Models (GMM) and Universal Background Models (UBM) are employed as they perform well with almost all the feature extraction techniques [8]. Yet, the main shortcoming of this arrangement is that the GMM and UBM are trained independently making them unaware of each other's learning rule. To overcome this shortcoming, the GMM and Support Vector Machines (SVM) were explored as they were more versatile and performed better than the GMM-UBM duo [12]. Additionally, the GMM classifier is limited to perform better with features of low dimensionality [13]. Furthermore, the Deep Learning models have been investigated and perform comparatively well in contrast to the shallow models [14]. Of course, they require a lot of labelled data for the training but also fail to capture temporal information simultaneously. This lead to exploring other deep models like Convolutional Neural Networks (CNN) [15], Recurrent Neural Networks (RNN) [16], Long Short-Term Memory (LSTM) [17], Gated Recurrent Units (GRU)[15] and Generated Adversarial Networks (GAN) [10]. In [16], [18], CNN-RNN are combined to explore effectiveness of both models individually. Thus, the CNNs are used when spatial learning is of importance while in case of temporal learning, RNNs and its variants are used. Moreover, the CNNs are powerful in handling large amount data at the cost of high training time and large number of parameters. To overcome this, the Light-CNN (LCNN) architecture is introduced to avoid repetition in the parameters and ultimately improve training resources [19]. The LCNN based fusion architecture achieved best results in the ASV spoof 2017 challenge with lower EER for replay attack [10]. Then onwards, LCNN gained popularity and has been tried for synthetic speech detection as well [20]. Following which, in the ASV spoof 2019 challenge, an improved version of this architecture with angular-softmax was presented for both LA and PA attack [21]. The major highlight of the LCNN is its ability to achieve generality for variation in data distribution such as recording conditions, noisy speech, speaker variations, etc [19]. This is possible by endowing in optimization of these networks. This leads to wider range of applications of the network along with impeccable theoretical proofs. Even though the optimization algorithms have been existing for more than two decades, through continuous refinement for highly complex networks with large data size, a defined reassessment of these state-of-the-art optimizers is the need of the day. Inspite of the popularity of LCNNs, according to author's knowledge, there is no significant work found in optimizing the LCNN for spoof detection task. Moreover, tuning of the hyperparameters improves the performance of the network with faster convergence. This work also dedicates its attention to various activation functions as they hold a crucial role in deciding the category of the unknown test sample which ultimately contributes to lower model loss and increase performance accuracy. Also, a precise choice of activation might prove to enhance the training time by making the network learn complex patterns easily. Different combinations of learning rates, activations and optimizers have been investigated in this work to determine most suitable model parameters for spoof detection task. Thus, the objectives of this work can be summarized as following:

i An extensive comparison of various optimizers is performed using ASVspoof 2019 dataset for LA and PA attacks. The common optimizers compared include Adaptive moment estimation (Adam), Adaptive-gradient (Adagrad), Adaptive-delta (Adadelta), Nesterov Accelerated Adam (Nadam), Root Mean Square Propagation (RMSprop), Stochastic Gradient Descent (SGD), SGD with nesterov accelerated gradient (NAG) and momentum. This unravels certain unexpected results as against the usual classification problem where the RMS prop performs equally well with the gradients and delta versions of Adam.

ii Exploring activation functions popularly used in training the transfer function are compared with variations in optimizers to suit the spoof detection application for the LCNN framework. These activations include Softmax, Rectified Linear Unit (ReLU) and Logistic function.

iii The experimental results are compared with state-of-the-art softmax-Adam optimizer and evaluated using Equal Error Rate (EER) along with Receiver Operating Characteristics (ROC) Curve.

The article is arranged as follows: Section III ASV based spoof detection framework and Section IV describes the LCNN architecture and hyperparameter testing. The Section V includes the experimental results and discussion while Section VI is the Conclusion of the work.

III. Spoof Detection System

The proposed spoof detection framework is portrayed in Fig. 1 showing two phases of operation. The training phase is also called enrolment phase where known authentic speaker samples are enrolled along with spoofed speech samples. Initially, the CQT features are extracted to obtain a spectrographic representation of speech [22]. The two dimensional spectrogram along with their labels is then fed to the LCNN architecture to obtain the trained speaker model using a loss function. Furthermore, in the testing phase, the unknown test samples which are a mixture of genuine and spoof speaker samples, are represented using CQT features. This feature set is then tested using LCNN classifier and categorized as spoof or genuine. The CQT based features extraction and LCNN classifier are explained in following sub-Sections III-A and III-B.

A. *Front-end CQT Features*

The CQT features were introduced few decades ago as an alternative to short-time Fourier transform (STFT) [22]. The STFT being a filtering technique for a long spectrum broken down into shorter windows leads to an increasing Q-factor towards higher frequencies. This is exactly opposite to the human speech perception model where the Q-factor is found to be constant from 500Hz to 20kHz. Thus, STFT fails to represent the human perception model and CQT is preferred.

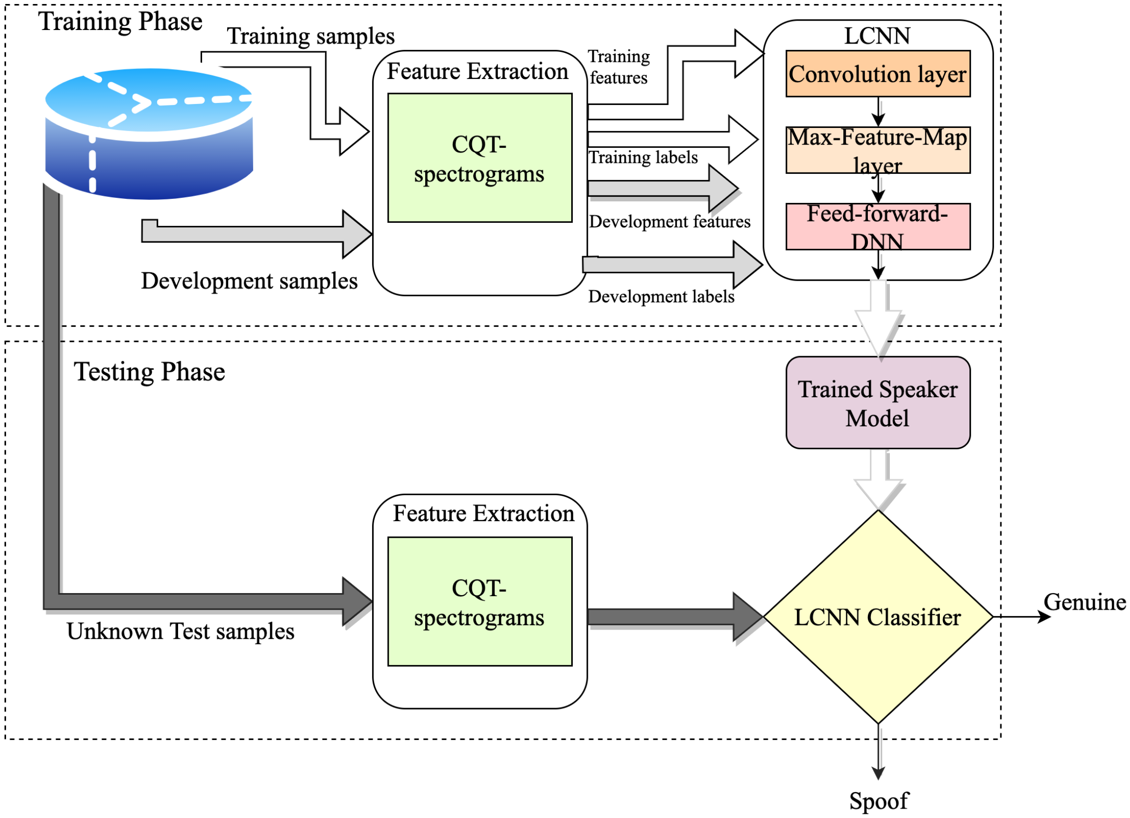


Fig. 1. Proposed Spoof Detection System.

The CQT based features, $C(i, m)$ are computed as shown in Eq. 1.

$$C(i,m) = \sum_{k=m-[M_i/2]}^{k=m+[M_i/2]} s(k) z_i^*(k - m + M_i/2) \qquad (1)$$

Where, $i = 1$ to I, is the index of frequency bins, M_i is window length which is a variable and z_i^* is complex conjugate of basis. Hence, at lower frequencies, a high resolution is obtained wrt frequencies and at higher frequencies, high temporal resolution is possible. Thus overcoming the shortcoming of STFT with fixed time-frequency resolutions [21], [23].

B. Back-end LCNN Classifier

The LCNN are popular due to their reduction in network parameters with nearly similar error rates as the CNNs [19]. In this work, we have employed a compact version of LCNN structure [21] using the Maximum Feature Mapping activation (MFM) layer. It is based on the Max-out activation which has the ability to choose the right features for problem solving purposes. The combination of MFM and multiple Batch Normalization (BN) layers form a LCNN structure with dense layer at end that wraps up the overall output from the previous layers. Also, after alternate MFM layer, a max pooling layer is added which picks out max value out a patch of feature map rather than input feature map. A more detailed information can be found in Section IV.

IV. LCNN Structure and Hyperparameter Tuning

The conventional CNN uses activation function in the convolution layer, typically ReLU [15]. The CNN with MFM activation triggers two neurons and ignores one (in case of 2/1 MFM). This is termed as competitive relationship; hence MFM acts as a fine feature selection algorithm embed inside a CNN. The LCNN network used in this work has nine MFM-convolution, 4 max Pooling, 7 BN layers and 2 Fully Connected (FC) layers as shown in Fig. 2.

The BN layer is appended after every convolution layer as it leads to faster convergence and improved accuracy. For an input convolution layer, $v^k \in R^{h \times w}$, where $k = \{1, 2, \ldots, 2K\}$, w is spatial width while h is spatial height. The MFM activation is given as shown in Eq. 2.

$$\hat{v}_{a,b}^m = \max(v_{a,b}^m,\ v_{a,b}^{m+K}) \qquad (2)$$

Where $2K$ is number of channels specific to input layer, m, a and b are indices for channel, width and height respectively. Therefore, the output dimension is $R^{h \times w \times K}$ and the gradients are calculated as shown in Eq. 3 and Eq. 4.

$$\frac{\delta \hat{v}_{a,b}^m}{\delta v_{a,b}^m} = \begin{cases} 1, & if \ \ v_{a,b}^m \geq v_{a,b}^{m+K} \\ 0, & elsewhere \end{cases} \qquad (3)$$

$$\frac{\delta \hat{v}_{a,b}^m}{\delta v_{a,b}^{m+K}} = \begin{cases} 0, & if \ \ v_{a,b}^m \geq v_{a,b}^{m+K} \\ 1, & elsewhere \end{cases} \qquad (4)$$

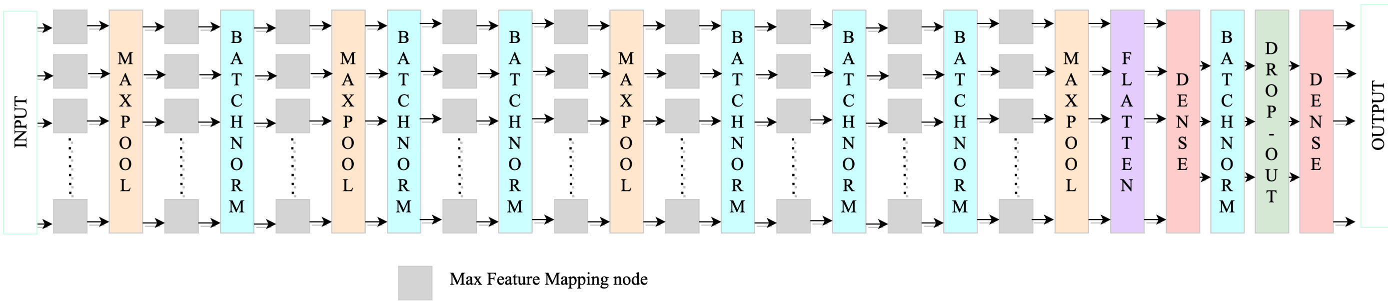


Fig. 2. Light Convolutional Neural Network Architecture for Spoof Detection System.

TABLE I. Various Optimizers with the Update Rule

Optimizer	Update Rule	Description
Adam	$\Theta_{s+1} = \Theta_s - \frac{\alpha}{\sqrt{\hat{v}_s}}\hat{m}_s$	s = time step, α = learning parameter, Θ_s = model parameters, m_s and v_s are decaying average gradients of past gradients and square of the gradients respectively.
Adadelta	$\Theta_{s+1} = \Theta_s + \triangle\Theta$ $\triangle\Theta = \frac{RMS[\triangle\Theta]_{s-1}}{RMS[g]_s} g_s$	g_s = gradient of objective function
Adagrad	$g_{s,i} = \triangledown_\Theta J(\Theta_{s,i})$ $\Theta_{s+1,i} = \Theta_{s,i} - \alpha g_{s,i}$	$J(\Theta_{s,i})$ = objective function i =. individual parameter
SGD	$\Theta = \Theta - \alpha \triangledown_\Theta J(\Theta; T^i; O^i)$	T^i = i^{th} training sample O^i = i^{th} label
SGD with momentum	$\Theta = \Theta - v_s$ $v_s = \sigma v_{s-1} + \alpha \triangledown_\Theta J(\Theta)$	σ =. momentum factor
SGD with NAG	$\Theta = \Theta - v_s$ $v_s = \sigma v_{s-1} + \alpha \triangledown_\Theta J(\Theta - \sigma v_{s-1})$	v_{s-1} = square of the previous decaying gradient
Nadam	$\Theta_{s+1} = \Theta_s - m_s$ $m_s = \sigma v_{s-1} + \alpha g_{s,i}$	α = learning rate Θ_{s+1} = future model parameter
RMSprop	$\Theta_{s+1} = \Theta_s - \frac{\alpha}{\sqrt{E(g^2)_s + \rho}} g_s$	$E(g^2)_s$ = running average at time step s, ρ = smoothing term

TABLE II. Details of ASV Spoof 2019 Dataset

	Logical Access		**Physical Access**	
	Genuine	**Spoofed**	**Genuine**	**Spoofed**
Training Set	2580	22800	5,400	48,600
Development Set	2548	22296	5,400	24,300
Total	5128	45096	10800	72900
	50224		83500	
	133724 Samples			

Apparently, half the information bearing neurons are acquired by $2/1$ MFM activation. Thus, implying 50% reduction in comparison to conventional CNN architecture. This is due to the element-wise maximum computation for all the feature channels. Hence, leading to sparser connections.

Additionally, this work experiments with different renowned optimizers to calculate the loss function including SGD, with momentum and NAG, Adaptive gradient techniques such as Adam, Adadelta, Adagrad and RMSprop optimizers [24]. Furthermore, we experimented by changing the activation functions from state-of the-art softmax-Adam optimizer [20] based LCNN architecture to ReLU and logistic activations. The aim to try out various optimizers and activations is to investigate the appropriate combination of individual optimization algorithms with respective activation functions. Many parameters have different working scenarios to perform best and this gives the reason to explore other optimizers and activations specific to spoof detection scenario.

A. Gradient Optimization Algorithms

The optimization of hyper-parameters is an essential step in training any Deep Learning framework. In this work, we have tested various gradient optimizers for overcoming the challenges of tuning learning rate, slow convergence, over-fitting of the model and lower accuracy. The SGD algorithms are derived from Gradient descent optimizers with noisy stochastic convergence at each iteration for a particular sample [25]. This implies that it can capture generality without the network to complete the training on the entire training set. On the other hand, the SGD algorithms might experience overshooting due to improper choice of learning rate. A small value of learning rate leads to slow convergence while a big value might lead to no convergence at all. To overcome this issue, momentum, NAG and adaptive optimizers are investigated. The momentum increases the speed of convergence towards steeper direction as against less steeper ones. The typical value of momentum is 0.9 [25]. Additionally, the NAG with momentum stores future gradients to speed-up the convergence by improving the learning rate to higher or lower values accordingly.

Apart from momentum and NAG, a vivid way of improving the performance of gradient optimizers is through adaptive gradient techniques. The AdaGrad [26] is one such optimizer that makes larger updates for not so frequent parameters while small updates for frequent ones. This also leads to accumulation of past gradients ultimately leading to a zero learning rate. In contrast to Adagrad, the Adadelta [26] uses a fixed

TABLE III. EXPERIMENTAL RESULTS FOR LA AND PA ATTACK VARIOUS OPTIMIZERS AND ACTIVATION FUNCTIONS

Type of System	Type of Attack		PA			LA		
	Activation	Optimizer	Epochs	Learning rate	EER	Epochs	Learning rate	EER
Baseline	**Softmax**	**Adam**	100	0.00001	11.949	50	0.00001	11.282
Proposed	**Softmax**	**Adadelta**	100	0.001	15.989	100	0.001	15.433
		Adagrad	100	0.001	11.022	50	0.001	11.559
		SGD	100	0.001	21.438	100	0.001	15.687
		SGD momentum	100	0.001	12.844	100	0.001	**9.055**
		SGD-nesterov	100	0.001	**9.951**	100	0.0001	10.671
		Nadam	100	0.00001	12.091	50	0.0001	10.312
		RMSprop	100	0.00001	11.834	100	0.00001	11.387
	ReLU	**Adam**	100	0.00001	53.87	50	0.00001	43.303
		Adadelta	100	0.001	38.978	100	0.001	45.721
		Adagrad	100	0.001	48.814	100	0.001	43.981
		SGD	100	0.0001	48.385	100	0.001	58.71
		SGD momentum	100	0.0001	59.073	100	0.001	50
		SGD-nesterov	100	0.0001	56.507	100	0.001	50
		Nadam	100	0.0001	44.681	100	0.0001	49.986
		RMSprop	100	0.00001	48.033	50	0.00001	54.692
	Sigmoid	**Adam**	100	0.00001	12.245	100	0.00001	**10.015**
		Adadelta	100	0.001	18.844	100	0.001	21.39
		Adagrad	100	0.001	15.85	100	0.001	13.984
		SGD	100	0.0001	30.07	100	0.001	14.379
		SGD momentum	100	0.0001	11.386	100	0.001	13.382
		SGD-nesterov	100	0.0001	12.27	100	0.001	10.535
		Nadam	100	0.0001	15.779	100	0.0001	18.275
		RMSprop	100	0.00001	**11.317**	100	0.00001	19.893

window to refrain from past gradient accumulation. Similarly, the RMSprop tries to fix the past gradient issue by averaging the square of the gradients. Furthermore, the Adam optimizer estimates learning rate for every parameter value. It is a fusion of RMSprop with momentum. Additionally, the amalgam of Adam with NAG is Nadam optimizer [25]. Further, we have applied an Early stopping condition by tracking the validation error with some patience to see if it is experiencing any changes; if not then, training is halted. Table I summarizes the update rule for all the discussed optimizers.

B. Activation Functions

The basis of any neural network to function the intended way is through activation. The activation functions lead the input to the output that speeds-up the training for capturing complex nature of the patterns within the input data. Usually the softmax and arg-softmax activations have been used in the LCNN architectures [21]. In this work, we have considered the combination of softmax, ReLU and logistic activation function to observe the loss characteristics with the chosen optimizers. The ReLU activation has been a popular choice amongst larger CNN as it overcomes the issue of vanishing gradients but at the same time experiences the dead neuron issue. While the logistic function is useful for binary classification tasks. The softmax is an extension of logistic activation as it works for multi-class problem [26].

V. EXPERIMENTAL RESULTS AND DISCUSSION

The performance of these optimizers is measured using the Equal Error Rate (EER) [27] and the ROC curve is plotted to check goodness of the classification algorithm signifying the Area Under Curve (AUC) value [28]. The EER must be ideally as low as possible. The system is trained and evaluated using the ASV spoof 2019 dataset that has synthetic (TTS and VC) and replay speech samples along with genuine speaker samples. There are 20 speakers (male and 12 female) including more than one hundred thousand samples with LA and PA attacks. This is currently the only large scale dataset with all three attack types and genuine samples. Also, in this work the development dataset is used for validation purposes. The test data or evaluation data has a lot of variation in environment conditions for replay speech and synthesizers of synthetic speech. Thus ensuring an unbiased testing scenario. Table II shows the details of ASV spoof 2019 dataset. Table III shows the EER along for various optimizers and activation functions for both the attacks while Fig. 3 and Fig. 4 portrays the ROC curve for individual activations and optimizers for LA and PA, respectively.

From Table III, the efficiency of sigmoid and softmax are similar in contrast to ReLU where it fails to capture generality in both the kind of attacks. The EER for ReLU is almost 50 for most of the optimizers implying that gradient is stuck in local minima rather than global minima. Additionally, the improper scaling of weights in ReLU function leads to loss of actual data being considered. The sigmoid and softmax are related as the latter is just an extension for multi-class problems. This results in similar performance by both the activations. Further, the spoofed samples have multiple types of attack generated from various VC and TTS sources. Hence, the softmax gives slightly better EER of 9.005 for LA and 9.951 for PA attack. Infact, the softmax is observed to converge faster with as low

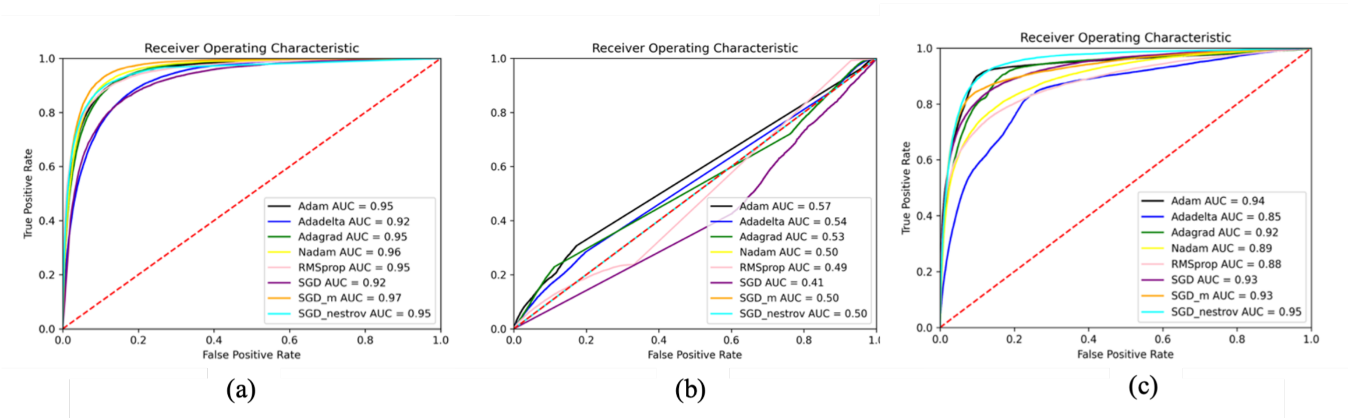


Fig. 3. ROC Curve for Various Activations and Optimizers for LA Attack (a) Softmax (b) ReLU (c) Sigmoid.

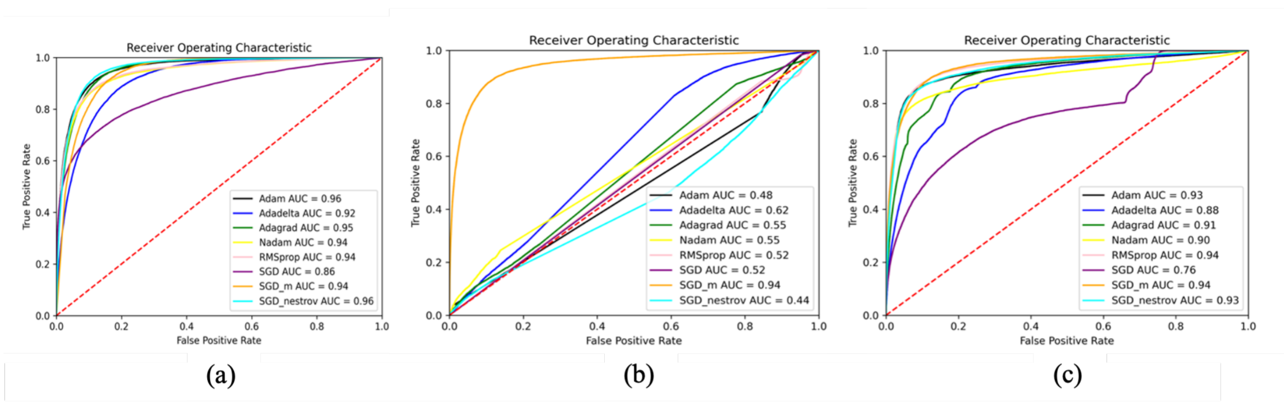


Fig. 4. ROC Curve for Various Activations and Optimizers for PA Attack (a) Softmax (b) ReLU (c) Sigmoid.

as 50 epochs for Adam and Adagrad for LA attack only.

In case of optimizer efficiency, Adam, RMSprop, SGD with momentum and NAG lead in EER for all the activations. The test condition comprise of noise in the dataset, hence RMSProp is the obvious performer for sigmoid activation with EER of 11.317. The SGD optimizer performs inconsistently with large variations in EER ranging from 14.379 to 30.070. This is due to lack of convergence and difficulty in adapting to convex problems. Thus as oppose to the SGD optimizer, the SGD with momentum and NAG are found to have a lower EER for both attacks. Hence, they are suitable for capturing generality like in the spoof detection task. The Adam optimizer performs consistently well with EER ranging from 10.015 to 11.949. So, it may be explored where generality is not of critical importance. Adadelta and Adagrad are not the shining performers but Adagrad gives a 0.3% improvement in EER than Adadelta; yet they perform poorly in comparison to Adam. The Nadam performs well for softmax optimizer while its performance worsens with increase in EER for sigmoid activation. The overall choice of activation will be softmax with any optimizer from the ones leading. Also, the EER for LA attack is lower than PA attack. Thus, the network efficiency is explicitly achieved for LA attack.

To confirm the performance of various optimizers the Receiver Operating Characteristics (ROC) curve with AUC are shown in Fig. 3 and Fig. 4 for LA and PA attack respectively. The required value of AUC is between 0 and 1 with values closer to 1 implying a good classifier. The Fig. 3(a) shows ROC for softmax function where all the optimizers perform well. The SGD with mometum has exceptional AUC of 0.97. In Fig. 3(b), none of the optimizers are able to form a learning rule in case of ReLU activation implying the the ReLU classifiers are not suitable for spoof detection task. The Fig. 3(c) confirms that the Adam and SGD with NAG have same AUC of 0.95 which is best amongst the other optimizers for sigmoid activation. Simiarly for PA attack, from Fig. 4(a) the ROC for softmax function shows all the optimizers perform well except SGD which has AUC of 0.86, while in Fig. 4(b), no significant efficiency is observed for ReLU activation. Lastly, Fig. 4(c), in case of sigmoid activation, the RMSprop and SGD momentum have same AUC of 0.94 which are better amongst the rest of the optimizers.

VI. Conclusion

The goal of conducting this study was proving that initialization of the network prior to training and tuning of parameters during the training improves the network accuracy. Thus in this work, a comprehensive comparison of various optimizers was carried out on LA and PA attack data. The rationale for conducting such a study was to signify the role of optimizers in classifying the test samples accurately.

Moreover, the activation functions were also considered in this comparative work to highlight their role based on nature of input—output data. The softmax and sigmoid prove to be better as against the ReLU function in the LA attack. Also, the networks converged faster with less number of epochs for Adam optimizers. In case of PA attack, the softmax function performed not so well and so did the ReLU function; while sigmoid showed significant improvement in accuracy in comparison to the other two. Further, it was evidently found that the RMSprop performed consistently well amongst all the others;while the SGD with momentum performed better than SGD but not so well against SGD with NAG. On the whole, the choice of optimizer, learning rate and activation affect the accuracy of the training network and thus the overall performance of the spoof detection system. In future, this work may be extended to experimenting with more activations like leaky-ReLU, Exponential linear unit and parametric ReLU; while optimizers such as AMSGrad may be explored to solve the issues of current adaptive algorithms.

ACKNOWLEDGMENT

The authors are grateful for research support from School of Computer Science and funding from Taylor's University, Malaysia.

REFERENCES

[1] H. Zeinali, H. Sameti, and L. Burget, "HMM-Based Phrase-Independent i-Vector Extractor for Text-Dependent Speaker Verification," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 7, pp. 1421–1435, 2017.

[2] P. Korshunov and S. Marcel, "A Cross-Database Study of Voice Presentation Attack Detection," in *Handbook of Biometric Anti-Spoofing*, 2019, pp. 363–389.

[3] T. Kinnunen, M. Sahidullah, H. Delgado, M. Todisco, N. Evans, J. Yamagishi, and K. A. Lee, "The ASVspoof 2017 Challenge: Assessing the Limits of Replay Spoofing Attack Detection," in *INTERSPEECH*. ISCA, August 2017, pp. 2–6.

[4] Z. Wu, A. Khodabakhsh, C. Demiroglu, J. Yamagishi, D. Saito, T. Toda, and S. King, "SAS: A speaker verification spoofing database containing diverse attacks," in *International Conference on Acoustics, Speech and Signal Processing - ICASSP*. IEEE, 2015.

[5] J. Yamagishi, M. Todisco, M. Sahidullah, H. Delgado, X. Wang, N. Evans, T. Kinnunen, K. Aik Lee, V. Vestman, and A. Nautsch, "ASVspoof 2019: Automatic Speaker Verification Spoofing and Countermeasures Challenge Evaluation Plan," in *ASV Spoof 2019 Challenge*, 2019.

[6] M. Todisco, X. Wang, V. Vestman, M. Sahidullah, H. Delgado, A. Nautsch, J. Yamagishi, N. Evans, T. Kinnunen, and K. A. Lee, "ASVSpoof 2019: Future horizons in spoofed and fake audio detection," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2019, pp. 1008–1012.

[7] B. Nasersharif and M. Yazdani, "Evolutionary fusion of classifiers trained on linear prediction based features for replay attack detection," *Expert Systems*, vol. 38, no. 3, 2021.

[8] M. Singh and D. Pati, "Usefulness of linear prediction residual for replay attack detection," *AEU - International Journal of Electronics and Communications*, vol. 110, p. 152837, 2019.

[9] Y. Zhang, F. Jiang, and Z. Duan, "One-class Learning Towards Synthetic Voice Spoofing Detection," *IEEE Signal Processing Letters*, vol. 28, pp. 937 – 941, 2020.

[10] M. Alzantot, Z. Wang, and M. B. Srivastava, "Deep Residual Neural Networks for Audio Spoofing Detection," in *Annual Conference of the International Speech Communication Association, INTERSPEECH*. ISCA, September 2019, pp. 1078–1082.

[11] J. Yang and R. K. Das, "Improving anti-spoofing with octave spectrum and short-term spectral statistics information," *Applied Acoustics*, vol. 157, p. 107017, 2020.

[12] S. Duraibi, W. Alhamdani, and F. T. Sheldon, "Voice Feature Learning using Convolutional Neural Networks Designed to Avoid Replay Attacks," in *IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, December 2020, pp. 1845–1851.

[13] X. Xiao, X. Tian, S. Du, H. Xu, E. S. Chng, and H. Li, "Spoofing speech detection using high dimensional magnitude and phase features: The NTU approach for ASVspoof 2015 challenge," in *International Speech Communication Association, INTERSPEECH*. ISCA, 2015.

[14] B. Chettri, D. Stoller, V. Morfi, M. A. M. Ramirez, E. Benetos, and B. L. Sturm, "Ensemble Models for Spoofing Detection in Automatic Speaker Verification," in *Annual Conference of the International Speech Communication Association, INTERSPEECH*. ISCA, September 2019, pp. 1018–1022.

[15] M. Dua, C. Jain, and S. Kumar, "LSTM and CNN based ensemble approach for spoof detection task in automatic speaker verification systems," *Journal of Ambient Intelligence and Humanized Computing 2021*, pp. 1–16, 2021.

[16] F. Ye and J. Yang, "A Deep Neural Network Model for Speaker Identification," *Applied Sciences*, vol. 11, no. 8, pp. 1–18, 2021.

[17] F. Fang, J. Yamagishi, I. Echizen, M. Sahidullah, and T. Kinnunen, "Transforming acoustic characteristics to deceive playback spoofing countermeasures of speaker verification systems," in *International Workshop on Information Forensics and Security (WIFS)*. IEEE, December 2018, pp. 1–9.

[18] C. Zhang, C. Yu, and J. H. L. Hansen, "An Investigation of Deep-Learning Frameworks for Speaker Verification Antispoofing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 4, pp. 684–694, 2017.

[19] M. Volkova, T. Andzhukaev, G. Lavrentyeva, S. Novoselov, and A. Kozlov, "Light CNN Architecture Enhancement for Different Types Spoofing Attack Detection," in *Lecture Notes in Computer Science*, 2019, vol. 11658, pp. 520–529.

[20] X. Li, N. Li, C. Weng, X. Liu, D. Su, D. Yu, and H. Meng, "Replay and Synthetic Speech Detection with Res2Net Architecture," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, June 2021, pp. 6354–6358.

[21] G. Lavrentyeva, S. Novoselov, A. Tseren, M. Volkova, A. Gorlanov, and A. Kozlov, "STC Antispoofing Systems for the ASVspoof2019 Challenge," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*. ISCA, September 2019, pp. 1033–1037.

[22] L. Liu and J. Yang, "Study on Feature Complementarity of Statistics, Energy, and Principal Information for Spoofing Detection," *IEEE Access*, vol. 8, pp. 141 170–141 181, 2020.

[23] M. Todisco, H. Delgado, and N. Evans, "A new feature for automatic speaker verification anti-spoofing: Constant Q cepstral coefficients," in *Odyssey 2016: Speaker and Language Recognition Workshop*, 2016.

[24] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, pp. 295–316, 2020.

[25] S. Ruder, "An overview of gradient descent optimization algorithms," *Lecture Notes in Computer Science*, vol. 11046, 2016.

[26] G. Habib and S. Qureshi, "Optimization and acceleration of convolutional neural networks: A survey," *Journal of King Saud University - Computer and Information Sciences*, 2020.

[27] Z. Chen, Z. Xie, W. Zhang, and X. Xu, "ResNet and Model Fusion for Automatic Spoofing Detection," in *Interspeech 2017*. ISCA, August 2017, pp. 102–106.

[28] I. Shahin, A. B. Nassif, N. Nemmour, A. Elnagar, A. Alhudhaif, and K. Polat, "Novel hybrid DNN approaches for speaker verification in emotional and stressful talking environments," *Neural Computing and Applications*, vol. 33, no. 23, pp. 16 033–16 055, 2021.

Hybrid Deep Learning Approach for Sentiment Classification of Malayalam Tweets

Soumya S
Department of Computer Applications
Cochin University of Science and Technology
Cochin, Kerala, India

Pramod K V
Department of Computer Applications
Cochin University of Science and Technology
Cochin, Kerala, India

***Abstract*—Social media content in regional languages is expanding from day to day. People use different social media platforms to express their suggestions and thoughts in their native languages. Sentiment Analysis (SA) is the known procedure for identifying the hidden sentiment present in the sentences for categorizing it as positive, negative, or neutral. The SA of Indian languages is challenging due to the unavailability of benchmark datasets and lexical resources. The analysis has been done using lexicon, Machine Learning (ML), and Deep Learning (DL) techniques. In this work, the baseline models and hybrid models of Deep Neural Network (DNN) architecture have been used for the classification of Malayalam tweets as positive, negative and neutral. Since, sentiment-tagged dataset for Malayalam is not readily available, the analysis has been done on the manually created dataset and translated Kaggle dataset. The hybrid models used in this study combine Convolutional Neural Networks (CNN) with variants of Recurrent Neural Networks (RNN). The RNN models are Long Short-Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM) and Gated Recurrent Unit (GRU). All these hybrid models improve the performance of Sentiment Classification (SC) compared to baseline models LSTM, Bi-LSTM and GRU.**

***Keywords*—Bi-LSTM; CNN; NLP; Malayalam; Twitter**

I. Introduction

Sentiment Analysis (SA) is one of the inevitable research domains in Natural Language Processing (NLP), ML and Linguistics. The public express their opinion about products, events, movies, political concerns, ideas, interests, and so on using various social media platforms like Facebook, Twitter, Blogs, and so forth. SA has a significant role in the automatic identification of sentiment hidden in the text [1]. SA classifies the text as positive, negative, or neutral based on the sentiment. SA has been done at different levels like sentence, document, and aspect levels. Since the Tweets are short messages with 280 characters long, sentence-level analysis is most suitable for SA of Tweets.

In the southern state of India, Malayalam is a prominent language spoken by Keralites and also in union territories like Lakshadweep and Puducherry. Most people in Kerala prefer Malayalam to express their opinion, ideas, comments, etc. The majority of youth in Kerala are using Twitter for expressing opinions. Thus the number of tweets is tremendously increasing over time. The Government has initiated the automation of regional languages such as automatic speech recognition, language translation, and character recognition to support the public. The policies made by Government can be modified or changed based on the opinion expressed by the people through social media platforms. Hence, the automation of SA of Malayalam is essential. SA of Indian languages has initiated in the year 2010. The first work in Indian languages reported for the Hindi language [3]. Research work on SA of Malayalam is at the beginning stage due to the lack of benchmark datasets and lexical resources. So far, only few works have been done on Malayalam based on ML, DL, and lexicon-based approaches [29]-[36]. Further important applications are in the field of business analytics, movie reviews, stock market prediction, and so forth.

Hinton et al. proposed different deep learning architectures [2]. The DL models are extensively applied in image processing, NLP, SA, and so on. The most generally used DL models are CNN, RNN, LSTM, and GRU.

The main objective of this study is to enhance the performance of the system and reduce computational costs. In order to achieve this, we extracted the best features by finetuning the hyperparameters and modeled the combined architecture of CNN and the varients of RNN. As a result, the sentiment classification accuracy is improved. In this paper, two different datasets have been analyzed using baseline models and hybrid models of DL architectures. The analysis shows that hybrid models performed better in SA of Malayalam tweets than baseline models.

Word embedding has a significant impact on text processing. Word embedding is the process of converting words or sentences into numerical vectors. These word vectors lie in different dimensions like 32, 64, 128, 300, etc. The words in the same context appear to lie near the same vector space. The word embedding is created either during the training process of embedding layers in the neural network or by using a pre-trained model. The available word embedding models are word2vec [4], BERT [5], fastText [6], etc. The pre-trained model available for the Malayalam language is fastText and BERT. Since the pre-trained model has not performed well in our dataset, the word vectors are formed during the training phase of the embedding layer. The contributions of this study are mentioned below:

- Two different datasets have been created for SA of Malayalam Tweets. Dataset I contains 6304 tweets, where 2907 are positive and 3397 are negative. The Dataset II includes 170000 tweets, among which 66357 are positive, 52798 are negative, and 50845 are neutral.

- Different DL methods like LSTM, Bi-LSTM and GRU have been applied for SA of tweets.
- Novel hybrid DL architectures have been developed by combining both CNN and varients of RNN models for the effective implementation of sentiment classification of Malayalam Tweets.

In the subsequent sections, Section II briefs the related works using hybrid deep learning models and SA of Malayalam language, whereas Section III is the proposed methodology of SA. Section IV describes the novel hybrid models. Section V is the experimental setup followed by results and discussion. Section VI concludes the work.

II. Related Work

The hybrid architecture combines different ML and DL algorithms for feature extraction and sentiment classification of datasets. The various works done using hybrid models are discussed as follows.

DL models are widely used in the analysis of social media content [7][8][9][10][11]. Hassan et al. [12] and Hedge et al. [13] proved that CNN and LSTM performed well in the analysis of short text messages . The recent study shows that the DL models such as CNN and RNN performed well in SA [7][8][14][15][16]. CNN combined with RNN shows potential improvement in the accuracy of SA of English [17]. Srinidhi et al. proposed the LSTM model combined with SVM for SC of IMDb dataset [18]. A similar architecture CNN combined with SVM was proposed by Akhtar et al. for SC of Hindi dataset [19]. SA on reviews/comments from e-commerce sites was proposed by Vo et al. using the hybrid architecture LSTM-CNN model [20]. The same model was proposed by Rehman et al. for SA of movie reviews [21]. All the above-mentioned works have been used single architecture for SC. Multiple DL models, CNN, LSTM and hybrid model CNN-LSTM was proposed by Kastrati et al. [22] . Facebook comments related to the COVID-19 pandemic were tested using this model. Pre-trained models like fastText and BERT were utilized for word embedding [23]. The same hybrid model was used for SC of IMDb dataset, social media content and SMS spam detection for Arabic and English Messages [24][25]. The CNN-LSTM with fastText word embedding was proposed by Ombabi et al. [26] . SVM was used at the final layer for SC. A hybrid model CNN and Bi-LSTM were proposed for sentiment and emotions analysis of Chinese product reviews [27]. Pandey et al. proposed a hybrid DL model merged with CNN and Bi-LSTM for SA of Tweets [10]. Dang et al. suggested multiple hybrid DL models with CNN, LSTM, and SVM for the classification of tweets and review datasets [14]. Salur et al. proposed hybrid DL architecture for SC of Turkish dataset [28]. Both CNN and LSTM models were combined for feature extraction and have obtained better accuracy compared to baseline models. The works are done in Malayalam language using rule-based, lexicon-based, Fuzzy logic, ML, and DL methods are shown in Table I. All the works have been done on the dataset which is created by corresponding authors on different domains like a movie review, novel, Tweets, etc. Preprocessing is an important step in SA, which depends on the domain of the dataset. All the works in Malayalam have been used as a single model for SC. Here, we have used hybrid architecture which extracts better features and has shown potential improvement in the SA tasks. The proposed hybrid DL models and their evaluation are explained in consecutive sections.

III. Proposed Methodology

In the previous works, lexicon-based, ML approaches like NB, SVM and RF, DL models like RNN, CNN, LSTM, BiLSTM and GRU have been applied for SA of Malayalam tweets [37][38][39]. Considering the significance of hybrid models in SA, three different hybrid architectures are developed in this work. As the first step of implementation, two different sentiment-tagged Malayalam datasets has been created. The pre-processing steps eliminate unnecessary information from the retrieved text. After that the feature vector is formed during the training phase of embedding layer. The feature vector is given to baseline as well as hybrid DL models. The sigmoid activation function for Dataset I and softmax for Dataset II are applied at the output layer in the baseline models and SVM is applied in the hybrid models. The proposed methodology for SC is shown in Fig. 1.

A. Dataset

Dataset I is created by retrieving Malayalam tweets using sentiment-oriented words [38]. Dataset II is created by translating the Kaggle dataset existing in English to Malayalam by using the Google document translator. The sample dataset is shown in Fig. 2. The sentiment distribution of the Dataset I and Dataset II are shown in Fig. 3 and Fig. 4, respectively. In Dataset I, 0 and 1 represents negative and positive sentiments, whereas, in the Dataset II, 0, 1, and 2 represent neutral, positive, and negative sentiment. The length of sentences under each category of Dataset I and Dataset II are shown in Fig. 5 and Fig. 6, respectively.

B. Preprocessing

The performance of the system has been improved by removing unnecessary information from the text. The following steps have been performed as part of data cleaning.

- Hyperlinks: Most of the tweets contain brief sentences followed by hyperlinks that do not provide meaningful information for SA. Hence all the hyperlinks have been removed from the corpus.
- Punctuations and Special Characters: Removed punctuations like :, , ',... etc. and special characters like @, $, #, etc. from the text.
- Stop Words: Stop words are often occurring in the input sentence, but are less information-oriented. Hence they are removed to reduce the vocabulary size.

Data cleaning has been done using regular expressions in Python language. Negative sentences are labeled with 0 and positive are labeled with 1. Preprocessing is an essential step in reducing the vocabulary size and removing unwanted information. Hyperlinks, special characters, punctuations, digits, foreign languages, etc. are removed by using the regular expression before extracting the features. Vocabulary is created by tokenizing the sentence based on Unicode standards within the range of 0D00–0D7F, which is shown in Fig. 7. Tokenized

TABLE I. SA OF MALAYALAM

Author	Domain with Size of Dataset	Preprocessing	Features	Classification Method	Accuracy
Mohandas et al. (2012) [29]	Novel	POS tagging	Tokens	SO-PMI-IR	63 %
Nair et al. (2014) [30]	Movie Review	Sandhi Splitter	Tokens, Negation	Rule Based	85 %
Anagha M et al.(2014) [31]	Movie Review	POS tagging	Wordnet	Lexicon	93.6 %
Anagha M et al. (2015) [32]	Movie Review	POS tagging	TnT tagger	Fuzzy Logic	91.6 %
Nair et al. (2015) [33]	Movie Review (30,000 tokens)	POS tagging	Tokens, Negation Intensifier	CRF, SVM	SVM: 91 %
Kumar et al. 2017 [34]	Tweets (12822)	Removing hyperlinks and punctuations	word embedding with different dimensions	LSTM, CNN	98.24 %
Rahul et al. (2018) [35]	Social Media	Removing hyperlinks and punctuations	POS tagging Positive and Negative Intensifier Negation	CRF, SVM	SVM: 52.75 %
Kumar et al. (2019) [36]	Tweets	Removing hyperlinks and punctuations	word embedding with different dimensions	RKS-RBF, LSTM CNN	86.5 % 89.3 %

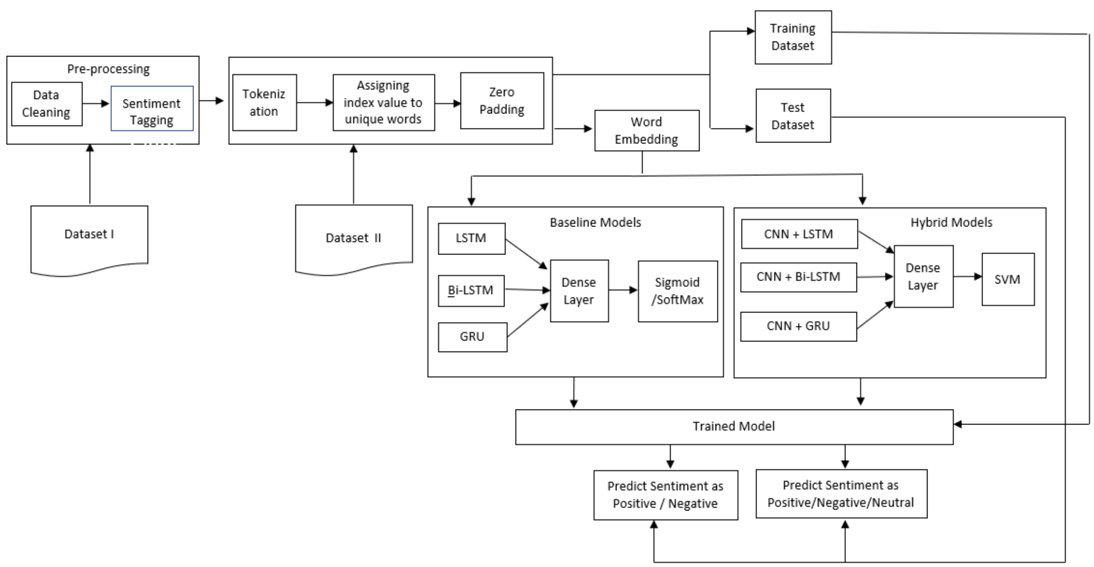


Fig. 1. Methodology for the Proposed System.

TABLE II. DATASET 1

Dataset I	
Sentiment	Number of Tweets
Positive	2907
Negative	3397
Total	6304

TABLE III. DATASET 2

Dataset II	
Sentiment	Number of Tweets
Positive	66357
Negative	52798
Neutral	50845
Total	170000

corpus is converted to a sequence of integers with variable length records and is represented in Fig. 8. The variable-length record is converted to fixed-length by zero-padding which is shown in Fig. 9. Word vector is formed during the training phase of embedding layer which takes embedding dimension, vocabulary size and maximum length of sentence as parameters. Embedding dimension takes different values like 64, 128 and 300, but it shows better performance when the dimension is set to 300. The 300 dimensional word vector is shown in Fig. 10.

IV. HYBRID ARCHITECTURE

In the related study, different hybrid models combined with CNN and LSTM show better results in SA of English languages. Here, we have used three different hybrid architectures for SA of Malayalam tweets. The word vector is created during the training phase of the embedding layer and then given to CNN and various RNN models. The features extracted by the CNN and RNN models are merged and corresponding output

Dataset I		
1	കലഹം നാലാംക്ലാസുകാരൻ മൂന്നാംക്ലാസുകാരനെ മണ്ണെണ്ണയൊഴിച്ച് തീകൊളുത്തി. (Riot A fourth-grader burned to death with kerosene)	0
2	അണലിയുടെ കടിയേറ്റ് ചികിത്സയിൽ കഴിയുന്ന വാവ സുരേഷിന്റെ ആരോഗ്യനില അതീവ ഗുരുതരം. (Vava Suresh is in serious condition after being bitten by a viper)	0
3	എഴുന്നേറ്റാലുടൻ ഒരുഗ്ലാസ് വെള്ളവും ഏറ്റവുമിഷ്ടമുള്ളൊരു പാട്ടും കേട്ടാൽ ശാരീരികവും മാനസികവുമായി ഉന്മേഷം ലഭിക്കും. (Getting up and drink a glass of water and listening to a favorite song will refresh you physically and mentally)	1
4	മൺസൂൺ നന്നായി പെയ്യുന്നു എന്നുള്ളത് വളരെ ശുഭകരമായ ഒരു വാർത്തയാണ് . (The good news is that the monsoon is getting better)	1
Dataset II		
1	ദൈവം നാളെ ഇന്ത്യയെ അനുഗ്രഹിക്കട്ടെ മോദി പ്രധാനമന്ത്രിയാകും (God bless India tomorrow Modi will be the Prime Minister)	0
2	മോഡിസ് ഇപ്പോൾ ഉപയോഗിക്കുന്ന അതേ വാക്കുകൾ ഇതിൽ നിന്ന് മാറ്റുന്നു (It replaces the same words that Modi now uses)	0
3	ആവേശമുണർത്തുന്ന ട്രെയിലർ ഗംഭീരമായ ശ്രമം കാത്തിരിക്കുന്നു നന്നായി കാണൂ കഠിനമായ ചോദ്യങ്ങൾ ചോദിക്കുന്നു ജനാധിപത്യം ശരിക്കും സംരക്ഷിച്ചു മോദി രാജ് എല്ലാ ആശംസകളും (Exciting trailer awaits great effort Watch well Ask tough questions Democracy is really protected Modi Raj All the best)	1
4	വിദേശ രാജ്യങ്ങളിലെ ഇന്ത്യൻ വംശജരെ മാന്യമായ ജീവിതം നയിക്കാൻ സഹായിച്ച മികച്ച നയതന്ത്ര മോദി അതാണ്. (Modi is the best diplomat who has helped Indians in foreign countries lead a dignified life.)	1
5	ജിഎസ്ടി പോലെ മോഡി കാലഘട്ടത്തിൽ ഉണ്ടാക്കിയ ബാങ്ക് അക്കൗണ്ടുകൾ നീക്കം ചെയ്യുന്നില്ല ആരാണ് ഈ വിഡ്ഢി പദ്ധതിക്ക് പണം നൽകുന്നത് . (Bank accounts created during the Modi era will not be removed like GST Who pays for this stupid scheme.)	2
6	വോട്ട് തേടുന്ന ഒരാളെ നിങ്ങൾ കാണുമ്പോഴെല്ലാം മോദിക്കെതിരെ അക്രമത്തിന് പ്രേരിപ്പിക്കുന്ന കുറ്റാരോപിതനായ ഗൗഡയ്ക്ക് ശക്തമായ അടി നൽകുക. (Whenever you see someone seeking votes, give a strong blow to Gowda, who is accused of inciting violence against Modi.)	2

Fig. 2. Sample Dataset.

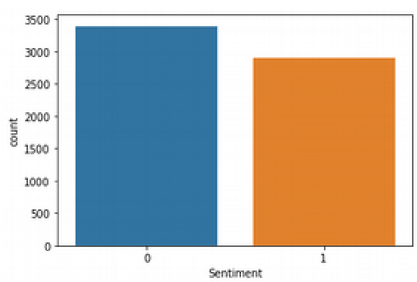


Fig. 3. Sentiment Distribution of Dataset I.

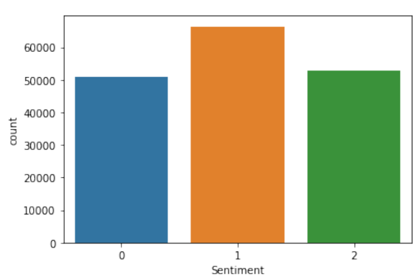


Fig. 4. Sentiment Distribution of Dataset II.

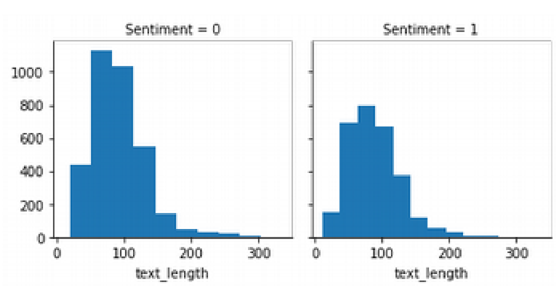


Fig. 5. Text Length of the Dataset I.

is given to a dense layer. Output of the dense layer is followed by a linear SVM. The three different hybrid models used in this study are as follows:

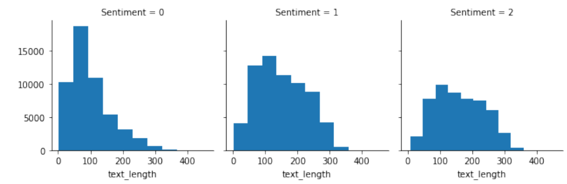


Fig. 6. Text Length of Dataset IIL.

```
['ഇന്നലെയും',
 'ഒരു',
 'ജീവൻ',
 'നഷ്ടപ്പെട്ടു',
 'കഴിഞ്ഞു',
 'ഈ',
 'നശിച്ച',
 'റോഡ്',
 'കാരണം',
 'മനഃപൂർവമായ',
 'നരഹത്യക്ക്',
 'കേസ്',
 'എടുക്കണം',
 'സർ',
 'ഈ',
 'പൈശാചികർക്കെതിരെ',
 '.'],

['വ്യാജരേഖകൾ',
 'ചമച്ച്',
 'സഹകരണ',
 'ബാങ്കുകളിൽ',
 'നിന്നും',
 'ഒരു',
 'കോടിയിലധികം',
 'രൂപ',
 'തട്ടിപ്പ്',
 'നടത്തി',
 '.'],
```

Fig. 7. Tokenized Corpus.

```
[5869, 5870, 5871, 5126, 5712, 5015, 2943, 13],
[4620, 137, 5872, 2943, 5873, 680, 5874, 13],
[5875, 5876, 5877, 5878, 5879, 5880, 5881, 13],
[5882, 5883, 5884, 2943, 5885, 5042, 1974, 5830, 5886, 13],
[5310, 140, 5887, 2943, 5888, 5267, 5268, 5889, 5890, 13],
[640, 644, 414, 5874, 674, 414, 5891, 2350, 5668, 2943, 5892, 5893, 13],
```

Fig. 8. Index Value Assigned to Each Word in a Sentence.

```
array([[    0,     0,     0, ...,    11,    12,    13],
       [    0,     0,     0, ...,    25,    26,    13],
       [    0,     0,     0, ...,    28,    29,    13],
       ...,
       [    0,     0,     0, ...,  1266, 20118,    13],
       [    0,     0,     0, ...,   165,  3397,    13],
       [    0,     0,     0, ...,  1703,  2287,    13]], dtype=int32)
```

Fig. 9. Sequence after Zero Padding.

- Word Vector – > CNN + LSTM – > Dense layer – > SVM
- Word Vector – > CNN + Bi-LSTM – > Dense layer – > SVM
- Word Vector – > CNN + GRU – > Dense layer – > SVM

Feature extraction is done using the convolution layer of CNN whereas feature reduction is done using the max-pooling

```
array([[ 0.36599999, 0.058 , -0.37479998, ..., -0.3773 ,
        -0.44199998, 0.1278 ],
       [-0.01850001, 0.10969999, 0.0535 , ..., -0.1031 ,
        -0.32610001, -0.29970001],
       [ 0.77069998, 0.5072 , -0.54899997, ..., -1.50040001,
        -0.66089999, -0.4816 ],
       ...,
       [ 0.3575 , 0.2539 , -0.79189999, ..., 0.06089999,
        -0.2482 , 0.2326 ],
       [ 0.67430001, 0.0322 , 0.27570002, ..., -0.6824 ,
        -0.4373 , 0.3796 ],
       [ 0.46149999, 0.0568 , -0.12809999, ..., -1.1345 ,
        -0.7934 , -0.0933 ]])
```

Fig. 10. Word Embedding.

layer [9]. The convolutional layer uses filters for extracting important attributes from the data. Different filters are utilized for various applications with distinct kernel sizes. Kernel size represents the n-gram representation. ReLU is the commonly used activation function in convolutional layers. The pooling layer consolidates the output from the convolutional layer by selecting optimal data from the previous layer. Thus, reducing the dimension of feature vectors. The final layer of CNN is a fully connected neural network. LSTM consists of different elements, including an input gate, forget gate, memory cell, hidden state and an output gate [40]. LSTM removes long-term dependency but keeps some useful information. Bi-LSTM [41] creates the exact copy of LSTM in a backward direction also. The output of both forward and backward LSTM hidden states are combined at each step. GRU [42] is the simplified model of LSTM by removing the output gate. Thus, it reduces the complexity involved in LSTM architecture.

A. Evaluation Measures

The model is evaluated for making standard metrics like Precision, Recall, F1-score, and Accuracy using the confusion matrix [43]. The evaluation measures are formulated using the following equations.

Precision (positive classification) = $\frac{TP}{TP+FP}$
Precision (negative classification) = $\frac{TN}{TN+FN}$
Recall (positive classification) = $\frac{TP}{TP+FN}$
Recall (negative classification) = $\frac{TN}{TN+FP}$

Where, TP represents true positive, TN: true negative, FP: false positive, and FN: false negative.

F1-score = $\frac{2*precision*recall}{precision+recall}$

Accuracy is measured based on how many sentences are correctly classified among total sentences.

Accuracy = $\frac{TN+TP}{TP+TN+FP+FN}$

B. Proposed Hybrid Architecture

The details of the proposed hybrid models are given as follows:

Algorithm 1: HYBRID MODEL OF CNN AND LSTM

Input: S // S is a corpus which consists a set of sentences $S_1, S_2, \dots S_n$
Output: Y // Labelled as 0 or 1 for Dataset I and 0, 1 or 2 for Dataset II
for each sentence S_i in S do
$W_i = Embedding(S_i)$ // Word vector in 300 dimensional space
end
for each W_i do
$C_i = CNN(W_i)$
$L_i = LSTM(W_i)$
end
for each C_i and L_i do
$O_i = Dense(C_i, L_i)$
end
for each O_i do
$Y_i = LinearSVM(O_i)$ // Linear SVM used as output layer activation function
end

Algorithm 2: HYBRID MODEL OF CNN AND BILSTM

Input: S // S is a corpus which consists a set of sentences $S_1, S_2, \dots S_n$
Output: Y // Labelled as 0 or 1 for Dataset I and 0, 1 or 2 for Dataset II
for each sentence S_i in S do
$W_i = Embedding(S_i)$ // Word vector in 300 dimensional space
end
for each W_i do
$C_i = CNN(W_i)$
$B_i = BiLSTM(W_i)$
end
for each C_i and G_i do
$O_i = Dense(C_i, B_i)$
end
for each O_i do
$Y_i = LinearSVM(O_i)$ // Linear SVM used as output layer activation function
end

1) Model 1: Model 1 combines both CNN and LSTM. The word vector is given to CNN and LSTM parallelly. Filter size and kernel size of CNN are 32 and 2 respectively. The LSTM has 50 neurons. The merged output is given to a dense layer with 20 neurons. Further, the output is given to the linear SVM. SVM classifies the Dataset I as positive or negative, whereas Dataset II as positive, negative or neutral. The psudo-code representation of Model I is shown in Algorithm 1.

2) Model II: The second hybrid architecture combines both CNN and Bi-LSTM model. All the other layers are the same as in Model I. The psudo-code representation of Model II is shown in Algorithm 2.

3) Model III: The third hybrid model combines both CNN and GRU. The merged output is given to the dense layer,

Algorithm 3: HYBRID MODEL OF CNN AND GRU

```
Input: S // S is a corpus which consists a set of
   sentences S_1, S_2, ... S_n
Output: Y // Labelled as 0 or 1 for Dataset I and 0, 1
   or 2 for Dataset II
for each sentence S_i in S do
W_i = Embedding(S_i) // Word vector in 300
   dimensional space
end
for each W_i do
C_i = CNN(W_i)
G_i = GRU(W_i)
end
for each C_i and G_i do
O_i = Dense(C_i, G_i)
end
for each O_i do
Y_i = LinearSVM(O_i) // Linear SVM used as output
   layer activation function
end
```

followed by linear SVM. The psudo-code representation of Model III is shown in Algorithm 3.

V. EXPERIMENTAL RESULTS AND DISCUSSION

The implementation of this work is carried out using Google Colab [46] with Keras [44] and TensorFlow [45] libraries. The three different baseline DNN models (LSTM, BiLSTM, GRU) and three hybrid models (CNN+LSTM, CNN+BiLSTM, CNN+GRU) are applied to two different datasets for SC. The dataset is split into 80:20 where 80% dataset is used for training while 20% is used for testing. For Dataset I, the vocabulary size is set to 8353 and for Dataset II, the vocabulary size is set to 58846 by removing the foreign words and least frequent words present in the text. The maximum sequence length of Dataset I is 32 and of Dataset II is 151. Table II represents the optimal value of hyperparameters chosen during the training phase of both baseline and hybrid models on Dataset I. The percentage of rightly classified and wrongly classified datasets are shown in Fig. 11 and Fig. 12. respectively. The confusion matrices of hybrid models like CNN+LSTM, CNN+Bi-LSTM, and CNN+GRU for Dataset I are shown in Fig. 11(a), Fig. 11(b), and Fig. 11(c), and for Dataset II is shown in Fig. 12(a), Fig. 12(b) and Fig. 12(c) respectively. Precision, Recall, F1-score and accuracy are used as evaluation measures. Table III represents the evaluation measures on Dataset I. Table IV represents the loss and accuracy of training, validation and test dataset. The optimal value of hyperparameters on Dataset II is depicted in Table V. Table VI is the evaluation measures on Dataset II. Table VII represents the loss and accuracy of training, validation and test dataset of Dataset II. Table IV and Table VII show that the CNN + GRU got better prediction accuracy of 87.23% for Dataset I and CNN + BiLSTM got an accuracy of 74% for Dataset II. The model architecture for CNN + GRU for Dataset I and CNN + BiLSTM for Dataset II are shown in Fig. 13 and Fig. 14, respectively. The bar chart shown in Fig. 15 and Fig. 16 compares the accuracy of baseline models with hybrid models for Dataset I and Dataset

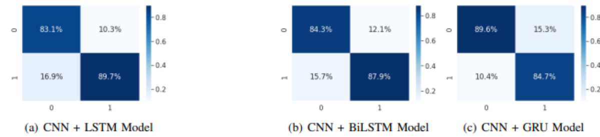


Fig. 11. Confusion Matrices of Hybrid Models on Dataset I.

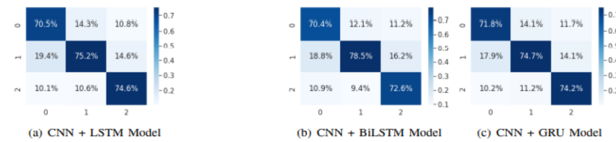


Fig. 12. Confusion Matrices of Hybrid Models on Dataset II.

II.

Discussion: Twitter is the most prominent platform for expressing opinions and suggestions about daily phenomena. Hence, tweets are considered for the SA. Both the baseline and hybrid DL models classified Dataset I as positive or negative and Dataset II as positive, negative and neutral. The best model is selected based on the accuracy of test dataset. The study of literature shows that the hybrid model performed well compared with baseline models for English and some Indian languages. This study analyzes that the hybrid models also show better prediction accuracy for the Malayalam language. Since the pre-trained vector fastText is not performed well on our datasets, the word embedding vector is created during the training phase of the embedding layer. After experimenting with various dimensions, the word vector is mapped to a 300-dimensional space for better prediction. Hybrid models improved the performance by nearly 2% to 3% compared with

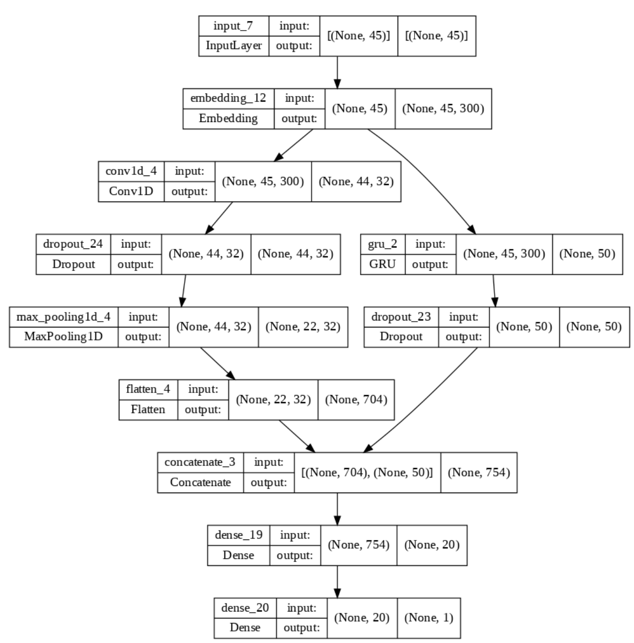


Fig. 13. Hybrid Architecture of CNN + GRU for Dataset I.

TABLE IV. VALUES OF HYPERPARAMETERS ON DATASET I

Model	Embedding Dimension	Number of Neurons in Dense Layer	Dropout	Optimizer	Kernel Size	Filter Size / Number of Neurons	Output Layer Activation Function	Loss Function
LSTM	300	20	0.4	Adam		50	sigmoid	binary_crossentropy
Bi-LSTM	300	20	0.4	Adam		50	sigmoid	binary_crossentropy
GRU	300	20	0.4	Adam		50	sigmoid	binary_crossentropy
CNN + LSTM	300	20	0.4	Adam	2	32, 50	linear SVM	squared_hinge
CNN + BiLSTM	300	20	0.4	Adam	2	32, 50	linear SVM	squared_hinge
CNN + GRU	300	20	0.4	Adam	2	32, 50	linear SVM	squared_hinge

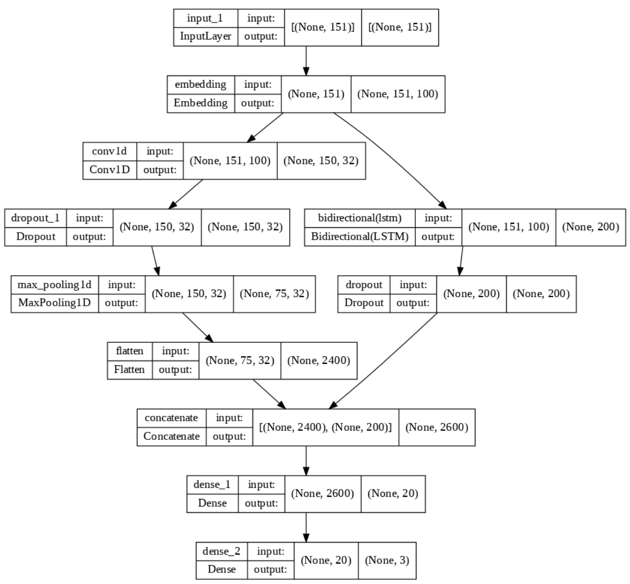


Fig. 14. Hybrid Architecture of CNN + BiLSTM for Dataset II.

TABLE V. EVALUATION MEASURES ON DATASET I

Model	Sentiment	Precision	Recall	F1-Score	Support
LSTM	Positive	0.84	0.78	0.81	581
	Negative	0.82	0.87	0.85	680
Bi-LSTM	Positive	0.80	0.87	0.83	581
	Negative	0.88	0.81	0.84	680
GRU	Positive	0.85	0.80	0.82	581
	Negative	0.83	0.88	0.86	680
CNN + LSTM	Positive	0.90	0.78	0.83	581
	Negative	0.83	0.92	0.87	680
CNN + Bi-LSTM	Positive	0.88	0.80	0.84	581
	Negative	0.84	0.91	0.87	680
CNN + GRU	Positive	0.85	0.88	0.86	581
	Negative	0.90	0.86	0.88	680

TABLE VI. TRAINING, VALIDATION AND TEST DATA ACCURACY OF DATASET I

Model	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy	Test Loss	Test Accuracy
LSTM	0.0892	0.9586	0.4433	0.8571	0.5120	0.831
Bi-LSTM	0.1359	0.9482	0.3740	0.8552	0.4126	0.838
GRU	0.2068	0.9167	0.3659	0.8373	0.3217	0.8410
CNN + LSTM	0.1078	0.9548	0.5915	0.8373	0.4169	0.8572
CNN + BiLSTM	0.1146	0.9473	0.5917	0.833	0.4288	0.858
CNN + GRU	0.1191	0.9486	0.4917	0.8472	0.4414	0.8723

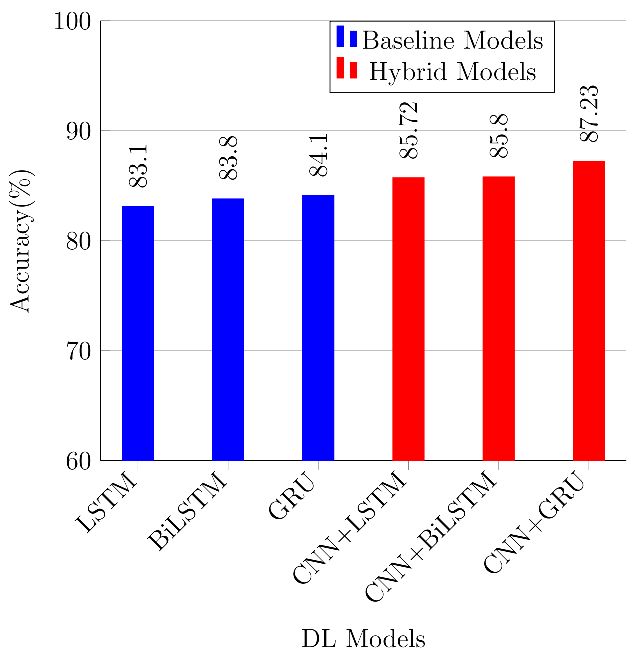


Fig. 15. Comparing Accuracy of DL Models on Dataset I.

baseline models for Dataset I and 1% to 2% for Dataset II. The combined architecture of DNN models extracts better features rather than a single model. CNN extracts the local features and variants of RNN extracts the long-term features from the text data. The hyperparameters like the number of neurons in LSTM, BiLSTM and GRU are set to 50 for Dataset I and 100 for Dataset II after several trials with varying neurons. The number of neurons at dense layer is selected as 20 for both datasets. The sigmoid activation function is used at the output layer for Dataset I and Softmax for Dataset II. But, for both datasets, LinearSVM is selected as the best activation function for hybrid models. The loss function binary_crossentropy, categorical_crossentropy and squared_hinge are applied for sigmoid, softmax and SVM activation functions. Adam optimizer is used for all the models. For small dataset, GRU performed well and for large dataset Bi-LSTM shows better accuracy in baseline models. Therefore, the combined architecture of CNN and GRU got better prediction accuracy for Dataset I and CNN + BiLSTM shows better accuracy for Dataset II.

VI. CONCLUSION

The expeditious growth of social media content in regional languages have led to the importance of SA in native languages. The advancement of ML and DL models has improved the performance of NLP applications. Since general people in Kerala use their native language to express their suggestions and opinions in social media platforms like Twitter, Facebook,

TABLE VII. VALUES OF HYPERPARAMETERS FOR DATASET II

Model	Embedding Dimension	Number of Neurons in Dense layer	Dropout	Optimizer	Kernel Size	Filter Size/ Number of Neurons	Output Layer Activation function	Loss function
LSTM	300	20	0.4	Adam		100	Softmax	categorical_crossentropy
Bi-LSTM	300	20	0.4	Adam		100	Softmax	categorical_crossentropy
GRU	300	20	0.4	Adam		100	Softmax	categorical_crossentropy
CNN + LSTM	300	20	0.4	Adam	2	32 , 100	linear SVM	squared_hinge
CNN + Bi-LSTM	300	20	0.4	Adam	2	32, 100	linear SVM	squared_hinge
CNN + GRU	300	20	0.4	Adam	2	32, 100	linear SVM	squared_hinge

TABLE VIII. EVALUATION MEASURES ON DATASET II

Model	Sentiment	Precision	Recall	F1-Score	Support
LSTM	Positive	0.75	0.70	0.73	13271
	Negative	0.69	0.79	0.74	10560
	Neutral	0.70	0.65	0.68	10169
Bi-LSTM	Positive	0.74	0.71	0.73	13271
	Negative	0.73	0.76	0.74	10560
	Neutral	0.69	0.70	0.69	10169
GRU	Positive	0.73	0.73	0.73	13271
	Negative	0.74	0.78	0.76	10560
	Neutral	0.71	0.67	0.69	10169
CNN + LSTM	Positive	0.75	0.73	0.74	13271
	Negative	0.75	0.77	0.76	10560
	Neutral	0.70	0.70	0.70	10169
CNN + Bi-LSTM	Positive	0.78	0.71	0.75	13271
	Negative	0.73	0.78	0.75	10560
	Neutral	0.70	0.73	0.72	10169
CNN + GRU	Positive	0.75	0.75	0.75	13271
	Negative	0.74	0.76	0.75	10560
	Neutral	0.72	0.69	0.70	10169

TABLE IX. TRAINING, VALIDATION AND TEST DATA ACCURACY OF DATASET II

Model	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy	Test Loss	Test Accuracy
LSTM	0.1243	95.76	0.7845	73.12	0.8245	71.42
BiLSTM	0.2698	89.56	0.8296	73.36	0.7982	72.74
GRU	0.2841	93.13	0.8123	72.47	0.8379	72.11
CNN + LSTM	0.1191	94.81	0.7565	73.51	0.7842	73.58
CNN + Bi-LSTM	0.2513	88.50	0.5437	74.38	0.6184	74.0
CNN + GRU	0.2132	91.40	0.6231	73.85	0.7123	73.72

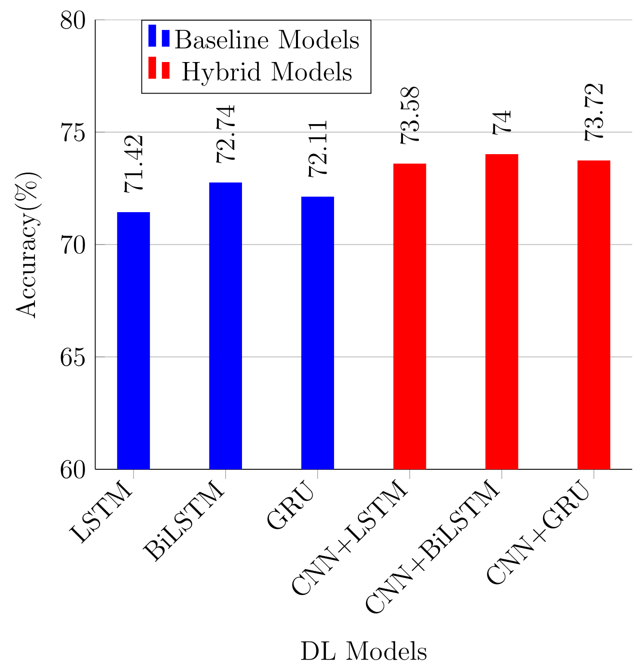


Fig. 16. Comparing Accuracy of DL Models on Dataset II.

etc. The automation of SA in Malayalam is essential for analysis. Also, DL models like LSTM, BiLSTM and GRU prevail over other methods of SC of text datasets. This paper proposes SA of Malayalam tweets using hybrid DNN models. Three different baseline DNN models, namely, LSTM, BiLSTM, GRU and further hybrid models combined with CNN and variants of RNN, including LSTM, Bi-LSTM and GRU, have been used for SC of tweets as positive, negative, or neutral. The experiments were conducted on two different datasets. Hybrid DL models performed well in both datasets. The combined architecture of CNN and the variants of RNN extract the best features compared to a single model. The word vector is formed during the training phase and mapped into a 300-dimensional vector space to achieve semantically similar words in the same space. Among the hybrid models, CNN + GRU shows the highest accuracy of 87.23 % for Dataset I and CNN + BiLSTM show better performance for Dataset II with an accuracy of 74%.

The major challenge was lack of benchmark datasets and lexical resources in Malayalam language. Hence, SA was done only on two different datasets, which is insufficient to standardize the results. Future work is aimed to develop benchmark datasets to achieve standardization of text analysis in Malayalam language.

REFERENCES

[1] Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up? Sentiment classification using machine learning techniques." arXiv preprint cs/0205070 2002.

[2] Hinton, Geoffrey E., Simon Osindero, and Yee-Whye Teh. "A fast learning algorithm for deep belief nets." Neural computation vol. 18, no. 7, pp:1527-1554, 2006.

[3] Joshi, Aditya, A. R. Balamurali, and Pushpak Bhattacharyya. "A fall-back strategy for sentiment analysis in hindi: a case study." Proceedings of the 8th ICON 2010.

[4] Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space." arXiv preprint arXiv:1301.3781 2013.

[5] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 2018.

[6] Grave, Edouard, et al. "Learning word vectors for 157 languages." arXiv preprint arXiv:1802.06893 2018.

[7] Abid, Fazeel, et al. "Sentiment analysis through recurrent variants latterly on convolutional neural network of Twitter." Future Generation Computer Systems Vol. 95, pp: 292-308, 2019.

[8] Alharbi, Ahmed Sulaiman M., and Elise de Doncker. "Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information." Cognitive Systems Research vol. 54, pp: 50-61, 2019.

[9] Khattak, Asad Masood, et al. “Tweets classification and sentiment analysis for personalized tweets recommendation.” Complexity 2020.

[10] Pandey, Avinash Chandra, Dharmveer Singh Rajpoot, and Mukesh Saraswat. “Twitter sentiment analysis using hybrid cuckoo search method.” Information Processing & Management vol. 53, no.4, pp: 764-779, 2017.

[11] Vateekul, Peerapon, and Thanabhat Koomsubha. “A study of sentiment analysis using deep learning techniques on Thai Twitter data.” 2016 13th international joint conference on computer science and software engineering (JCSSE). IEEE, 2016.

[12] Hassan, A., & Mahmood, A. “Deep learning approach for sentiment analysis of short texts.” In 2017 3rd international conference on control, automation and robotics (ICCAR) pp. 705-710, 2017.

[13] Hegde, Siddhanth U., A. S. Zaiba, and Y. Nagaraju. “Hybrid cnn-lstm model with glove word vector for sentiment analysis on football specific tweets.” 2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT). IEEE, 2021.

[14] Dang, Cach N., María N. Moreno-García, and Fernando De la Prieta. “Hybrid Deep Learning Models for Sentiment Analysis.” Complexity 2021.

[15] Martín, Carlos Alberto, et al. “Using deep learning to predict sentiments: case study in tourism.” Complexity 2018.

[16] Monika, R., S. Deivalakshmi, and B. Janet. “Sentiment analysis of US airlines tweets using LSTM/RNN.” 2019 IEEE 9th International Conference on Advanced Computing (IACC). IEEE, 2019.

[17] Chen, Tao, et al. “Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN.” Expert Systems with Applications vol. 72, pp: 221-230, 2017.

[18] Srinidhi, H., G. M. Siddesh, and K. G. Srinivasa. “A hybrid model using MaLSTM based on recurrent neural networks with support vector machines for sentiment analysis.” Engineering and Applied Science Research vol. 47, no. 3, pp: 232-240, 2020.

[19] Akhtar, Md Shad, et al. “A hybrid deep learning architecture for sentiment analysis.” Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. 2016.

[20] Vo, Quan-Hoang, et al. “Multi-channel LSTM-CNN model for Vietnamese sentiment analysis.” 2017 9th international conference on knowledge and systems engineering (KSE). IEEE, 2017.

[21] Rehman, Anwar Ur, et al. “A hybrid CNN-LSTM model for improving accuracy of movie reviews sentiment analysis.” Multimedia Tools and Applications vol. 78, no.18, pp: 26597-26613, 2019.

[22] Kastrati, Zenun, et al. ”A deep learning sentiment analyser for social media comments in low-resource languages.” Electronics vol. 10, no. 10, pp: 1133, 2021.

[23] Kaur, Harleen, et al. “A proposed sentiment analysis deep learning algorithm for analyzing COVID-19 tweets.” Information Systems Frontiers vol. 23, no. 6, pp: 1417-1429, 2021.

[24] Ghourabi, Abdallah, Mahmood A. Mahmood, and Qusay M. Alzubi. “A hybrid CNN-LSTM model for SMS spam detection in Arabic and english messages.” Future Internet vol. 12, no.9 pp: 156, 2020.

[25] Yenter, Alec, and Abhishek Verma. “Deep CNN-LSTM with combined kernels from multiple branches for IMDb review sentiment analysis.” 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON). IEEE, 2017.

[26] Ombabi, Abubakr H., Wael Ouarda, and Adel M. Alimi. “Deep learning CNN–LSTM framework for Arabic sentiment analysis using textual information shared in social networks.” Social Network Analysis and Mining vol. 10, no. 1, pp: 1-13, 2020.

[27] Zhou, Kai, and Fei Long. ”Sentiment analysis of text based on CNN and bi-directional LSTM model.” 2018 24th International Conference on Automation and Computing (ICAC). IEEE,2018.

[28] Salur, Mehmet Umut, and Ilhan Aydin. “A novel hybrid deep learning model for sentiment classification.” IEEE Access 8 pp: 58080-58093, 2020.

[29] Mohandas, Neethu, Janardhanan PS Nair, and V. Govindaru. “Domain specific sentence level mood extraction from Malayalam text.” 2012 International Conference on Advances in Computing and Communications. IEEE, 2012.

[30] Nair, Deepu S., et al. ”SentiMa-sentiment extraction for Malayalam.” 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE, 2014.

[31] Anagha, M., et al. ”Lexical resource based hybrid approach for cross domain sentiment analysis in Malayalam.” Int J Eng Sci 15 (2014): 18-21.

[32] Anagha, M., et al. ”Fuzzy logic based hybrid approach for sentiment analysisl of malayalam movie reviews.” 2015 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES). IEEE, 2015.

[33] Nair, Deepu S., et al. ”Sentiment Analysis of Malayalam film review using machine learning techniques.” 2015 international conference on advances in computing, communications and informatics (ICACCI). IEEE, 2015.

[34] Kumar, S. Sachin, M. Anand Kumar, and K. P. Soman. ”Sentiment analysis of tweets in malayalam using long short-term memory units and convolutional neural nets.” International Conference on Mining Intelligence and Knowledge Exploration. Springer, Cham, 2017.

[35] Rahul, M., R. R. Rajeev, and S. Shine. ”Social Media Sentiment Analysis For Malayalam.” International Journal of Computer Sciences and Engineering, E-ISSN (2018): 2347-2693.

[36] Kumar, S., M. Anand Kumar, and K. P. Soman. “Identifying Sentiment of Malayalam Tweets Using Deep Learning.” Digital Business. Springer, Cham, pp: 391-408, 2019.

[37] Soumya, S., and K. V. Pramod. ”Fine Grained Sentiment Analysis of Malayalam Tweets Using Lexicon Based and Machine Learning Based Approaches.” 2021 4th Biennial International Conference on Nascent Technologies in Engineering (ICNTE). IEEE, 2021.

[38] Soumya, S., and K. V. Pramod. “Sentiment analysis of Malayalam tweets using machine learning techniques.” ICT Express vol. 6, no. 4, pp: 300-305, 2020.

[39] Soumya, S., and K. V. Pramod. “Sentiment analysis of Malayalam tweets using different deep neural network models-case study.” 2019 9th International Conference on Advances in Computing and Communication (ICACC). IEEE, 2019.

[40] Hochreiter, Sepp, and Jürgen Schmidhuber. “Long short-term memory.” Neural computation vol. 9, no. 8, pp: 1735-1780, 1997.

[41] Graves, Alex, Navdeep Jaitly, and Abdel-rahman Mohamed. “Hybrid speech recognition with deep bidirectional LSTM.” 2013 IEEE workshop on automatic speech recognition and understanding. IEEE, 2013.

[42] Cho, Kyunghyun, et al. “Learning phrase representations using RNN encoder-decoder for statistical machine translation.” arXiv preprint arXiv:1406.1078 2014.

[43] Su, Louise T. “Evaluation measures for interactive information retrieval.” Information Processing & Management vol. 28, no.4, pp: 503-516, 1992.

[44] Francois, C. H. O. L. L. E. T. ”Keras: The Python deep learning library.” keras. io (2015).

[45] Abadi, Martín, et al. ”Tensorflow: Large-scale machine learning on heterogeneous distributed systems.” arXiv preprint arXiv:1603.04467 (2016).

[46] Google. Google Colab. Accessed: Jan. 10, 2020. [Online]. Available: https://colab.research.google.com

IoDEP: Towards an IoT-Data Analysis and Event Processing Architecture for Business Process Incident Management

Abir Ismaili-Alaoui
Université de Lorraine,
CNRS, Inria, LORIA, F-54000 Nancy, France
Rabat IT Center,
ENSIAS, Mohammed V University,
Rabat, Morocco

Karim Baina
Rabat IT Center, ENSIAS,
Mohammed V University,
Rabat, Morocco

Khalid Benali
Université de Lorraine,
CNRS, Inria, LORIA, F-54000 Nancy,
France

***Abstract*—IoT is becoming a hot spot area of technological innovations and economic development promises for many industries and services. This new paradigm shift affects all the enterprise architecture layers from infrastructure to business. Business Process Management (BPM) is a field among others that is affected by this new technology. To assist data and events explosion resulting, among others, from IoT, data analytic processes combined with event processing techniques, examine large data sets to uncover hidden patterns, unknown correlations between collected events, either at a very technical level (incident/anomaly detection, predictive maintenance) or at business level (customer preferences, market trends, revenue opportunities) to provide improved operational efficiency, better customer service and competitive advantages over rival organizations. In order to capitalize the business value of data and events generated by IoT sensors, IoT, Data Analytics and BPM need to meet in the middle. In this paper, we propose an end-to-end IoT-BPM integration architecture (IoDEP: IoT-Data-Event-Process) for a proactive business process incident management. A case study is presented and the obtained results from our experimentations demonstrate the benefit of our approach and allowed us to confirm the efficiency of our assumptions.**

Keywords*—Business process management; internet of things; machine learning; complex event processing; data analytics

I. Introduction

Nowadays, Business Process Management (BPM) is a well-established discipline in both academia and industry. It is considered as a powerful solution that helps organisations adapt to strategic, tactical and operational changes and gain more visibility and control over their business processes, so that they can continuously improve and optimise their activities and resources. Organizations use Business Process Management systems as an activity-based workflow manager that allow them to track the optimized functioning of their activities in order to gain in terms of agility, efficiency and performance. This method is mainly based on the concept of business processes. A business process is the structure of activities and actions as they occur in the real world. It defines all the possible paths in the real process and the rules that determine the path to follow and the actions to perform.

In most cases, business processes are isolated either from each other or from the organization's external ecosystem. Thus they don't benefit from the different added values that could be created from sensor data for example, and the useful knowledge that could be extracted from event logs and historical data from previous executions. Further, BPM works in a reactive way [1] which is not sufficient when facing new radical or incremental changes. Early anticipation is crucial to either avoid the occurrence of the problem or respond to it quickly, with no latency, and in an efficient way. This lack of proactivity and predictability is remarkable in three main steps of the BPM life cycle [2] [3]: the design and redesign step, the implementation step, and the execution step. To address this problem, proactive-oriented concepts start to be used in the BPM glossary such as proactive business process management [1] [4], process forecasting or future-oriented BPM [3] and context-aware business processes [5]. Therefore, switching from a reactive to a proactive and adaptable business processes becomes mandatory for every organization. With this new digitization of industrial processes, comes also the age of assistance, which mean that companies should be focused on customers, in order to offer a personalized and adaptable services, and even predict their needs in almost real time. Dealing with such a continuous changing environment requires an intelligent, adaptive and flexible business processes. As a result of all these new changes, organizations nowadays find that the traditional BPM systems present several limitations [6]. In the literature, different approaches have bean proposed to improve business process by applying diverse techniques and technologies such as recommended systems [7] [8], Ontologies [9], data analysis, data mining and process mining [10] [11], complex event processing [1] [12], Ubiquitous Computing [13], Internet of things - IoT [14] [15], just to name few.

Processes are executed within application systems belonging to the real world, where humans, cooperative computer systems and even physical objects are involved. In fact, connected objects are becoming progressively more prominent in the business process execution environment. IoT represents the inter-networking of physical objects [16] (also referred to as ”things”, ”connected devices”, ”smart devices”, ”ubiquitous devices”), vehicles and other items embedded with sensors, electronics, actuators, and network connectivity that enable these ”things” to collect and exchange data when interacting and sensing their environment. At the execution level, Business processes can be classified into Person-to-Person (P2P), Person-to-Application (P2A), and Application-to-Application (A2A) processes [2]. Recently with the emergence of these IoT devices, new Business process interactions are emerging such as Person-to-Thing, Thing-to-person, and Thing-to-Thing due to the advent of Internet of things (IoT) technologies [17]. However, IoT-BPM integration is still at its infant age. Most of the current research work on BPM-IoT integration, propose new approaches that target a specific aspect of the BPM life-cycle. For example, updating business process models by enriching business process model and notation (BPMN) with new elements that correspond to IoT domain and that can explicitly define IoT devices within a business process [18], improving resource Optimization and monitoring and task execution via IoT context-specific knowledge provisioning [19], improving business process execution via an IoT-aware business process execution that exploits IoT for BPM by providing IoT data in a process-aware way [20], or proposing an architecture for IoT-BPM integration in order to cope with the issues and limitations raised by the recent case studies in both industry or academia.

In this paper, we focus on how to improve BPM through IoT integration via an end-to-end architecture. This integration will help us to proactively manage the business process instances, that are launched by different IoT devices, based on their priority level. Although academia and industry have taken an interest in this integration, there is still a lot of research work to be done in order to propose effective methodologies, design patterns and architectures to ensure efficient and smooth integration and communication between the IoT domain and BPM.

In the literature, several research works are paving the way for BPM and IoT combination and integration, in order to optimize BPM using IoT and allow BPM to benefit from this new advanced technology. To go into more detail about our proposal related to IoT-BPM integration/Communication, it is appropriate to ask some research questions in order to define our problem in more concrete way.

- **Q01** : What is the state-of-the art regarding the integration/communication between IoT and BPM?
- **Q02** : What design strategy or methodology can we follow in order to achieve a successful integration/communication between IoT and BPM?
- **Q03** : What are the encountered issues when integrating IoT and BPM?
- **Q04** : And given that data and event are the common points between IoT and BPM, to what extent can the functionalities offered by data analysis and complex event processing be exploited for an end-to-end IoT-BPM architecture?

The reminder of this paper is organized as follow. Section 2 presents our context of work, illustrated with a real-life scenario. Section 3 overviews BPM, IoT and the integration of these two technologies. In Section 4, we present in more details our approach. In Sections 5, 6 and 7 we present an overview of our proposed IoDEP architecture from different perspectives. In Section 8 we describe the implementation of our approach (via an initial block validation), and we discuss and evaluate our results. And finally, we conclude our paper and present some future perspectives in Section 9.

II. Context and Motivating Scenario

In this section, we present a real-life scenario to illustrate our problematic and highlight the challenges we are trying to solve with the proposed approach.

The case study of our research work belongs to silver economy domain, which is a new industrial sector officially launched in 2013 in France. The aim is to create personalized services and new technologies to improve disability-free life expectancy and to help dependent elderly people as well as their care-givers on a day-to-day basis. Most countries all over the world live the demographic transition of aging population. According to the united nations, the number of people with 80+ years old will triple between 2015 (126.4 millions) and 2050 (446.6 millions). If we take France as an example, in 2015 the number of people with more than 60 years old is 12 thousands, which represents 18% of the French population and they will represent more than 1/3 of the population by 2060[1]. Since the demographic change is becoming a global phenomenon, several companies are focusing on developing products and services to create age-friendly societies.

The risk of diseases, loss of capacity and falls increase with age. Losing physical capacities due to age or some kind of accidents can lead to serious falls of elderly people and those falls can have adverse repercussions. In fact, The physical consequences of a fall differ from one individual to another. They can represent a decrease in mobility and an increase in daily life activities dependency. Falls have also some psychological consequences such as a loss of self-confidence, which can accelerate functional capacities decline. Falls among seniors result in a significant number of hospitalizations, with hip fractures being the main cause. Besides, falls are the leading cause of injury-related death.

Several studies have been conducted in the field of silver economy, in order to determine a standard definition of a fall and the number of falls over a specific period to consider an old person as a repetitive case. In [25] [21] [22] [23][24] a fall represents “an unintentional change in position resulting in coming to rest at a lower level or on the ground”. To characterize the repetitive aspect of a fall, we must determine the number of falls and the time interval between falls. As we can see in Table I, the majority of published studies consider at least two falls to retain repetitive character, with an interval between two falls ranging from 6 to 12 months on average. Quick intervention after a fall, using a fall detector for example, could avoid 26 % of hospitalizations, i.e. 160 M Euros and 9,400 deaths per year. There are several solutions for Fall Detection such as:

- Passive Solutions: where the senior must press a beeper to notify in case of an incident.
- Active solutions: these solutions require the use of sensors (Accelerometer, biological signals) or environmental detectors (presence, ground, doors...). In case of a particular variation of the signals, the device triggers an alert.

[1] https://www.insee.fr/fr/accueil

TABLE I. REPETITIVENESS ASPECT OF A FALL

References	Data Collected	Participants	Study plan	Falls repetition
[21]	Questionnaire A follow-up period of one year	N = 730 Age >= 55 years	Transversal	Yes >= 2/12 months
[22]	Postal questionnaire A follow-up period of one year	N = 1660 Age >= 70 years	Transversal	Yes >= 2/12 months
[23]	Telephone questionnaire Interval : 6 weeks A follow-up period of 36 weeks	N = 311 Age >= 70 years	Observational cohort	Yes >= 2/9 months
[24]	A follow-up period of 3 years Participants report their falls weekly on a fall calendar Phone contact in case the person is incapable of filling in his calendar	N = 1365 Age > 65 years	Observational cohort	Yes >= 2/6 months
[25]	Participants interviews Retrospective (12 months)	N = 377 Age =78 +− 3 years	Transversal	Yes >= 2/12 months

- Video-surveillance Solutions: the camera sensor analyzes the senior's behavior and triggers the alert accordingly.

Some of these solutions (bracelets, presence detectors, active floors...) are "blind". This means that they do not allow to know if a fall is a serious one or not, based on the received alert. Only the image delivered by the video fall detectors, allows us to remove doubt about the incident and therefore avoid unnecessary interventions and therefore minimize the overall cost of the service. Predicting and preventing falls among elderly, is the main objective of our case study, in fact to apply our approach, we will use a data set and a business process model from a Video surveillance company. This company edits an automatic falls detection system for elderly people and offers a 24/7 automatic alert solution and a quick rescue without the intervention of the person in danger.

Fig. 1 depicts our incident management process, using BPMN, which aims to manage falls alerts from detection to assistance and resolution. This incident management process is based on an analysis in real-time of alerts received from 24/7 streaming cameras (IoT devices) for detecting elderly people's falls. Waiting too long (sometimes even for few minutes) can be so risky as it can complicate the situation and also it can be so painful for the person. That is why a quick rescue is mandatory to assist the person after a fall or an incident. To achieve this prompt intervention, smart video surveillance cameras are installed at client's home or patients rooms at geriatric services. These devices detect suspicious scenes that may be a fall or an incident, take a picture of the scene and then send automatically an alert to the video-surveillance center. This received alert is handled by a human agent, who qualify the alerts into four categories, as described below, and after that he/she determines whether an assistance action is necessary or not according to the criticality level of the alert. That is why, each alert received requires a quite vigilant treatment, in order to be sure of its category, because the margin of error in this type of system must be very small, as those falls, in case of a delayed intervention or an incorrect qualification, may have an adverse impact on the person concerned: 1) False alerts (level 0): Empty place. 2) False alerts (level 1): Active person. 3) Alerts with average level (level 2): Seated person. 4) High level alerts (level 3): Person lying down.

Video-surveillance systems have proven their efficiency, as they can detect hundreds of falls and risky situations and then assist the person in danger in less than five minutes. However, these ambient cameras videos generates a lot of false alerts, triggered by an active person or a moving curtains for example, that are send to the back-office alert workflow management system. Given that the human agent handles the received alerts in a first in first out (FIFO) order, sometimes true and critical alerts may stay on a waiting list for few minutes until the agent handles all the false alerts or less critical ones that were received before the true one, as the incoming events are intercepted, queued and launch the process instances . So if we do not integrate a mechanism to help the human intelligence by prioritizing the event generated by a (very) serious case, high latency will induce delays that can be disruptive, depending on the severity of the case.

III. BACKGROUND AND LITERATURE OVERVIEW

In the literature, several research works are focusing on combining BPM and IoT. We start this section with a general BPM and IoT background. Then we present a literature Overview about BPM and IoT integration.

A. *Introduction to Internet of things*

Smart objects swept in our life to facilitate it in so many ways and in different domains such as transportation, health care, hospitalization, civil protection, smart home, smart cities, emergency, and individual automation. From smart phones to new smart objects that interact not only with people but with other machines (Machine to machine communication). The concept of the "Internet of Things" first emerged in a presentation by K. Ashton, on the connection of Radio Frequency Identification (RFID) to the supply chain of Procter and Gamble in 1999. Since then, IoT has been exploding and invading our daily lives in different aspects (smart phones, smart door locks, self-driving smart cars, smart cameras, smart medical devices, etc.). This new technology started to thrive right after the development of the web in the 1990's and mobile Internet in the 2010's. The internet of things (IoT) is growing by leaps and bounds. It is made up of billions of smart devices that use wireless technology to communicate with each other and with us. IoT infrastructures can range from connected, instrumented devices providing data to intelligent, standalone systems. IoT enables "a world where things can automatically

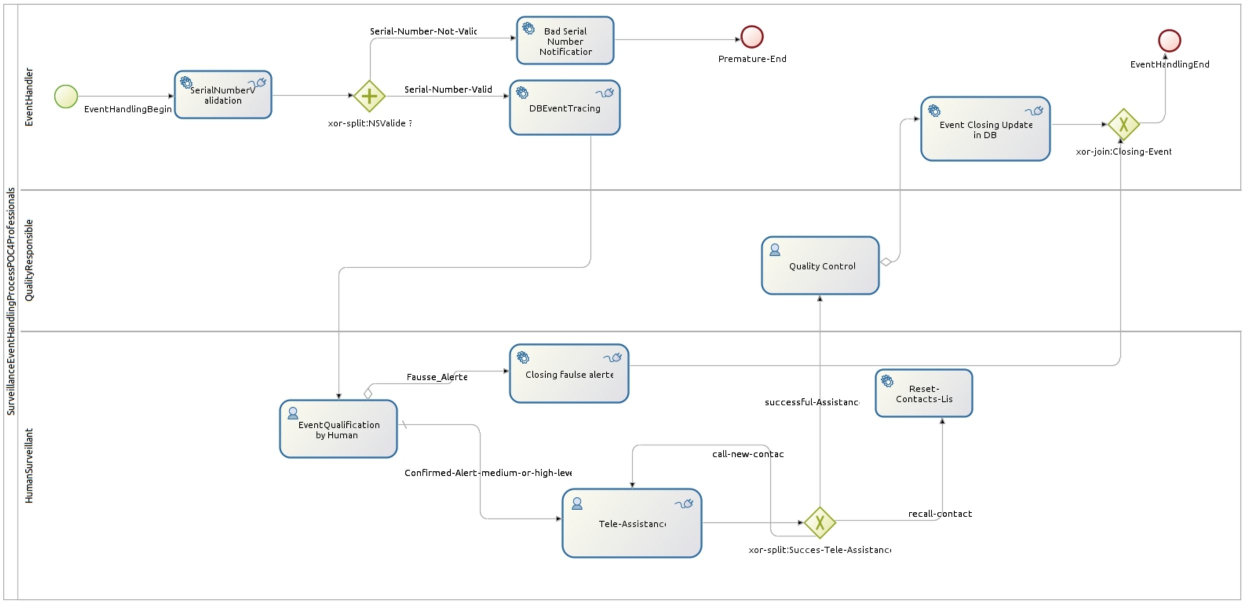


Fig. 1. Motivating Scenario: Qualification and Assessment of the Risk Level of Incidents Process.

communicate to computers and each other providing services to the benefit of the human kind" [26].

Connected devices collect data by sensing their environment, and exchange data with other devices and humans. All devices (things) act locally withing their environment. Whereas, the IoT allows them to be remotely monitored via existing network infrastructures, including the Internet [19]. IoT contribute to continuously feeding "big" data to every node [27]. Besides, the evolution of the IoT invokes significant opportunities for private data exchange enabling new business models across heterogeneous networks [28]. However, it cannot generates value. That is why, it is necessary to couple it with other technologies to transform this huge amount of data into useful knowledge, in order to make meaningful decisions.

B. Introduction to Business Process Management

The Process approach has been increasingly adopted by companies since the 1980s, leading to a new organizational model and a new way of operating in different organizations. Faced with a changing and competitive environment, traditional approaches that treat the company as a closed environment are no longer appropriate. Indeed, the process approach is a systemic approach that aims to transform the hierarchical and vertical structure of an organization into a transverse structure whose ultimate goal is the satisfaction of external and internal customers. It is a method of analysis and modeling intended to ensure collaborative work in order to control and improve the efficiency and smooth running of an organization.

This method is mainly based on the concept of business processes. A business process is the structure of activities and actions as they occur in the real world. It defines all the possible paths in the real process and the rules that determine the path to follow and the actions to perform [29]. The International Standards Organization (ISO) defines processes as a set of interrelated or interacting activities that transform inputs into outputs elements. For Harrington in [30], he presents business processes as an activity or set of activities that uses an input, adds some value to it, and delivers it as an output to an internal or external customer. Dumas et.al [31] represent a business process as a collection of inter-related events, activities, and decision points that involve a number of actors and objects, which collectively lead to an outcome that is of value to a customer. All these definitions and many others present business processes as a set of activities and tasks that exploit the different resources (human and/or machine) of the organization to achieve one or more objectives previously defined, in order to satisfy an internal or external customer. Each business process is attached to a single enterprise, but in some cases it may interact with other business processes belonging either to the same organization or to other organizations. In order to achieve its objectives and ensure efficient performance, the organization must subject its business processes to a continuous improvement mechanism. This mechanism represents the BPM life-cycle [31]. In fact, BPM life cycle is about discovery, modeling/(re)desing, executing, analysis and monitoring of business processes in a perpetual repetition in order to optimize and automate these processes as much as possible, and also to accommodate the ever-changing business requirements.

Business processes therefore occupy a very important place in the field of information systems, because they play a core role within every organization. Moreover, the performance level of any organization is indirectly linked to the efficiency of its processes and the quality of their models. In fact, a proper management of business processes within an organization can have a very positive impact on the efficiency and smooth running of its activities, as it allows this organization to have a clear vision of its objectives in order to better meet the requirements of competitiveness that are constantly increasing.

BPM provides already different methods and solutions to manage and analyse data and events. Among these methods we have Business Process Intelligence (BPI). In fact BPI systems provides solutions to enhance decision making throughout a wide range of business activities, by analysing, predicting, monitoring, controlling and optimizing business processes [32]. Although, BPI has proven to be efficient for post-execution prediction of future process behavior, it is unable to manage and process huge amounts of real-time data and events that are generates from different sources [33]. This becomes more difficult when integrating IoT devices in a BPM architecture. Another solution provided by the BPM field, when dealing with real-time event data is Business activity monitoring (BAM). In fact, BAM is used in order to analyse data related to activities that have been executed. It complements ex-post analysis of process execution by continuously identifying specific situations at run-time and responding to them by triggering specific actions [34]. However, this technology remains less effective in use cases that includes IoT generated events. The limitations of traditional BAM in IoT case studies can be seen from two aspects: prediction and proactivity. The first aspect is manifested in the complex event correlation identification [35]. In fact, by sensing their environment, IoT devices generates a massive volume of event data that need to be processed and analyzed in order to extract useful information and to detect (complex) event patterns in real-time. However, traditional BAM does not provide rule-based engines. This limitation becomes more apparent when the events are generated from diverse data sources, because BAM lacks flexibility in integrating multiple heterogeneous data sources [33] [36] [37]. The second aspect is linked to the absence of proactivity in BAM solutions. In fact, using BAM in a reactive way is no longer sufficient, especially when we have this huge amount of real life data and events.

In business environment where every single event is important and need to be processed, Event Driven Architecture (EDA) needs to be adopted. Now with the emergence of IoT, events are becoming increasingly important for the current information systems (SI), especially for organizations that integrate IoT devices and senors in their business operations (video surveillance, Health care, ...). EDA is the successor of service Oriented Architecture (SOA). The idea behind this paradigm is that everything is an event, all the different components of this architecture interact with each other by events. In this architecture each component is either an event consumer or an event provider. Event consumers subscribe to an intermediate event handler, and providers publish to that handler. When the event handler receives an event from a providers, the handler forwards it to the consumer [38]. The difference between EDA and SOA is that SOA is based on the *"request/response"* concept where the consumer of the service sends a request to the producer, and the producer sends a response that contains either the result or feedback. While EDA is based on the *"publish/subscribe"* concept where the communication pattern between the consumer and the provider is reversed. So in this architecture consumers do not start the communication channel, but they receive the events published by the event providers, which means that the communication is made in an unidirectional way [39].

The main interest of the EDA is to manage real-time process events and data in an efficient way. This message-driven architecture enables the introduction of a higher level of event processing using the Complex Event Processing (CEP) engine. CEP is used to exploit and correlate large event streams generated by heterogeneous data sources in order to produce useful information.

C. Event-Driven Business Process Management

Although, IoT is becoming the hot spot area of technological innovations and economic development promises for many industries and services, it still at its infant age, as we have seen previously. During the last years, both academic and industrial world have been interested in this field and its integration with other domains. However, there is still a lot of research work to do to propose methodologies, design models and architectures in order to ensure an efficient and smooth integration and bidirectional communication between IoT field and BPM. Dealing with event in BPM field is not recent. Event-Driven Business Process Management (EDBPM) concept represents an enhancement of Business Process Management, by including other concepts such as SOA (Service Oriented Architecture), EDA (Event Driven Architecture), SaaS (Software as a Service), BAM (Business Activity Monitoring) and CEP (Complex Event Processing) [40]. This concept was first used in 2003 in a white paper of Bruce Silver Associates in connection with the FileNet P8-BPM platform [41]. The idea behind EDBPM was limited to a single event processing, because the concept of CEP was not well known back then. So early applications of EDBPM was mainly focused on business process monitoring and on Key Performance Indicators (KPIs), or metrics measurement [42]. Despite the fact that the concept of CEP was used for the first time in 2002 by David Luckham is his book "the power of events" [43]. It was until 2007 that an integration of CEP in BPM field has been considered for the first time [44]. The integration of CEP technology within BPM is then known as EDBPM.

Recently, the EDBPM research area has been growing significantly again due to the omnipresence of IoT devices. We find applications that integrate IoT, CEP and (ED)BPM in several research area and business sectors such as Health-care, logistics, manufacturing, banking, smart cities/homes, cultural heritage, agriculture, etc. ([45], [46]). Hence the interest to propose an architecture that integrates all these technologies side by side, in order to achieve a proactive event/instance management.

D. Integration BPM with IoT: Literature Overview

During the last years, both academic and industrial world have been interested in this field and its integration with other fields. However, there is still a lot of research work to do to propose methodologies, design models and architectures in order to ensure an efficient and smooth integration and bidirectional communication between IoT and BPM.

We discuss this related work from two perspectives. The first perspective is the impact of IoT on business process. The second one is the different problems and difficulties encountered when we try to integrate IoT and BPM.

1) How IoT Can Impact Business Process?: The alignment of IoT and BPM is the focus of several research works. However, this alignment comes with various challenges that

TABLE II. BUSINESS PROCESS RELATED LIMITATIONS, ADDRESSED VIA IOT INTEGRATION

BPM Limitations	BPM Life-cycle phases				References
	Process modeling	Process analysis	Process execution	Process monitoring	
Physical surrounding	X		X		[18] [47] [48]
Context-insensitivity			X		[47] [20] [49] [50] [51]
Data-input			X		[51] [16] [20] [47]
Models complexity	X		X		[51]
Transparency				X	[51]
Latency			X		[51]
Event-logs quality		X			[51] [52]
Context-awareness	X	X	X	X	[53] [54]
Knowledge feedback loop	X	X	X	X	[55] [56]

need to be tackled. Both technologies will affect and of course benefit from each other [19] [51]. However, in this paper, we focus only on the influence / benefit of IoT on BPM.

The integration of IoT can provide several benefits for BPM. Besides, it can also address some business process related limitations. Among these limitations we have:

- **Physical surrounding:**Business processes have no access (or limited access) to physical surrounding, as they operate in a cyber surrounding. Taking into consideration the physical surrounding at modeling phase can lead to correct resolution and execution of business processes [18] [47] [48].
- **Context-insensitivity:** Business processes are insensitive to context, they are considered as blind and stateless, which mean that in each business process execution we do not take into consideration neither the results from last process instances nor the context (a context-aware business processes execution). So they need to know the conditions and situations in which IoT operates, given that IoT is by default context sensitive through devices/sensors [47] [20] [49] [50] [51].
- **Data-input:** Business processes have no direct access to data generated by different devices ans sensors. This data could be exploited to BP execution to progress via taking actions (e.g., IoT-based trigger events/alerts, IoT-based decisions...) [51] [16] [20] [47].
- **Models complexity:** Integrating IoT technology can reduce the complexity of process models (for example, replace elements or patterns, ...). It Can also extend and enrich process models. As a result, we have more precise process definitions that accurately reflect the operational reality [51].
- **Transparency:** Integrating IoT technology enhances business process monitoring by increasing BP transparency through data provided by IoT sensors [51].
- **Latency:** The incorporation of IoT technology can ultimately lead activity run time reduction and significant latency that can result in an overall performance enhancement [51].
- **Event-logs quality:** IoT sensors produce a huge amount of data that enrich process event logs. Given that event logs are the fuel of process mining technology enriched event logs provide enriched process models [51] [52].

Table (II) summarises and classifies these Business process related limitations according to the different phases of BPM life-Cycle.

2) Problems Encountered in IoT/BPM Integration : The integration of IoT with BPM certainly contributes to business process improvement. However, this integration comes with several issues and challenges. In [57], authors present the challenges that need to be addressed in Business Process Management Systems (BPMS) to achieve an efficient integration of IoT, such as the absence of direct interaction between the business layer and the edge network, or the problem of complex and inflexible business process models due to a lack of standardization and interoperability when modeling IoT elements and components in BPM. The unexploited potential of extracted data from sensing environment represents a challenging issue when integrating IoT with BPM [58]. Security and data privacy represent another level of IoT/BPM integration concerns [19]. Many other challenges exist and need to be tackled when dealing with an IoT and BPM integration and alignment, these challenges have been summarized by C.Janiesch et.al in their manifesto [19].

In this article, we focus mostly on the challenges linked to event processing and also learning aspect withing an IoT-BPM architecture. The next section details this proposed approach.

IV. PROPOSED APPROACH

Event-driven business process management is mainly adapted in organizations that have a real-time based activities which involved some sensors or some IoT devices that collect data and generate new events by sensing their environment [55]. However, a real-time system must have three main characteristics to ensure better functioning inside any organization [59]: 1) *High availability*, 2) *Low latency* and 3) *Horizontal scalability*. Those three characteristics are mandatory to achieve a real-time and efficient scheduling and event management in BPM.

So in order to ensure a (near) real-time priority-based business process instances management, we resort to an integration of four concepts: IoT (to sense the environment), CEP (to detect situations of interest since it is considered as the standard course for real-time analysis and situation detection [60]), Machine Learning (to analyse our data, find patterns in it and then make predictions, to facilitate decision making), and BPM (to manage our business processes). The idea behind this integrated quaternity of technologies (see Fig. 2) is: 1) to accompany data an event explosion resulting from IoT, 2) to

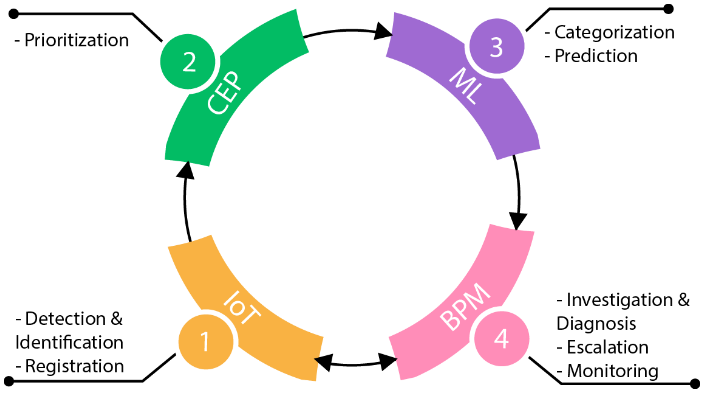


Fig. 2. Quaternity View of BPM, IoT, Machine Learning and CEP.

examine large data sets in order to uncover hidden patterns, unknown correlations between collected events, either at a very technical level (incident/anomaly detection, predictive maintenance) or at business level (customer preferences, market trends, revenue opportunities) 3) to capitalize business value from data generated by IoT sensors, 4) to provide improved operational efficiency, better customer service, competitive advantages over rival organizations.

To turn this conceptual integrated quaternity of technologies into a concrete reality, we propose an end-to-end IoT-BPM architecture (IoDEP: IoT-Data-Event-Process). This integration architecture follows Haze Architecture and Cascading Analytics [61], incarnated by a DIKW (Data Information Knowledge Wisdom) discovery pattern crossing the architecture from device then Fog/Edge to the cloud, and a learning feedback loop that feeds forward insight to adjust either Fog/Edge or device algorithms [55]. IoT-BPM integration involves bidirectional communication. It is possible to acquire data/event from sensors (e.g., monitor and control IoT devices) to manage business process instances and to send instructions to those devices (e.g., reset, adjustment or shut them down). One of the requirement of our proposed architecture is scalability without imposing an architecture redesign. That is the reason behind using Haze Architecture and Cascading architecture as it ensures fluidity and dynamism.

A. *Research Design / Modeling Methodology*

As we said previously, in order to handle the challenges faced by BPM when dealing with IoT objects, IoT and BPM need to meet in the middle. The Incorporation of two heavy paradigms such as IoT and BPM generates, with no doubt, a modeling methodology issue. This issue becomes more complicated when this integration involves other technologies (CEP and ML in this case). There are three types of modeling methodologies: - The Top-down approach, - The Bottom-up approach, - and the Meet-in the- middle approach. We propose in this article a meet-in-the-middle approach, to facilitate the integration of IoT and BPM. In fact, The meet in the middle approach is considered as a method of refinement going alternately from top to bottom to bottom to top. The combination of deductive and inductive iterative sprints in this approach allows both re-use/mutualization and disruptive thinking. In this paper, we propose an integration of IoT and BPM via an end-to-end architecture aiming to provide a meet-in-the-middle environment capable to capture data and event from IoT sensors, when they are sensing their environment, create actionable and useful knowledge, and allow this knowledge to be used in the business layer through business processes.

B. *Functional Requirements*

Business processes are supposed to be smoothly executed under different business situations and context. This constantly changing environment, requires having business processes that can easily be adapted to the appropriate action taken. However, without being coupled to other technology, business processes are still deficient regarding the critical ability to provide assistance to their users [62] due to a lack of two important aspect:

- **Context-awareness**: The emergence or even the omnipresence of IoT solutions in different businesses forces organizations to adapt their processes to a high level of connectivity. Context-awareness is a fundamental characteristic of ubiquitous computing [53], and it is the key to benefit from sensors collected raw data, as it allows to store contextual information related to these raw data and to decide which data should be processed, in order to facilitate the interpretation [54] especially at the level of business processes.
- **Knowledge feedback loop**: Traditional BPM systems present different limits, as they do not facilitate the use of knowledge extracted/generated from data by business processes after their execution. As a result, tremendous amount of data and event data that are constantly collected within the organization is not exploited to improve business processes. As a matter of fact, these data represent for enterprises a real engine of growth. However, a large amount of raw data is not valuable; data must go through a whole process to extract value from it [63]. The analysis of huge data helps organizations to extract information and then knowledge, because the real value is in how organizations will use that data and turn their organization into an information-centric company that relies on insights derived from data analyses for their decision-making.

To accompany data an events explosion resulting, among others, from IoT, data analytic processes combined with event processing techniques, examine large data sets to uncover hidden patterns, unknown correlations between collected events, either at a very technical level (incident/anomaly detection, predictive maintenance) or at business level (customer preferences, market trends, revenue opportunities) to provide improved operational efficiency, better customer service, competitive advantages over rival organizations. In order to capitalize business value of data and events generated by IoT sensors and business process execution, IoT and BPM need to meet in the middle, as we said previously. One critical use case for IoT is to warn organizations when a product or service is at risk. Early detection is essential to either remedy the issue before it becomes a real problem or quickly do cleanup when failure

		IoDEP Architecture				
		IoDEP Architecture Front-End		IoDEP Architecture Back-End		IoDEP Architecture Continuum
		Connected Device/Sensor Processing & Analytics	Fog/Edge Processing & Analytics	Cloud Data Processing & Analytics	Human Processes	Learning Feed back loop
Incident Management Process Functions	Detection/Identification	X	X			X
	Registration		X	X		
	Categorization		X	X		X
	Prioritization			X		X
	Investigation & Diagnosis				X	X
	Escalation			X	X	
	Resolution & Recovery				X	
	Closure				X	

Fig. 3. ITIL Incident Management Process Functions vs IoDEP Architecture Analysis Matrix.

hits[2].

Supervision and incident management processes, especially in health-care domain, are considered as an event-driven business process. The instances of these processes are: In some cases, launched by IoT generated events ('Big' data push paradigm: i.e. data is triggering processes). In order to handle incident management in IoT-BPM architecture, Information Technology and Infrastructure Library (ITIL) repository has been considered. In fact, according to ITIL V3 [64], incident management processes are composed of the following functions:

- **Detection / Identification:** Detect the incident and identification of first elements of classification.
- **Registration:** Record basic details of the incident and propagate the incident alert as necessary.
- **Categorization:** Categorize incidents, Assign impact and urgency, and thereby define priority and match against known errors and problems.
- **Prioritization:** The incident is prioritized for better utilization of the resources and the Support Staff time
- **Investigation and Diagnosis:** Assess the incident details, Collect and analyse all related information, and resolve, (including any work around) or route to online support.
- **Escalation:** Escalate (functionally or hierarchically) where/when necessary.
- **Resolution and recovery:** Resolve the incident and take recovery actions.
- **Closure:** When the Incident has been resolved, the system should ensure that details of the action taken to resolve the incident are concise and readable, classification is complete and accurate according to root cause, resolution/action is achieved.

[2]https://www.informationweek.com/big-data-analytics/when-internet-of-things-meets-big-data

Beside incident management process, IoT-BPM Architecture integrates additional features to target highest IoT maturity levels. Among those features, we have:

- **Monitoring and Communication:** All information, metrics, and key performance indicators applicable to the incident are assessed, recorded and reported (time spent on the incident, support actors, date and time of closure, number and type of reoccurring incidents, average time to achieve incident resolution, percentage of incidents resolved at first line support that meet the Service Level Agreement, etc.).
- **Prediction:** Predicting an incident before happening, will enable anticipatory incident management. This may help avoiding the incident by actionning problem management, or at least, this moves the predicted incident resolution closer to the incident detection insuring proactive incident management. Thus, by collecting and combining connected devices/sensors data with historical context data, IoT-BPM Architecture could provide a wide variety of ad hoc, proactive and anticipatory incident, anomaly, and problem management.

Fig. 3 represents these functional requirements according to each layer of our proposed architecture.

V. Overview of the Layered Architecture

This section provides a high level overview about the different layers of our proposed IoDEP architecture.

Fig. 4 presents the layers of our architecture. In fact, the purpose of this layered architecture is to meet the requirements detailed in the previous section. Our proposed architecture requires five layers (see Fig. 4). IoT sensing layer or Edge Layer, IoT sensor data acquisition Layer, Detection, identification and registration Layer, Categorization and Prioritization Layer, and Cloud layer. In the following we will present a detailed explanation of each layer of this architecture.

A. The Edge Layer

This layer concerns all IoT sensors and devices, each one of these have a precise role depending on their environment, location and the purpose behind using them, such

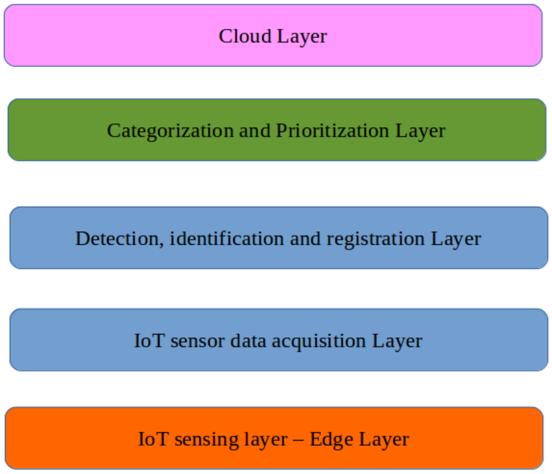


Fig. 4. The Architecture Layers.

as hospital, home, geriatric services, retail store, facilities. IoT devices collect data and generate new events by sensing their environment, after that those events could launch specific business process instances according to the particularities of the detected situation.

As we have seen before, early detection is essential to either remedy the issue before it becomes a real problem or quickly do cleanup when failure hits [65]. The time lapse between sensing the environment and sending a notification, a message or a signal is very critical. In fact, latency is one of the most challenging requirements for connected IoT devices. One of the most effective solution to reduce this lapse of time, is putting some computing on the device or at least bringing computing near the device (edge computing). Pushing data to the edge eliminates long-distance data transmissions to the cloud, which reduces network congestion and latency. Edge nodes are endowed with fall detection capabilities. In fact, these nodes implements several advanced algorithms for image processing and interpretation ([66], [67]). However, this is out of scope of this paper. These algorithms allow the automatic detection and qualification of risky events, following these steps:

- **Image Qualification**: This step allows to qualify the captured image of the camera. Several types of errors are identified at this stage (Error of recovery of the image, Image of incorrect size ...).
- **Movement detection**: This step allows to identify the movements captured by the image, and then render the behaviors of these movements. Mathematical features are identified and extracted from the images and videos. Their exploitation allows to model the behaviors and the events.
- **Movement tracking**: This step consists of tracking objects in their attention zones. Tracking their evolution allows to recognize an abnormal event or behavior and to make a decision.
- **Decision**: At this step a decision about launching an alert or not is taking, based on the results of the previous steps.
- **Pre-qualification**: This step is reinforced with the learning feed back loop [55], by learning from past decisions, mainly false alerts. This final step will consider and exploit the overall context to correct future risk assessments.

At the end of these steps, the detected alerts/events are sent to the cloud. Intercepted events are queued in order to be qualified by a human resource.

B. The Fog Layer

When dealing with IoT devices and sensors, the cloud by itself cannot connect, process and analyze data from thousands and millions of objects and devices of different type and nature spread out over vast areas. To overcome this issue, Fog layer was introduced by Cisco in 2012, in order to offload the cloud through the injection of smart devices at the network layer to provide limited computational resources at the edge of the device layer [68]. Fog computing and edge computing seem similar since they both consist of bringing intelligence and processing closer to data creation. However, the location of the intelligence, processing and computing power is the key difference between these two layers. Generally, intelligence and computing power are placed in devices such as Smart Cameras with embedded vision software (used in our case study). While in the fog layer intelligence and computing facilities are placed at the local area network (LAN).

In our proposed architecture, the fog layer is dedicated to real-time stream processing with CEP as we can see in Fig. 6.

a) Complex Event Processing - CEP: An event is a record of an activity in a system [69] and it represents any change that occurs or will occur in this system, whereas a complex event is a set of events that are related to each other by aggregation, causality or time [70].

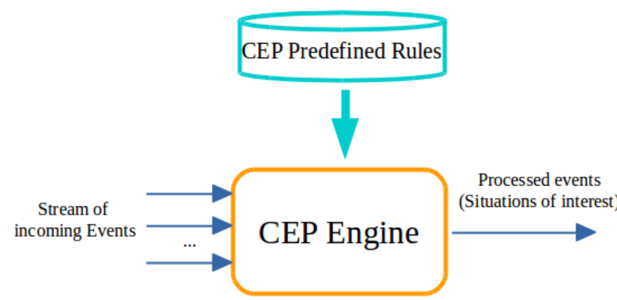


Fig. 5. CEP Basic Architecture.

Complex event processing is a widely used technology, it becomes an increasingly active research field [71] especially with the popularity of publish/subscribe systems in 1997 [72] as stated in [1]. Complex Event Processing (CEP) represents a set of methods, techniques, and tools for processing events while they occur [73]. It aims to process and analyze events generated from different sources in order to extract useful information [74]. CEP is widely used to detect and deal with different business anomalies, threats and opportunities [74] by analyzing event streams instead of traditional static data stored in databases. CEP engines provide the scalability and the (near) real-time processing to filter, to combine and to extract actionable knowledge, known as situation of interest, from a stream of events (see Fig. 5).

Complex situation of interest can be easily expressed using CEP engines and rules. For example, in our case study, queries such as: *prioritize a case if the patient has some particular needs (Wheelchair, walker, etc.)* or *trigger an alert if the patient is a recidivist faller*.

The combination between CEP and BPM is not recent. In fact, it has been widely used to control and monitor business processes in real time in order to improve the effectiveness of business operations by keeping track of what is happening now and raising awareness of issues as soon as they can be detected [75]. CEP helps to monitor not only process instances and activities during run time, but also different events that are related to business process but not necessarily generated from the process instances [76]. Several approaches and solutions based on the integration of CEP with BPM have been proposed for run time or design time ([77], [78], [79]) just to name a few.

The starting point of this incident management approach is the real time analysis of the incoming generated events, using CEP engine to detect the events with the highest priority withing all the incoming ones, that will launch our business process different instances. And based on this estimated priority we can schedule those instances. In fact, the business process instances triggered by those events will then have the highest priority to be assigned to the available human resources. In order to process the incoming events and then detect meaningful patterns concerning the important situation of interest (which represent for us the events with the highest priority), CEP engine needs a set of rules that are determined in advance. A rule for us in this approach represents a condition (IF...THEN ...) that characterizes the event source (the different sources that generate each event). So in the processing step, whenever a condition is satisfied by the event, the priority of this event increases.

CEP strength relies basically on concept of rules and operators. CEP engines are mainly based on a set of rules provided by a rule engine. A rule engine represents a part of a CEP engine that generates rule models. These rules are used in order to create and/or modify business logic in a Business Process Management System (BPMS) [80] for example. CEP rules are based on CEP operators. Among these operators we have: *aggregation operators, sequencing operators, logic operators, single-item operators, windowing operators, and flow management operators*. Those rules are, in most of the time, manually predefined by domain experts, and after that implemented in CEP systems such as Esper, Siddhi, FlinkCEP[3], or Oracle. Since defining these rules manually can be error prone and time consuming, there are many recent approaches that propose an automatic CEP rules learning and generation [1] [81] [74]. However, the automatic CEP rules learning and generation is out of the scope of this paper, because the main purpose of this paper it to pave the way for the use of CEP in order to achieve a real-time analysis of incoming events in order to detect situations of interest, about the priority of the business process instances that need to be executed.

C. The Cloud Layer

Both fog and edge computing are extensions of cloud networks. The majority of enterprises are already familiar with cloud since it is the de facto standard in most industries. The concept of ”cloud” was used in several contexts in the 1990s, but only in 2006 when it became associated with offering services over the internet [82]. Cloud computing is ”a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.” [83]. Cloud layer facilitates storing and accessing data and programs over the internet (as a service) rather than on servers of the enterprise as we can see in Fig. 6. In fact, it offers the ability to drastically outgrow an organization’s normally available storage, without having to host any additional servers. In our IoDEP architecture, Data and intelligence are pushed through layers from edge to cloud in order to be analysed and processed. The knowledge produced goes through a learning feedback-loop that feeds forward insight to adjust either Fog/Edge or device algorithms. To produce this knowledge, we use machine learning algorithm. In the following, we introduce this technology and present in more details the algorithm used in our approach.

a) Machine Learning: Integrating IoT in a BPM environment generates several challenges, among these challenges, we have the huge amount of data and event data that are continuously gathered. Data and event data are the key to get a better understanding of the functioning of business processes. This data represent for enterprises a real engine of growth. However, a large amount of raw data is not valuable; data must go through a whole process to extract value from it. In fact, pre-processing and exploring data before using it help to get correct assumptions and insights in order to make correct predictions and finally take correct and accurate actions and decisions such as instances scheduling, resources management, or business process models redesign. The analysis of this huge generated data helps organizations to extract information and then knowledge, because the real value is in how organizations will use that data and turn their organization into an ”information-centric company that relies on insights derived from data analyses for their decision-making”[4]. Hence integrating data analysis techniques, process mining, data mining, machine learning algorithms or even deep learning in each step of business process life cycle is very crucial for the process improvement. In our approach, we have chosen to exploit the machine learning algorithms in order to enhance one aspect of business process life-cycle, which is the instance management.

Machine learning is a branch of the artificial intelligence research domain. Using mathematical methods, Machine learning enables systems to learn from data and generate knowledge from experience. With time and more experience, the system can learn and improve and sharpen a model that can be used to predict outcomes of questions using previous learning

[3]https://nightlies.apache.org/flink/flink-docs-release-1.14/docs/libs/cep/

[4]https://datafloq.com/read/3vs-sufficient-describe-big-data/

[84]. Machine learning algorithms are organized into different categories based on the learning type: supervised learning, unsupervised learning, semi-supervised learning, Reinforcement learning, and Transduction learning. Supervised learning and unsupervised learning algorithms are the most used in real world problems. They can be further grouped into different categories that we can find in real world machine learning problems. However, the choice of the appropriate algorithm depends on your case study, the type of data manipulated, and the purpose of your analysis.

The first step of our approach aims to ensure an efficient acquisition, filtering, and analyzing incident/event data generated by IoT devices. The second step consists on scoring the generated events from IoT devices, based on the result of the first step, using unsupervised machine learning algorithm. We opted for clustering algorithms to discover groups in our dataset, in order to achieve a categorization of the event sources that trigger our process instances. We choose K-means clustering algorithms, which is a partitioning technique used to analyse data based on the distance between different data points in the input dataset. This algorithm was described by Hartigan in 1975. The idea behind K-means algorithm is to divide a dataset composed of M data points in N dimensions into K clusters, in such a way that the within-clusters sum of squares is minimized [85]. The most complicated part of this algorithm is the determination of the right value of K which represents the number of clusters. In the literature, we can find several methods for selecting the most optimal number of clusters for this algorithm such as: The Elbow method [86]: The Average Silhouette method [87], or the GAP statistic method [88]. When applying K-means algorithm in our approach, we have used the Elbow method to determine the value of K. The basic steps of K-means algorithm are shown in the following pseudo-code:

Algorithm 1 K-Means Clustering Algorithm

Input:
S = s_1, s_2..... s_m // list of data points (list of sources which generate the different events)
K //Number of clusters

```
choose K Random data points from S as initial clusters centroids
repeat
    Assign each data point s_i to the cluster which has the closest centroids.
    Calculate the new centroids of each cluster.
until Convergence //no more changes for centroids
Output: Set of K clusters
```

We tested several criteria such as the frequency of falls or total number of falls, in order to have the most representative clustering for our data. We apply K-means algorithm on a set of events sources in order to classify those sources on different clusters using a score that we calculate for each event's source (a patient in our case) based on the frequency of previously generated events and their criticality value given previously by the agents (human resources) in the qualification step.

VI. Overview of the Front-end and Back-end of the Architecture

From a front-end and back-end perspective, our architecture can be seen as follows (see Fig. 3):

The IoDEP Front-end architecture is composed of two parts:

- **Connected Device/Sensor Processing and Analytics:** insure acquisition of incident data, and incident data filtering and Simple classifier.
- **Fog/Edge Processing and Analytics:** insure the following functionalities:
    1) **incident data processing**: Detection of anomaly incident.
    2) **incident data analytics**: Pattern recognition/correlation/scoring (advanced supervised time-based analysis algorithms here need smaller training set but may need more performance resources like GPU).
    3) **incident data routing**: Transmission of the anomaly information through an Edge Spooler.

The IoDEP Architecture Back-End is represented by several components insuring cloud side processing and analytic of incident management:

- **Cloud Data Processing and Analytics:** Ensure the following functionalities:
    1) **incident data routing**: Transmission of the anomaly information to the relevant back end processing and analytics system – ESB/CEP.
    2) **incident data analytics and intelligence**: Extraction, cleaning and annotation, Integration, aggregation and representation, Modelling and analysis Pattern recognition/correlation/scoring (more sophisticated supervised machine learning algorithms (e.g. deep learning) may here need big training sets) (Big Data).
    3) **incident data processing**: Anomaly Human Processes (Human qualification of the anomaly information) and Enterprise Business Processes (BPMS based on Big Data analytics), and Interpretation: Through on Reporting incident KPI Scoreboards based on (Data Warehousing and Data visualisation).

VII. Application Architecture Overview

In this section, we present a concrete application of our architecture (see Fig. 6). The content of the Edge layer can differ from one case study to another. For our case study, IoT devices were represented by smart cameras that detect the falls of patients. In other context, these devices and sensors could be either Smart locks, Fire and smoke alarms, or smart wristband, etc.

Our Fog layer represents the real-time stream processing layer using CEP technologies. We propose the use of Apache Flink framework as it facilitates complex event processing

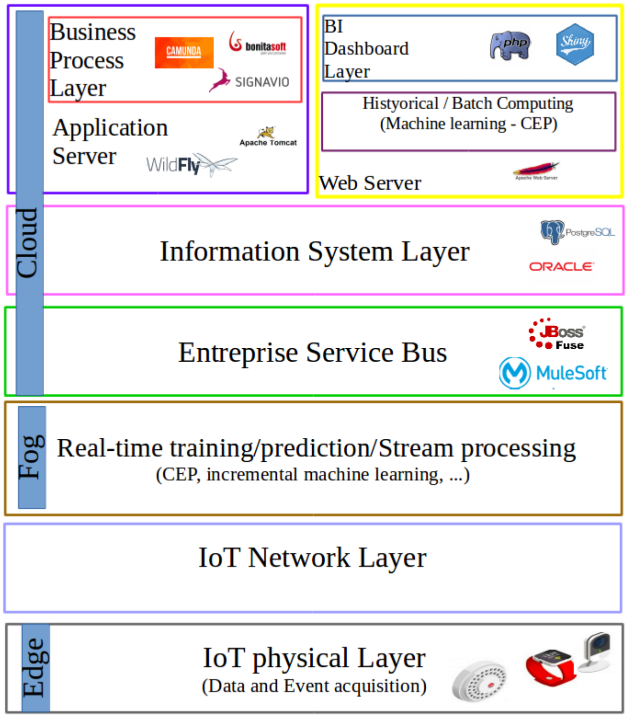


Fig. 6. IoDEP Applicative Architecture.

for real time analysis. Our choice is based on the benchmark that we have conducted. The Enterprise Service Bus layer represents a message broker. It is an integration solution implementing a totally distributed architecture where the applications or services to be integrated are distributed on different machines or information systems, and its role is to ensure communication and interoperability between these different applications whatever their communication protocols. ESB is mostly adapted to asynchronous communications, publish/subscribe messaging, and message queues. Solutions such as Mule ESB, PEtALS, JBoss ESB, Glassfish ESB, or Apache Camel, could be used in this architecture. Historical data in our architecture are stored in a PostgreSQL database. However other Data Base Management System (DBMS) can be used. For the Business Process Layer, several Business Process management Systems can be used such as Bonitasoft[5] (that we have used in our case), Signavio, or Camunda. As an application server, solutions such as Apache Tomcat, or WildFly can be used in this architecture. As a web server to distribute web content, examples such as Apache web Server, can be integrated in this architecture. And finally, for communicating results a BI Dashboard layer, interactive Dashboards can be build using PHP, or Shiny R package given that we have used R language in our experiments.

VIII. Experiments and Results

We opted for à block validation in order to conduct an initial validation of the efficiency of our IoEDP architecture.

Experimental settings:

We worked with a dataset of patients falls (from 01-02-2016 to 12-06-2017), this dataset contains 238228 observations generated by 81 patients.The historical data in this dataset are gathered from our incident management process past instances, and they are partitioned as follows: 89312 alerts are of level 0 (low), 148466 of level 1(average), 275 of level 2 (serious) and 175 of level 3 (very serious). Since that the serious and very serious alerts are the most important in our study, we have applied our clustering algorithms on these two levels only.

The dataset was stored in a PostGreSQL database and all analyses and algorithms implementation were conducted using R and different R packages[6], such as: (tidyverse, RPostgreSQL, ggplot2, dplyr, caret, ...). And All our experiments were conducted on an Intel(R) Core(TM) i5- 540M2.53GHz. All data have been anonymized.

Experimental results for K-means:

As we have seen before, to determine the value of K, which represents the number of clusters, we have used the Elbow method (see Fig. 7).

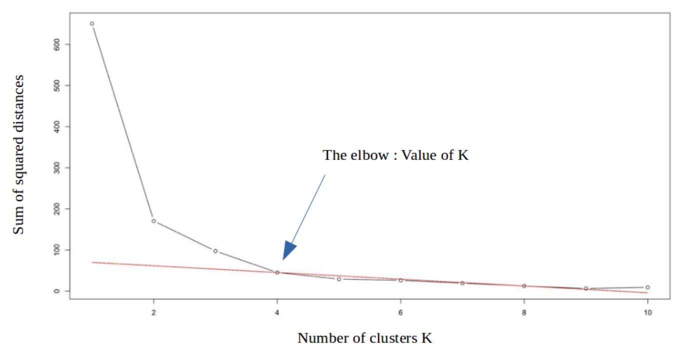


Fig. 7. Elbow Method - Sum of Squared Distances.

The plot represents the variation of the sum of squares with the number of classes. We notice that the most adequate number of clusters would be 4, since this point corresponds to an inflection point and the sum of squares seems to stabilize from this point on.

As we can see in Fig. 8, patients used in this clustering are divided into 4 clusters: patients with fewer than 8 falls (the blue cluster), 9 to 12 (the green cluster), 13 to 18 (the red cluster) and more than 24 falls (the black cluster). These latter represent the most critical cases. As we have seen before, The clustering algorithm helps us to categorize our event sources. In the following we will present some experimental results that proves the interest of integrating CEP into the IoDEP architecture, in order to manage the event that will trigger the process instances.

Experimental results for CEP integration:

The purpose of this series of experiments is to show the interest of integrating CEP in the IoT-BPM architecture. So we will compare two solutions. *Solution 1* represents incident management in our architecture without the integration of CEP, and *Solution 2* represents the integration of CEP in the architecture.

[5]https://fr.bonitasoft.com/

[6]https://www.tidyverse.org/

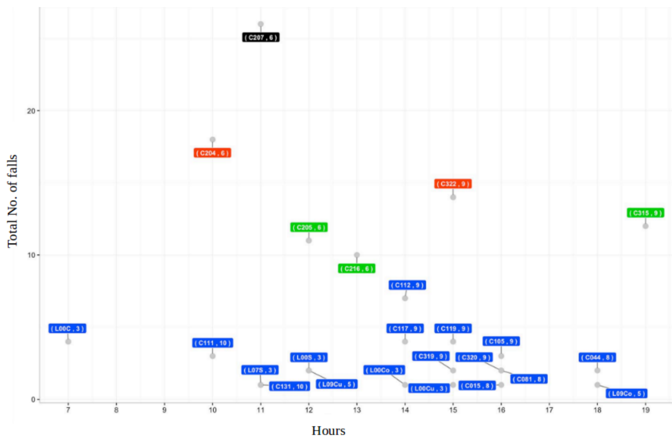


Fig. 8. Clustering with K-means.

CEP Rules:

Based on the context of our case study, we have defined some rules that help us to estimate the priority level of the incoming event. To define those rules we take into consideration the available information about the patient, his/her past incidents (falls) and the cluster to which this patient (event source) belongs. Taking into consideration the results of our clustering (detailed in the beginning of this section), we tried to manually define some rules. So basically we come up with the following rules:

- **IF** the event source belongs to cluster 4 (the critical cluster) **THEN** the new event generated by this source might be serious.
- **IF** The patient has some particular needs (Wheelchair, walker, etc.) **THEN** the new event generated by this source might be serious.
- **IF** The last event generated by this patient within one month was a serious or very serious alert **THEN** the new event generated by this source might be serious.

CEP Message broker:

As we have seen before, in order to manage the different incoming events within a CEP solution, we need a message broker. In this experiment, we have chosen RabbitMQ[7]. It is an open source message broker, lightweight and easy to deploy.

Global Schema of Event-pattern detection with CEP solution:

Fig. 9 illustrates our first attempt to integrate CEP engine in our IoT-BPM architecture.

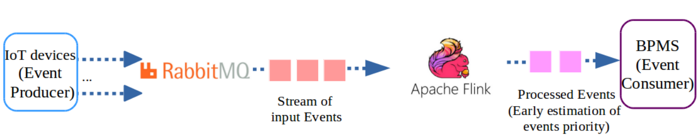


Fig. 9. Priority-based Event Management with CEP.

[7]https://www.rabbitmq.com/

The different modules of this architecture operate as follows:

- **Event Producer:** Events in this approach are mostly generated by some sensors or some IoT devices by sensing their environment (Smart cameras).
- **Lightweight Message Broker:** To manage the amounts of events received and that need to be processed by the CEP engine, we use a message broker that ensure the communication between the source and the target based on a publish/subscribe mechanism. This asynchronous mechanism implemented by message brokers allows source and target messages to be completely decoupled. Besides the message brokers can as well store the messages locally until they can be processed by the target element. That is why we have chosen RabbitMQ.
- **CEP engine:** Flink CEP is used in this case to filter and process incoming events based on the predefined rules, in order to detect the events with highest priority among the incoming stream of events.
- **Event consumer:** It represents in this approach a business process management system (BPMS) where the processes are managed, executed and monitored.

In our experiment for both solutions, we have simulated several event streams with different total number of events (100, 200, 300, 400, 500) (generated from the historical events in our data set). Our objective is to compare the evolution of the computation time of both solutions (as described in the following tables) as a response to increasing the number of input events (non-concurrent access (see Table III and Table IV) and concurrent access (see Table V and Table VI)). Besides, in both solutions we have used the results of the clustering step.

Non-Concurrent Access (NCA):

TABLE III. COMPUTATION TIME (SEC) FOR SOLUTION 1 - NCA

Total event number	Computation time (sec) - Solution 1
100	19.0
200	19.45
300	25.48
400	31.3
500	37.4

TABLE IV. COMPUTATION TIME (SEC) FOR SOLUTION 2 - NCA

Total event number	Computation time (sec) - Solution 2
100	22.07
200	32.14
300	44.25
400	55.6
500	70.62

Concurrent Access (CA) :

As we can see in Fig. 10, solution 1 presents better results comparing to solution 2 when we have a non-concurrent access of the incoming events. However, when we have a concurrent access, the CEP-based approach (solution 2) presents better results, especially when we increase the number of incoming events (see Fig. 11).

TABLE V. COMPUTATION TIME (SEC) FOR SOLUTION 1 - CA

Total event number	Computation time (sec) - Solution 1
200	48
400	72.1
600	108

TABLE VI. COMPUTATION TIME (SEC) FOR SOLUTION 2 - CA

Total event number	Computation time (sec) - Solution 2
200	27
400	34
600	60.3

Although solution 1 seems to be more efficient at low input events volumes, the CEP solution can performs better especially if implemented in an effective IoT architecture with 'Big' Data requirements.

For incident management systems providing a balance between (near) real-time event processing and scalability, it is very important to achieve an efficient and optimized business process instances scheduling and event management in BPM. Moreover, in real cases, we deal in most of the time with concurrent access of incoming events. So this confirm the efficiency of our assumptions that CEP can provide better results when integrated to an IoT-BPM architecture, and it can also provide better results comparing to traditional approaches for business process instances scheduling. Those initial results are encouraging to implement the entire end-to-end IoDEP architecture.

Interested readers can check the complete solution that we have implemented from GitHub[8].

IX. CONCLUSIONS AND PERSPECTIVES

Integrating IoT and BPM as a step towards an improved process management that benefits from data and event power, is possible via an integrated architecture (IoDEP). The idea behind this architecture is to manage data and events at the same time via an integration approach that includes four concepts: IoT (to sense the environment), CEP (to detect situations of interest since it is considered as the standard course for real-time analysis and situation detection), Machine Learning (to analyse our data, find patterns in it and then make predictions, to facilitate decision making), and BPM (to manage our business processes instances.

We introduced in this article the reason behind using machine learning algorithms and CEP techniques in our IoT-BPM communication approach. In fact, this bi-directional communication is established through event in one direction and data in another direction. That is why conducting a data analysis approach with event management can facilitate this communication/integration. Moreover, we have presented the different functionalities proposed by this architecture and also the different requirements that should be addressed. Throughout the different overviews of our proposed architecture, we argue that our approach is generic and can be used in multivariate settings, and most importantly in normal and strict environments where time and priority matter.

[8]https://github.com/Abir-IA/CEPFlink_EventManagement

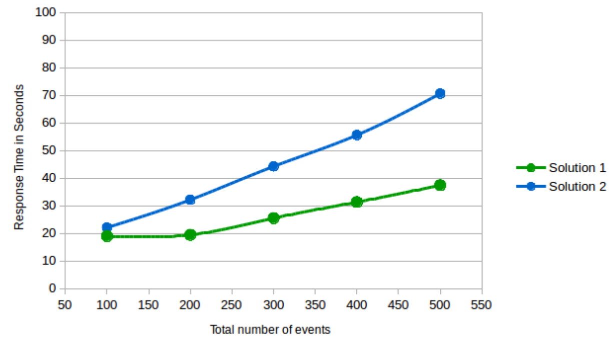


Fig. 10. Input Events with Non-concurrent Access.

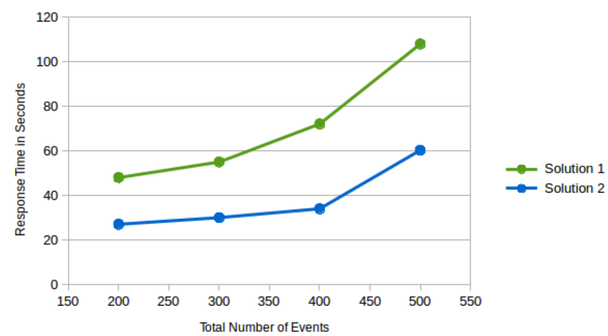


Fig. 11. Input Events with Concurrent Access.

Fully implementing this architecture, extending to test more machine learning algorithms, and exploring the automatic CEP rules learning in order to enrich the business rules engine for proactive event processing and process instances execution, are all plans to realize on our future schedule.

REFERENCES

[1] R. Mousheimish, *Combining the Internet of things, complex event processing, and time series classification for a proactive business process management.* PhD thesis, Université Paris-Saclay, 2017.

[2] W. M. Van der Aalst, "Business process management: a comprehensive survey," *International Scholarly Research Notices*, vol. 2013, 2013.

[3] R. Poll, A. Polyvyanyy, M. Rosemann, M. Röglinger, and L. Rupprecht, "Process forecasting: Towards proactive business process management," in *International Conference on Business Process Management*, pp. 496–512, Springer, 2018.

[4] A. Ismaili-Alaoui, K. Benali, K. Baïna, and J. Baïna, "Business process instances scheduling with human resources based on event priority determination," in *International Conference on Big Data, Cloud and Applications*, pp. 118–130, Springer, 2018.

[5] M. Anastassiu, F. M. Santoro, J. Recker, and M. Rosemann, "The quest for organizational flexibility: driving changes in business processes through the identification of relevant context," *Business Process Management Journal*, vol. 22, no. 4, pp. 763–790, 2016.

[6] F. Bergenti, G. Caire, and D. Gotta, "Interactive workflows with wade," in *2012 IEEE 21st International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises*, pp. 10–15, IEEE, 2012.

[7] I. Barba, B. Weber, C. Del Valle, and A. Jiménez-Ramírez, "User recommendations for the optimized execution of business processes," *Data & knowledge engineering*, vol. 86, pp. 61–84, 2013.

[8] M. A. Setiawan, S. Sadiq, and R. Kirkman, "Facilitating business process improvement through personalized recommendation," in *Inter-*

national Conference on Business Information Systems, pp. 136–147, Springer, 2011.

[9] M. Khanbabaei, F. M. Sobhani, M. Alborzi, and R. Radfar, "Developing an integrated framework for using data mining techniques and ontology concepts for process improvement," *Journal of Systems and Software*, vol. 137, pp. 78–95, 2018.

[10] W. Van Der Aalst, A. Adriansyah, A. K. A. De Medeiros, F. Arcieri, T. Baier, T. Blickle, J. C. Bose, P. Van Den Brand, R. Brandtjen, J. Buijs, *et al.*, "Process mining manifesto," in *International Conference on Business Process Management*, pp. 169–194, Springer, 2011.

[11] S. Aguirre, C. Parra, and J. Alvarado, "Combination of process mining and simulation techniques for business process redesign: a methodological approach," in *International Symposium on Data-Driven Process Discovery and Analysis*, pp. 24–43, Springer, 2012.

[12] D. Redlich and W. Gilani, "Event-driven process-centric performance prediction via simulation," in *International Conference on Business Process Management*, pp. 473–478, Springer, 2011.

[13] A. Yousfi, K. Batoulis, and M. Weske, "Achieving business process improvement via ubiquitous decision-aware business processes," *ACM Transactions on Internet Technology (TOIT)*, vol. 19, no. 1, pp. 1–19, 2019.

[14] C. Stoiber, "Exploiting internet of things for business process improvement," 2021.

[15] M. M. Queiroz, S. F. Wamba, M. C. Machado, and R. Telles, "Smart production systems drivers for business process management improvement: An integrative framework," *Business Process Management Journal*, 2020.

[16] S. Cherrier and V. Deshpande, "From bpm to iot," in *International conference on business process management*, pp. 310–318, Springer, 2017.

[17] R. Wehlitz, I. Rößner, and B. Franczyk, "Integrating smart devices as business process resources–concept and software prototype," in *International Conference on Service-Oriented Computing*, pp. 252–257, Springer, 2017.

[18] S. Meyer, A. Ruppen, and C. Magerkurth, "Internet of things-aware process modeling: integrating iot devices as business process resources," in *International conference on advanced information systems engineering*, pp. 84–98, Springer, 2013.

[19] C. Janiesch, A. Koschmider, M. Mecella, B. Weber, A. Burattin, C. Di Ciccio, G. Fortino, A. Gal, U. Kannengiesser, F. Leotta, *et al.*, "The internet of things meets business process management: a manifesto," *IEEE Systems, Man, and Cybernetics Magazine*, vol. 6, no. 4, pp. 34–44, 2020.

[20] S. Schönig, L. Ackermann, S. Jablonski, and A. Ermer, "An integrated architecture for iot-aware business process execution," in *Enterprise, business-process and information systems modeling*, pp. 19–34, Springer, 2018.

[21] P. A. Stalenhoef, J. P. Diederiks, L. P. de Witte, K. H. Schiricke, and H. F. Crebolder, "Impact of gait problems and falls on functioning in independent living persons of 55 years and over: a community survey," *Patient Education and Counseling*, vol. 36, no. 1, pp. 23–31, 1999.

[22] P. A. Stalenhoef, J. P. Diederiks, J. A. Knottnerus, L. P. de Witte, and H. F. Crebolder, "The construction of a patient record-based risk model for recurrent falls among elderly people living in the community," *Family Practice*, vol. 17, no. 6, pp. 490–496, 2000.

[23] P. Stalenhoef, J. Diederiks, J. Knottnerus, A. Kester, and H. Crebolder, "A risk model for the prediction of recurrent falls in community-dwelling elderly: a prospective cohort study," *Journal of clinical epidemiology*, vol. 55, no. 11, pp. 1088–1094, 2002.

[24] V. S. Stel, S. M. Pluijm, D. J. Deeg, J. H. Smit, L. M. Bouter, and P. Lips, "A classification tree for predicting recurrent falling in community-dwelling older persons," *Journal of the American Geriatrics Society*, vol. 51, no. 10, pp. 1356–1364, 2003.

[25] K. A. Faulkner, M. S. Redfern, J. A. Cauley, D. P. Landsittel, S. A. Studenski, C. Rosano, E. M. Simonsick, T. B. Harris, R. I. Shorr, H. N. Ayonayon, *et al.*, "Multitasking: association between poorer performance and a history of recurrent falls," *Journal of the American Geriatrics Society*, vol. 55, no. 4, pp. 570–576, 2007.

[26] A. Furness, "A framework model for the internet of things," *GRIFS/CASAGRAS: Hong Kong, China*, 2008.

[27] J. Rifkin, *The zero marginal cost society: The internet of things, the collaborative commons, and the eclipse of capitalism*. St. Martin's Press, 2014.

[28] E. Natalizio, *Mobility as a Primitive to Improve Communications in Wireless Networks*. PhD thesis, UTC Compiègne, 2017.

[29] D. Grigori, *Eléments de flexibilité des systèmes de workflow pour la définition et l'exécution de procédés coopératifs*. PhD thesis, Université Henri Poincaré-Nancy 1, 2001.

[30] H. Harrington, "P. american society for quality control quality, business process improvement: the breakthrough strategy for total quality, productivity, and competitiveness," 1991.

[31] M. Dumas, M. La Rosa, J. Mendling, H. A. Reijers, *et al.*, *Fundamentals of business process management*, vol. 1. Springer, 2013.

[32] D. Grigori, F. Casati, M. Castellanos, U. Dayal, M. Sayal, and M.-C. Shan, "Business process intelligence," *Computers in industry*, vol. 53, no. 3, pp. 321–343, 2004.

[33] B. Schwegmann, M. Matzner, and C. Janiesch, "A method and tool for predictive event-driven process analytics.," in *Wirtschaftsinformatik*, p. 46, Citeseer, 2013.

[34] W. Schmidt, "Business activity monitoring (bam)," in *Business Intelligence and Performance Management*, pp. 229–242, Springer, 2013.

[35] O. Etzion and P. Niblett, "Event processing in action, manning publications," 2011.

[36] R. von Ammon, O. Etzion, H. Ludwig, A. Paschke, and N. Stojanovic, "Introduction to the second international workshop on event-driven business process management (edbpm09)," in *International Conference on Business Process Management*, pp. 345–346, Springer, 2009.

[37] C. Janiesch, M. Matzner, and O. Müller, "Beyond process monitoring: a proof-of-concept of event-driven business activity management," *Business Process Management Journal*, vol. 18, no. 4, pp. 625–643, 2012.

[38] D. Bo, D. Kun, and Z. Xiaoyi, "A high performance enterprise service bus platform for complex event processing," in *2008 Seventh International Conference on Grid and Cooperative Computing*, pp. 577–582, IEEE, 2008.

[39] D. Luckham, "Soa, eda, bpm and cep are all complementary," *Retrieved May*, vol. 20, p. 2011, 2007.

[40] A. Estruch and J. A. H. Álvaro, "Event-driven manufacturing process management approach," in *International Conference on Business Process Management*, pp. 120–133, Springer, 2012.

[41] R. v. Ammon, C. Emmersberger, F. Springer, and C. Wolff, "Event-driven business process management and its practical application taking the example of dhl," 2008.

[42] C. Costello, O. Molloy, and K. Hinkelmann, "Towards a semantic framework for business activity monitoring and management.," in *AAAI spring symposium: AI meets business rules and process management*, pp. 17–27, 2008.

[43] D. Luckham, *The power of events*, vol. 204. Addison-Wesley Reading, 2002.

[44] K. Danielsson and G. Trotta, "Key requirements for event-driven bpm and soa."

[45] W. Yao, C.-H. Chu, and Z. Li, "Leveraging complex event processing for smart hospitals using rfid," *Journal of Network and Computer Applications*, vol. 34, no. 3, pp. 799–810, 2011.

[46] R. Mousheimish, Y. Taher, K. Zeitouni, and M. Dubus, "Pact-art: Enrichment, data mining, and complex event processing in the internet of cultural things," in *2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pp. 476–483, IEEE.

[47] Z. Maamar, E. Kajan, I. Guidara, L. Moctar-M'Baba, and M. Sellami, "Bridging the gap between business processes and iot," in *Proceedings of the 24th Symposium on International Database Engineering & Applications*, pp. 1–10, 2020.

[48] C. Friedow, M. Völker, and M. Hewelt, "Integrating iot devices into business processes," in *International conference on advanced information systems engineering*, pp. 265–277, Springer, 2018.

[49] R. Song, Y. Wang, W. Cui, J. Vanthienen, and L. Huang, "Towards improving context interpretation in the iot paradigm: a solution to integrate context information in process models," in *Proceedings of the 2018 2nd International Conference on Management Engineering, Software Engineering and Service Sciences*, pp. 223–228, 2018.

[50] S. Schönig, L. Ackermann, S. Jablonski, and A. Ermer, "Iot meets bpm: a bidirectional communication architecture for iot-aware process execution," *Software and Systems Modeling*, vol. 19, no. 6, pp. 1443–1459, 2020.

[51] S. Schönig, L. Ackermann, and S. Jablonski, "Internet of things meets bpm: A conceptual integration framework.," in *SIMULTECH*, pp. 307–314, 2018.

[52] C. Sturm, S. Schönig, and C. Di Ciccio, "Distributed multi-perspective declare discovery.," in *BPM (Demos)*, 2017.

[53] D. Evans, "The internet of things: How the next evolution of the internet is changing everything," *CISCO white paper*, vol. 1, no. 2011, pp. 1–11, 2011.

[54] Z. A. Almusaylim and N. Zaman, "A review on smart home present state and challenges: linked to context-awareness internet of things (iot)," *Wireless networks*, vol. 25, no. 6, pp. 3193–3204, 2019.

[55] A. Ismaili-Alaoui, K. Baïna, K. Benali, and J. Baïna, "Towards smart incident management under human resource constraints for an iot-bpm hybrid architecture," in *International Conference on Web Services*, pp. 457–471, Springer, 2018.

[56] C. Soto, "Bpm and iot: The signal of true disruption," *Redmond Media Group*, 2015.

[57] C. Chang, S. N. Srirama, and R. Buyya, "Mobile cloud business process management system for the internet of things: a survey," *ACM Computing Surveys (CSUR)*, vol. 49, no. 4, p. 70, 2017.

[58] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of things (iot): A vision, architectural elements, and future directions," *Future generation computer systems*, vol. 29, no. 7, pp. 1645–1660, 2013.

[59] J. Pielmeier, S. Braunreuther, and G. Reinhart, "Approach for defining rules in the context of complex event processing," *Procedia CIRP*, vol. 67, pp. 8–12, 2018.

[60] D. C. Luckham, *Event processing for business: organizing the real-time enterprise*. John Wiley & Sons, 2011.

[61] D. Canty, "Architecting for dynamic analytics in iot.." https://dzone.com/articles/architecting-for-dynamic-analytics-in-iot, 2014. [Online; accessed July-2021].

[62] S. Tjoa, S. Jakoubi, G. Goluch, G. Kitzler, S. Goluch, and G. Quirchmayr, "A formal approach enabling risk-aware business process modeling and simulation," *IEEE Transactions on Services Computing*, vol. 4, no. 2, pp. 153–166, 2011.

[63] A. Ismaili-Alaoui, K. Baïna, and K. Benali, "Harnessing the power of data and event data for business process improvement," in *CAISAM 2019-Complexity Analysis of Industrial Systems and Advanced Modeling*, 2019.

[64] U. ITIL, "Introducing service operation.." https://www.ucisa.ac.uk. [Online; accessed January-2021].

[65] J.Kaplan, "When internet of things meets big data." https://www.informationweek.com/big-data/big-data-analytics/when-internet-of-things-meets-big-data/a/d-id/1298137?,. [Online; accessed July-2021].

[66] J. Gutiérrez, V. Rodríguez, and S. Martin, "Comprehensive review of vision-based fall detection systems," *Sensors*, vol. 21, no. 3, p. 947, 2021.

[67] Y. Chen, X. Kong, L. Meng, and H. Tomiyama, "An edge computing based fall detection system for elderly persons," *Procedia Computer Science*, vol. 174, pp. 9–14, 2020.

[68] Cisco, ""cisco fog computing solutions: Unleash the power of the internet of things", syst. inc., san jose, ca, usa, white paper, 2015." https://www.cisco.com/c/dam/en_us/solutions/trends/iot/docs/computing-solutions.pdf. [Online; accessed July-2021].

[69] D. Luckham, "A short history of complex event processing," *Part 1: Beginnings. 1-a-short-history-of-cep-part-1. pdf*, 2007.

[70] D. Robins, "Complex event processing," in *Second International Workshop on Education Technology and Computer Science. Wuhan*, pp. 1–10, Citeseer, 2010.

[71] G. Cugola and A. Margara, "Processing flows of information: From data stream to complex event processing," *ACM Computing Surveys (CSUR)*, vol. 44, no. 3, p. 15, 2012.

[72] D. S. Rosenblum and A. L. Wolf, *A design framework for Internet-scale event observation and notification*, vol. 22. ACM, 1997.

[73] M. Eckert and F. Bry, "Complex event processing (cep),"

[74] A. Margara, G. Cugola, and G. Tamburrelli, "Learning from the past: automated rule generation for complex event processing," in *Proceedings of the 8th ACM International Conference on Distributed Event-Based Systems*, pp. 47–58, ACM, 2014.

[75] M. von Rosing and S. Polovina, "Business process trends," 2015.

[76] M. Daum, M. Götz, and J. Domaschka, "Integrating cep and bpm: how cep realizes functional requirements of bpm applications (industry article)," in *Proceedings of the 6th ACM International Conference on Distributed Event-Based Systems*, pp. 157–166, ACM, 2012.

[77] M. Weidlich, H. Ziekow, J. Mendling, O. Günther, M. Weske, and N. Desai, "Event-based monitoring of process execution violations," in *International conference on business process management*, pp. 182–198, Springer, 2011.

[78] F. Koetter and M. Kochanowski, "A model-driven approach for event-based business process monitoring," *Information Systems and e-Business Management*, vol. 13, no. 1, pp. 5–36, 2015.

[79] P. Soffer, A. Hinze, A. Koschmider, H. Ziekow, C. Di Ciccio, B. Koldehofe, O. Kopp, A. Jacobsen, J. Sürmeli, and W. Song, "From event streams to process models and back: Challenges and opportunities," *Information Systems*, vol. 81, pp. 181–200, 2019.

[80] B. Mazon-Olivo, D. Hernández-Rojas, J. Maza-Salinas, and A. Pan, "Rules engine and complex event processor in the context of internet of things for precision agriculture," *Computers and electronics in agriculture*, vol. 154, pp. 347–360, 2018.

[81] N. Mehdiyev, J. Krumeich, D. Enke, D. Werth, and P. Loos, "Determination of rule patterns in complex event processing using machine learning techniques," *Procedia Computer Science*, vol. 61, pp. 395–401, 2015.

[82] Q. Zhang, L. Cheng, and R. Boutaba, "Cloud computing: state-of-the-art and research challenges," *Journal of internet services and applications*, vol. 1, no. 1, pp. 7–18, 2010.

[83] P. Mell, T. Grance, *et al.*, "The nist definition of cloud computing," 2011.

[84] J. Bell, *Machine learning: hands-on for developers and technical professionals*. John Wiley & Sons, 2020.

[85] J. A. Hartigan and M. A. Wong, "Algorithm as 136: A k-means clustering algorithm," *Journal of the royal statistical society. series c (applied statistics)*, vol. 28, no. 1, pp. 100–108, 1979.

[86] R. L. Thorndike, "Who belongs in the family," in *Psychometrika*, Citeseer, 1953.

[87] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.

[88] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 2, pp. 411–423, 2001.

Route Planning using Wireless Sensor Network for Garbage Collection in COVID-19 Pandemic

Javier E. Ramirez, Caleb M. Santiago, Angelica Kamiyama
Systems Engineering Department,
University of Lima, Lima 15023, Peru

Abstract—**Garbage collection is a responsibility faced by all cities and, if not properly carried out, can generate greater costs or sanitary problems. Considering the sanitary situation due to the COVID-19 pandemic, it is necessary to take sanitary safety measures to prevent its spread. The challenge of the present work is to provide an efficient and effective solution that guarantees a garbage collection that optimizes the use of resources and prioritizes the attention to garbage containers located in or near contagion risk zones. To this end, this research proposes the integration of a basic garbage monitoring system, consisting of a wireless sensor network, and a route planning system that implements the decomposition of the Vehicle Routing problem into the subproblems of clustering and sequencing of containers using the K-Means and Ant Colony algorithms. For the monitoring of garbage, a significant reduction in the measurement error of waste level in the containers was achieved compared to other authors. About route planning, adequate error ranges were obtained in the calculation of the optimal values of distance traveled and travel time indicators with respect to an exhaustive enumeration of routes.**

Keywords*—*K-Means; ant colony optimization; route planning; vehicle routing problem; garbage collection; wireless sensor network

I. INTRODUCTION

Garbage collection is one of the most complex tasks of urban governments. Often, this task requires that collection vehicles travel along a specific route and manually check all the garbage containers along that route to empty them if necessary. The selection of routes followed by the vehicles is usually done at the discretion of the drivers which causes inefficient use of resources [1]. In general, there is no permanent monitoring of the containers to alert them about their possible overflow to avoid potential sanitation problems. In Peru, on average more than seven million tons of garbage are generated annually, of which about 50% comes from the capital city [2]. In addition, the city of Lima is ranked seventh with the highest traffic congestion worldwide [3]. Since March 2020, Peru has been in a state of health emergency due to the COVID-19 pandemic whose infection and mortality rates, as of January 2022, were 11.64% with respect to the number of tests and 7.22% with respect to the level of infected people in the whole country, respectively. Metropolitan Lima has a case fatality rate of 6.89% with respect to the number of people infected by coronavirus nationwide [4]. Waste collection costs in Peru are close to one billion soles annually, in addition to the work of thousands of operators, and the employment of hundreds of fleets of vehicles [5]. It is important to have an efficient waste collection service capable of dealing with the current health situation. Otherwise, the inefficient use of resources by the municipalities will continue, as well as the increase in the contagion and mortality rate due to the spread of the pandemic. This article presents a solution to the problem of optimizing the garbage collection, considering not only the associated costs but also the impact on public health; it will focus on minimizing collection costs, but at the same time guaranteeing timely attention to the containers by not allowing them to overflow. In this way, by saving collection costs, more resources can be allocated to other municipal processes, meanwhile contributing to the containment of the pandemic and additionally having a positive impact on the environment. It will be assumed that the containers belong to the municipality and it is possible to insert wireless sensors in the containers. This solution includes two systems, the first is a basic monitoring system composed of a network of wireless sensors to detect the volume and level of decomposition of garbage in the containers, with which it will be possible to alert about the imminent overflow of the containers. This monitoring system feeds into a route planning system that uses a heuristic approach to determine the optimal routes for the collection trucks, i.e. the most economical routes. In this solution, preferential attention will be given to containers located in contagion risk zones[1] as follows: for the contents of a container to be emptied, its contents must exceed a certain threshold; for containers located in an contagion risk zone or close to one, this threshold is lower. In section two, related work is reviewed. In section three, the solution methodology is explained. The results and discussion are presented in sections four and five, respectively. In the last sections, conclusions are formulated and future work is proposed.

II. RELATED WORK

A. Waste Monitoring Sensors

An efficient garbage collection process requires knowing the current status of the garbage containers, which can be achieved through the use of different sensors and image processing methods, for which it is necessary to take into account the measurement range, accuracy, sensitivity, consistency, energy consumption, ease, and cost of the device [8]. Several solutions for this process use different sensors such as infrared [9] or ultrasound (HC-SR04) to determine the amount of stored waste, load cells to calculate the weight of the container [10], humidity sensors (DHT11), temperature sensors (LM35) for detecting the presence of fire [11], gas sensors (QS-01) for air quality monitoring [12], and tilt sensors (Itead) to detect possible container overturning. Among these reviewed works,

[1] A contagion risk zone refers to an area facing a health emergency and having a high rate of population density and/or mobility over time [6], [7].

several issues are reported that impact on the quality of the information, such as the angle of measurement and the position of the sensors inside the container, this issue can be solved by using high cost devices to reduce the percentage error in the measurement; therefore, a horizontal measurement is proposed to reduce the percentage error in the measurement through the use of low cost sensors, the mentioned served as motivation to seek to address these challenges evidenced in the research of other authors. Container status information needs to be sent and stored in a database for further analysis and monitoring. There are multiple data transmission techniques, e.g. the use of the Ethernet SPI module via network cable [13], as well as low-power communication technologies such as LoRa, which provides a long transmission range over long distances [14], and ZigBee, which improves data transfer speed [12]. However, these solutions generate a bottleneck by centralizing the sending of information. Another solution is a GSM transmission sensor, which allows data to be sent via SMS to a telephone number [9]. Finally, the NodeMCU device with a built-in Wi-Fi module allows each node of the sensor network to operate independently in data transmission [15]. Container status information can be used in the design of collection truck routes.

B. Route Optimization

The determination of adequate garbage collection routes is crucial for efficient service. This is a combinatorial optimization problem known as the Vehicle Routing Problem (VRP); it is an extension of the classical traveling salesman problem, where multiple vehicles travel through a network, starting from and returning to an origin. Its objective is to find optimal routes for multiple vehicles visiting a set of locations [16]. The VRP is a difficult problem to solve exactly, but it can also be solved with approximate methods [17]. The exact method guarantees the computation of an optimal solution but can be computationally inefficient with large numbers of nodes to visit, while the approximate method can provide a good quality solution in a reasonable time, independent of the number of nodes [18]. Route planning can be approached with the mathematical programming method whose objective is to maximize or minimize an outcome by choosing the values of real or integer decision variables [1]. For example, [19] developed a mixed-integer linear programming model that integrates a continuous perspective, using formulas to provide asymptotic estimates of routing costs, with a discrete perspective, through an iterative stochastic approximation procedure. To solve capacitated vehicle routing problems, when vehicles do not necessarily return to the starting point at the end of their trip, a model was proposed through the use of LINGO and the Branch and Bound technique to obtain the optimal routing [20]. Regarding approximate methods, [21] propose an approach based on NSGA-III, a multi-objective evolutionary algorithm taking into consideration indicators such as distance, travel time, and the number of traffic lights, to find multiple solutions to different levels of traffic congestion. Likewise, [22] seeks to address the problems related to the adequate provision of garbage collection services, through the use of a heuristic model based on the Pagerank web ranking algorithm, which is complemented by the application of constructive heuristics to the entire system. Meanwhile [23] apply the Ant Colony Algorithm (ACO) adapted to a set of routes to search for those that consume the least amount of energy, in addition to a Pareto solution that is deployed to find the optimized route in terms of energy and distance. On the other hand, [24] proposes a hybrid metaheuristic algorithm called GACO that adopts the path construction procedure of the ACO algorithm and incorporates the crossover and mutation operations used in the genetic algorithm to improve the local and global search for the solution. Some authors employ the strategy of decomposing the routing problem into two subproblems, the grouping of nodes into clusters or groups and the sequencing of nodes in each cluster. For example, [16] uses the Modified Capacity K-Means algorithms to group the new garbage cans concerning their distance and demand, and Variable Neighborhood Search to reorder the collection sequence of the garbage cans. In his work [25] uses the K-Means algorithm to group the garbage cans to form small clusters and the ACO algorithm to calculate the sequencing of the garbage cans in each cluster.

1) Clustering Problem: K-Means Algorithm: The clustering problem consists of gathering objects into groups or clusters according to some measure of similarity so that objects must be homogeneous within a group and heterogeneous in other groups [26]. Clustering techniques are used in various disciplines such as software engineering, statistics, data mining, image analysis, machine learning, web clustering engines, and text mining [27]. Among these techniques is the K-Means algorithm, which is widely used as an unsupervised learning technique [28]. This technique starts with the random selection of several centroids, where the centroid is defined as a point equidistant from the objects that belong to that group, each object is assigned to the nearest centroid, then the centroid in each group is recalculated as well as the distances of the objects to the new centroids, and new centroids are assigned again until the function F, called inertia, shown in the following expression, stabilizes [27], [29].

$$F = \sum_{k=1}^{K} \sum_{i \in C_k} \|x_i - m_k\|^2 \tag{1}$$

Notation:

- K: Number of groups.
- C_k: Partitioning of group k (set of objects that compose it).
- x_i: Object data i for partition C_k.
- m_k: Centroid of group k.

One method to compute the number of groups to be formed is the Elbow method, which evaluates the inertia shown in expression (1) so that, if you add another group, the inertia does not improve significantly [30]. Researcher in [31] argue that the choice of the number of groups will be determined at the point of highest curvature after multiple iterations.

2) Sequencing Problem: Ant Colony Optimization: Metaheuristics are defined as an iterative process that guides heuristics to explore and exploit the search space, are not specific to a particular problem, and can be applied to a wide variety of optimization problems [18]. Among these techniques is the Ant Colony Method (ACO), which is a development paradigm used to solve optimization problems, inspired by the behavior of the ant colony [17], [32]: to reach the food source, ants

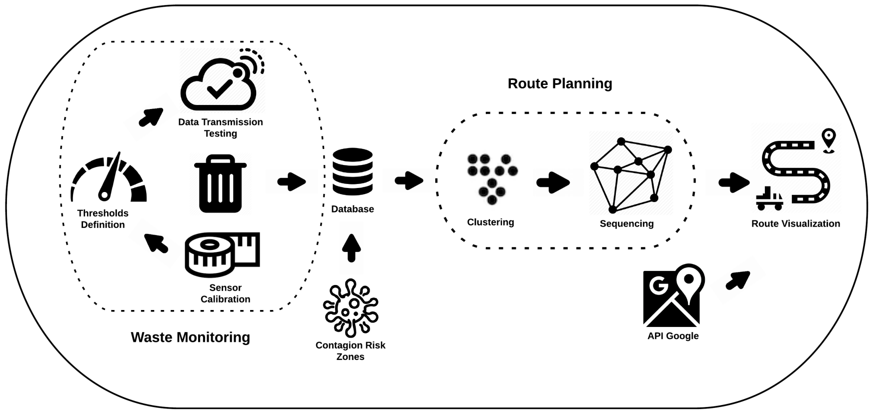


Fig. 1. Solution Diagram.

are guided by pheromone information, which is a substance that ants leave on the path as they pass through it, with ants preferring to follow the paths with the highest amount of pheromones. Artificial ants start at nodes chosen according to some initialization rule. Each ant incrementally builds a path by applying a state transition rule [33]. For this path to be visited by more ants, the local pheromone update rule is applied each time an ant crosses an arc in which, a pheromone level is calculated based on the total number of nodes and the visibility to the nearest node which are affected by a decay parameter will be added. Finally, when the ants complete their path, the global pheromone update rule is applied, this change is only performed on the best of the constructed paths. Since the Routing Problem can be solved with exact methods or with approximate methods, many authors have compared both types of methods in terms of computational efficiency. In [34], an exact solution method using the LINGO software and the Clark & Wright metaheuristic algorithm were compared, registering a processing time of 611 hours and 0.14 seconds respectively, which shows the higher computational efficiency of the latter. In the same way, [35] performed a comparison between the results obtained by applying exact methods using the LINGO program and the Simulated Annealing algorithm where vehicles must reach a customer before their competitors. For large instances, the former was not able to obtain an optimal solution in a reasonable time, while the latter achieved a quality result in an acceptable processing time. On the other hand, in [33], to minimize resources and provide efficient management of the garbage collection process focused on a vehicle routing problem with stochastic demand, a solution based on ant colony optimization was implemented and compared with the simulated annealing algorithm under different values of vehicle capacity. The results indicated that the former showed better performance for all ranges and sizes of the problem, and even has an additional advantage over the latter and the genetic algorithm because it can adapt the model to parameter changes in real-time. In this literature review, no published papers have been found that report solutions to the garbage collection problem that combine the advantages of sensors to monitor the containers and the effectiveness of existing methods that solve the vehicle routing problem. In this paper, a solution with such characteristics is proposed and will be described in the next section.

III. Methodology

As explained in the previous section, this work proposes the implementation of a network of sensors that record and transfer the information from the containers through wireless links to a digital repository. This network would consist of low cost and ease of use ultrasound sensors for waste monitoring to measure the level of waste, as well as gas sensors to identify those containers with a higher level of biodegradable waste; for data transmission, the NodeMCU device was selected to work with each container independently to avoid bottlenecks, which ensures greater availability of the monitoring system. For route planning, the use of approximate methods was chosen, through a decomposition scheme and making use of the K-means algorithm and the Ant Colony algorithm to obtain adequate solutions in an acceptable processing time. Fig. 1 shows the outline of the solution to the problem of the garbage collection consisting of two components, Waste Monitoring, and Route Planning. Therefore, the framework is composed of two stages, the first one corresponds to the implementation of the smart container, that is, the container provided with ultrasound and gas sensors that allow monitoring of the volume and CO2 concentration of the waste inside. This implementation involves sensor calibration, the definition of thresholds, and data transmission tests. The information obtained by the wireless sensor network, the location of the waste bins, and contagion risk zone will be stored in a database for later use. In the second stage, the Route Planner is implemented using a heuristic approach that decomposes the vehicle routing problem into the subproblems of container clustering and container sequencing. For clustering, the Elbow method and the K-Means algorithm will be used to obtain the appropriate number of container clusters and their distribution, respectively. On the other hand, sequencing will be implemented using the Ant Colony algorithm. Finally, the routes will be visualized on a map using Google Maps API Directions application.

A. Waste Monitoring

To obtain quality data, it is necessary to calibrate the sensors. To do this, the values obtained by the devices must be sampled, then the least-squares method is applied to obtain the linear relationship between the values provided by the sensors and the real values, obtaining the coefficients to calculate the waste and CO2 levels. Thresholds must be defined to identify the containers to be collected. In the works reviewed, the so-called vertical measurement is used in which a waste level sensor is used; in this work, the horizontal measurement of the waste level will be carried out by inserting four waste level sensors which will be placed horizontally inside the container and, depending on the dimensions of the container, the low, medium and high thresholds will be defined. For the contents of a container to be emptied, it must exceed the high threshold, but if the container is in an contagion risk zone or sufficiently close to an area at risk, it will be sufficient for it to exceed the medium threshold. On the other hand, for the gas sensor, taking into account the CO2 concentration standards in the air, a value between 250 and 400 particles per million (ppm) represents adequate air quality. Values between 400 and 1000 ppm define a reduction in air quality being able to perceive unpleasant odors and possible discomfort. And a value higher than 1000 ppm represents a critical state capable of generating serious problems in air quality and serious diseases in the locations where it is perceived [36]. Therefore, the threshold for CO2 concentration is 1000 ppm, i.e., any container exceeding this threshold will be emptied. Finally, to send data, a series of tests must be performed to confirm the transmission of information from the NodeMCU device to a database through a wireless connection to the Internet, to effectively analyze and monitor the status of the garbage containers.

B. Route Planning

1) Clustering: Previously, based on the data obtained by the Waste Monitoring system, the containers that will be identified will be emptied; a container will be attended only if its content exceeds at least one of the established thresholds. Subsequently, the K-Means algorithm will be applied to group these containers to be served in K clusters, based on the distances between them. The algorithm is as follows:

1) Randomly select K centroids
2) Repeat while there are new centroid assignments
    a) Repeat for each node
        i) Find the nearest centroid
        ii) Assign the node to the cluster of that centroid

        End repeat
    b) Repeat for each cluster $k \leftarrow 1$ to K
        i) New centroid $\leftarrow$ "Average of the elements assigned to the cluster"

        End repeat

    End repeat

2) Sequencing: After grouping the containers, the routes or sequences of attention of the containers in each of the groups or clusters will be defined using the Ant Colony metaheuristic. The corresponding algorithm to build a route for each cluster is shown below [37].

1) Initialize parameters and pheromone level in each arc
2) Repeat while the iteration limit is not exceeded
    a) Repeat until each ant has built a solution
        i) Repeat for each ant k
            A) Choose the next node by applying the state transition rule
            B) Apply local pheromone update rule

            End repeat

        End repeat
    b) Update the best solution
    c) Apply global pheromone update rule

    End repeat

In this algorithm, during the construction of a route, the ant k currently positioned at node i, moves to a node n according to the following state transition rule: A random variable q between 0 and 1 is evaluated, and if $q > q_0$ then it is chosen to the next node n, through a biased scan of the arcs, based on expression (2), otherwise, a probabilistic random proportional rule is applied that uses the best weight, product of the pheromone level and the visibility in the arc, to move to a next node n, provided that it belongs to $N_i k$, described in expression (3) [28].

$$p_{in}^{k} = \frac{(\tau_{in})^{\alpha} * (\eta_{in})^{\beta}}{\sum_{l \in N_i^k} (\tau_{il})^{\alpha} * (\eta_{il})^{\beta}} \tag{2}$$

$$n = \arg\max_{j \in N_i^k} \left\{ (\tau_{ij}) * (\eta_{ij})^{\beta} \right\} \tag{3}$$

Notation:

- τ_{in}, represents the level of pheromones in the arc (i, n).
- η_{in}, represents visibility (inverse of distance/cost) in the arc (i, n).
- N_i^k, represents the list of previously unvisited nodes that the ant k, positioned at node i, can access.

IV. Results

The San Borja district of the city of Lima was selected as a test case for the experimentation since it had 74 underground containers, which are suitable for inserting the monitoring sensors since they are not exposed to the environment or the public [38]. However, due to the impossibility of experimenting with real containers, due to the health situation of the pandemic, a dataset (waste level) of 32 containers, generated by an Australian government project and applied to Wyndham City Council, was used [39]. Therefore, 32 of the 74 containers in the district of San Borja were randomly selected and assigned the data from the dataset. As for the contagion risk zones the "Peru en tus Manos" application was used to determine the exact locations with confirmed cases of COVID-19. Using the Google Maps API Distance Matrix application, the distances between the containers and between the areas at risk of infection and the containers were obtained.

A. Smart Container

1) Waste Level Sensor: First, an algorithm was developed to calculate the distance between the objects and the ultrasonic sensor, a cylindrical test container, 34 cm high and 20 cm in diameter, was experimented with, and a sample of 84 values measured both real and by the sensor was taken, as shown in Fig. 2 the relationship between these values is approximately a linear function ($y = a*x+b$). Subsequently, the least-squares

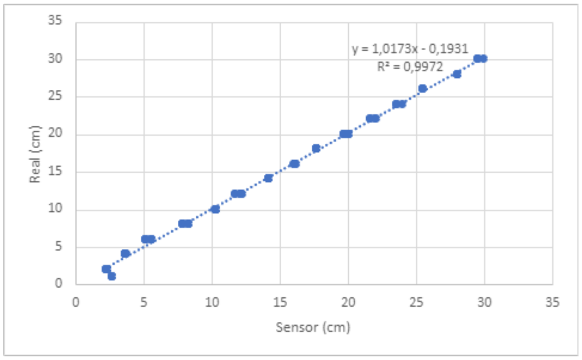


Fig. 2. Relationship Between Sensor Measurement and Real Measurement.

method was applied to obtain the values of the coefficients a and b of this function. Table I shows the results of the application of the developed algorithm and the Ultrasonic 1 [40] and Ultrasonic 2 [41]. In Table I it can be seen that the percentage error of the algorithm developed is 0.1%, while the library errors are 0.14% y 0.13% respectively, which shows that the algorithm developed provides a better approximation to the real distance. For experimental purposes, four ultrasonic sensors, a gas sensor, a NodeMCU device, and a portable battery were incorporated into the test container to monitor the level and decomposition of waste. The distribution of the ultrasound sensors was set horizontally about the dimensions of the garbage bin, i.e., for the low level at 3 cm from the base, for the medium level two sensors were placed at 1/3 and 3/5 of the height of the garbage can, finally, for the high level at 5 cm below the height of the garbage can alert the collection action before the garbage can overflow. As shown in the left container in Fig. 3, the sensors are distributed on the sides of the container according to the defined thresholds. Likewise, the present solution called horizontal measurement

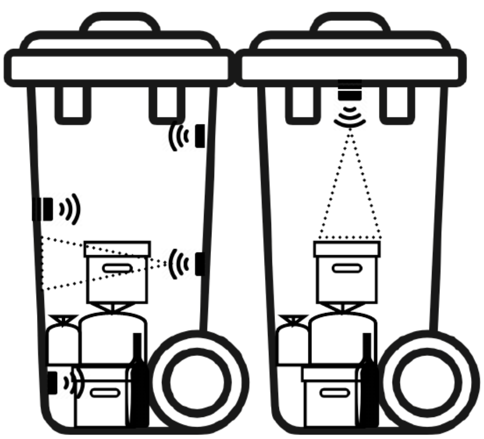


Fig. 3. Horizontal and Vertical Measurement of Container Contents.

was compared with the classic vertical measurement proposed by other authors [42]. In the right container in Fig. 3, a sensor located in the lid of the container can be seen to perform the vertical measurement. The results are shown in Table II, where the real measurement indicates the different levels of waste inside the container, symbolized by L, M, and H (low, medium, and high) respectively; the real location refers to the position in which the waste is distributed, which can be in the center, to one side or uniformly, symbolized by C, L, and U, respectively. Finally, the horizontal and vertical measurements are the waste levels obtained by the sensors for the respective methods. According to Table II, the horizontal measurement reduces the number of erroneous results by 23%; the errors in the vertical measurement are generated because the value H or M can be obtained instead of M or L respectively since the waste is not necessarily uniformly distributed within the garbage bin. For example, in Fig. 3, the horizontal measurement would give the value B (low level of waste), while the vertical measurement would give M, when in fact it is L. It can be stated that with horizontal measurement better results are achieved than with vertical measurement. Moreover, adequate results are achieved by using low-cost sensors.

2) CO2 Level Sensor: A preheating stage was performed for 24 hours to establish the appropriate environmental conditions and eliminate excess impurities or humidity in the sensor [43]. Next, the sensor output resistance (R_s) was measured over five minutes to obtain an average equivalent of 13480.265, which will be the final value of R_s. Then, the CO2 gas values specified in the sensor datasheet were used, obtaining an exponential function ($y = a * x^b$), shown in Figure 4, where $x = R_s/R_o$, where R_o is the initial resistance. Subsequently, the least-squares method was applied, obtaining the values of the coefficients a and b, so that $y = 114.308 * (R_s/21285.605)^{-2.829}$. Table III shows the

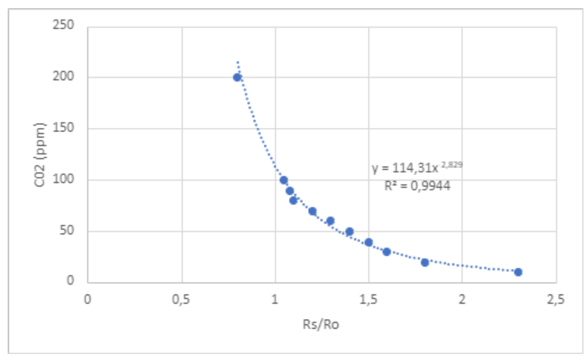


Fig. 4. CO2 Gas Measurement Curve.

average values obtained with the algorithm developed and with the MQ135 Library [44] for three scenarios: Scenario 1 refers to a container with organic waste, in Scenario 2 a container with non-organic waste was used and in Scenario 3 an empty container was used. While in scenarios 1 and 2 the variations of the results are 10.76% and 12.98% respectively, in scenario 3 a significant difference of 35.14% is observed. However, the average concentration of CO2 in free air is between 250 and 400 ppm [36], so when comparing the results in this scenario, it can be verified that the algorithm yields adequate values that are in this range.

B. Route Planning

1) Clustering: With respect to the sample of 32 containers indicated above, only 12 exceeded the corresponding thresholds to be served. The Elbow method was used to determine the appropriate number of clusters to use in the clustering. Fig. 5 shows the plot of the inertia function concerning the

TABLE I. Results of the elaborated algorithm and libraries for the ultrasound sensor

Test Number	1	2	3	4	5	6	7	8	9	10	11	12
Real Values (cm)	20	18	16	14	12	10	8	6	4	2	1	0
Elaborated Algorithm (cm)	20	18	16	14	12	10	8	5	4	2	2	1
Ultrasonic Library 1 (cm)	19	17	16	14	12	10	7	5	4	2	2	51
Ultrasonic Library 2 (cm)	20	18	16	14	12	10	8	5	4	2	2	357

TABLE II. Horizontal and Vertical Measurement Results

Test Number	1	2	3	4	5	6	7	8	9	10	11	12	13
Real Measurement	L	L	L	L	L	M	M	M	M	M	M	H	H
Real Location	U	C	U	C	L	U	C	L	U	C	L	U	C
Horizontal Measurement	L	L	L	L	M	L	L	M	M	M	H	H	H
Vertical Measurement	L	M	M	H	M	M	H	M	M	H	H	H	H

TABLE III. Results of the Elaborated Algorithm and the Library for the Gas Sensor

Scenario	1	2	3
Elaborated Algorithm (ppm)	9578.98	613.41	318.83
MQ135 Library (ppm)	8648.64	533.81	491.59
Variation (%)	10.76	12.98	35.14

number of clusters where a sharp change in slope, similar to that of an arm and its elbow, is observed between two and three clusters, given that in the following numbers of clusters no substantial improvement in inertia is observed. Therefore, the 12 trash containers will be gathered in two clusters (which also coincides with the Municipal Tax Ordinance of San Borja [45]). Using the K-means algorithm, a group of the 12 garbage

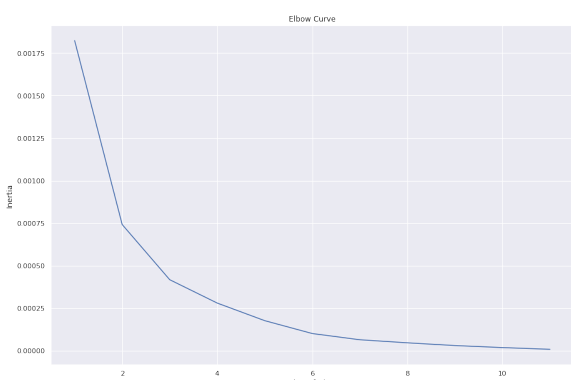


Fig. 5. Elbow Curve.

cans into two clusters of 6 garbage cans each was obtained as shown in Fig. 6. In each cluster, one more node appears which is the centroid, which is not an element of the cluster but is calculated as the average of the cluster elements.

2) Sequencing: 5000 iterations of the Ant Colony algorithm was run to find the best sequence of containers to serve, by optimizing each of the following two indicators: distance traveled and travel time separately, and for each of the two groups. Table IV shows the values obtained for each indicator in each group: the minimum travel distance to empty all the containers in group 1, starting from the Lurin depot and finally returning to the same depot is 101.122 kilometers, which requires the collection vehicle to be used for a time of 2.11 hours with a fuel cost of 61.4821 soles, calculated as the

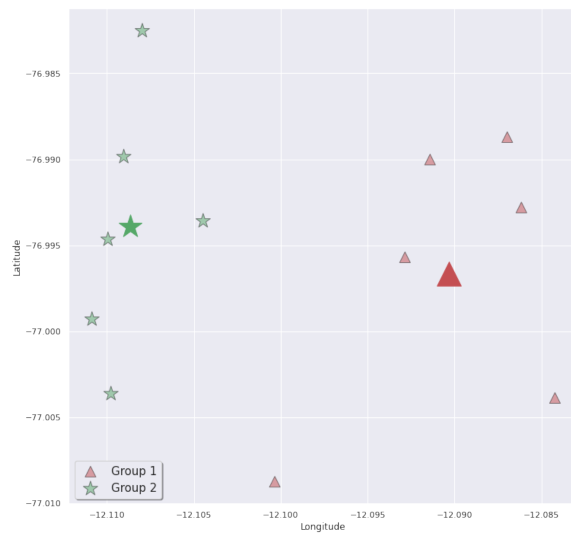


Fig. 6. Container Clustering.

product of the travel time and the unit cost of fuel. Likewise for group 2 with its corresponding indicators. Fig. 7 shows the

TABLE IV. Route Planning Indicators

Indicator	Group 1	Group 2
Distance traveled (km)	101.122	97.015
Fuel cost (S/.)	61.482	58.985
Travel time (Hs)	2.11	2.12

sequence in which the containers of group 1 are visited by the collection truck that starts its route in the garbage depot located in the district of Lurin, being container A and F the first and the last to be emptied. To visualize this route, various tools provided by Google Maps, such as API Directions, markers, etc. were used.

V. Discussion

To evaluate the results obtained by the proposed solution, we proceeded to the exhaustive enumeration (EE) of the routes for each of the two groups of 6 containers each that were obtained in the clustering phase, which allowed us to calculate the corresponding optimal values of the 2 indicators distance traveled and travel time, shown in Table V. Two additional scenarios were evaluated, scenario 2 with two groups of 8

TABLE V. Optimization by ACO and EE for Each Group and Each Scenario

Indicator	Distance traveled (km)						Travel time (Hs)					
Scenario	1		2		3		1		2		3	
Group	1	2	1	2	1	2	1	2	1	2	1	2
EE Optimal Value	100.807	96.830	97.971	98.608	98.440	97.831	2.03	1.99	2.12	1.87	1.93	2.11
ACO Optimal Value	101.122	97.015	98.361	98.608	99.097	97.831	2.11	2.12	2.18	1.87	1.99	2.11
Error (%)	0.31	0.19	0.40	0	0.67	0	3.84	6.70	2.43	0	2.43	0

TABLE VI. Optimization by ACO and EE for All Containers and Each Scenario

Indicator	Total distance traveled (km)			Total vehicle usage time (Hs)			Total travel time (Hs)		
Scenario	1	2	3	1	2	3	1	2	3
EE Optimal Value	197.637	196.579	196.271	4.02	3.99	4.03	2.03	2.12	2.11
ACO Optimal Value	198.137	196.969	196.928	4.23	4.04	4.10	2.12	2.18	2.11
Error (%)	0.253	0.198	0.335	5.25	1.29	1.70	4.38	2.43	0

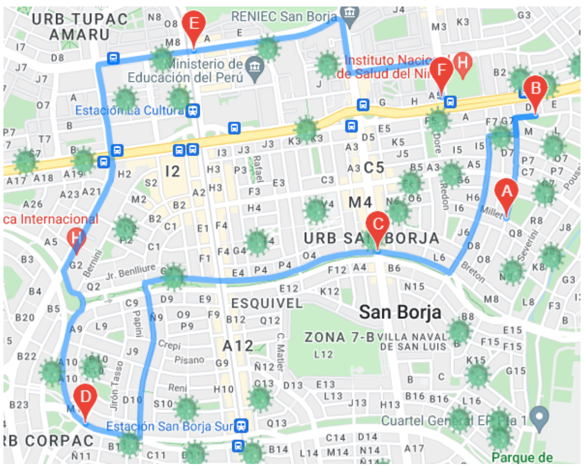


Fig. 7. Route for Group 1.

and 4 containers, and scenario 3 with two groups of 7 and 5 containers respectively. In scenario 1, 0.31% and 0.19% error was obtained in the results obtained by the ant colony algorithm (ACO) with respect to the exhaustive enumeration (EE) results about the distance traveled indicator for group 1 and 2, respectively, while for the travel time indicator, errors of 3.84% and 6.70% were obtained for group 1 and 2. Lower percentages error were observed in scenarios 2 and 3 for the distance traveled indicator. These results are evidence of the quality of the solutions computed by the ACO metaheuristic and are reported in a multitude of papers, some of which were referred to in section 2. Although acceptable results have been obtained for each of the two groups, it is necessary to evaluate the integrated solution that covered all 12 containers, analyzing the indicators of total distance traveled, total vehicle use time, and total travel time (time taken for the garbage collection process); the three indicators are referred to the set of all containers and in the same three scenarios defined in the previous analysis. The indicators are shown in Table VI as well as the percentages error. A fourth indicator would be the fuel cost indicator, however; as it is directly related to the distance indicator, it will have the same percentage error. In scenario 1 we noticed that the optimal ACO value of the total travel distance (and fuel costs), has an error of 0.253% with respect to the EE Optimal Value, while for the indicators of time of use of the collection vehicles and total travel time the errors were 5.25% and 4.38%, respectively. Improvements in the percentages error can be seen in the other scenarios.

VI. Conclusion

Taking into account the limitations of the experimentation due to the sanitary situation, and the quality of the results obtained, it can be affirmed that the objective of presenting a solution that optimizes the resources used in the collection of solid waste, keeping in mind the impact on public health, has been adequately fulfilled. On the one hand, the use of wireless sensors allows efficient monitoring of waste and contamination levels inside a container, since it is no longer necessary to check the garbage bins in situ to know the state they are in. Better quality data was obtained with the algorithms developed than with the standard libraries, as a lower percentage error was achieved in the measurement of waste and contamination levels. Likewise, evidence was found that through the horizontal measurement, the number of erroneous results generated by the vertical measurement is reduced, since not only the maximum level is considering when the waste is not uniformly distributed within the containers. In summary, by using low-cost devices, adequate results have been obtained. On the other hand, the low percentage error obtained with the model based on the K-Means algorithm and the Ant Colony metaheuristic evidences the quality of the results concerning parameters such as container location, location of contagion risk zones, fuel cost, thresholds for waste level and contamination inside the container, etc. The implementation of Route Planning using a heuristic approach guarantees its scalability by providing adequate results using a reasonable execution time. Moreover, by integrating the basic Waste Monitoring system with the Route Planning system, a continuous flow of information has been established to ensure traceability of results. This integration allows for agile planning of garbage collection since route planning is based on online information. Finally, the integrated system can be transformed into an online system, i.e., while the garbage collection is in progress, a container not scheduled to be collected might exceed one of the alert thresholds. In this case, the remaining containers to be collected would be rescheduled.

VII. Future Work

Future work includes the incorporation of special sensors in the smart container to monitor the humidity and temperature inside the containers since the speed of sound and the concentration of carbon dioxide (CO2) are sensitive to these variables. To preserve the energy and mobility of the smart container, it recommends using an independent long-life battery (LIPO), as well as a library that is characterized by setting the ESP8266 microcontroller in sleep mode to reduce the battery consumption and guarantee a long lifetime for the smart container. The proposed solution can be transformed into an online system through real-time route planning so that if any container not scheduled to be collected reaches a monitoring threshold during the route of the collection trucks, it will be integrated into the most convenient collection route. This will provide more timely attention. The proposed route planning model is essentially a Vehicle Routing model (VRP). This model could be made more realistic by taking into account other parameters that affect actual waste collection problems, e.g. number of vehicles, vehicle capacity, time windows, etc. Its application to larger cases is recommended due to the scalability of the Route Planning model.

REFERENCES

[1] A. Sulemana, E. A. Donkor, E. K. Forkuo, and S. Oduro-Kwarteng, "Optimal routing of solid waste collection trucks: A review of methods," *Journal of Engineering*, vol. 2018, p. 4586376, Oct. 2018.

[2] Ministerio del Ambiente, "Sistema Nacional de Información Ambiental." 2020. [Online]. Available: https://sinia.minam.gob.pe/informacion/tematicas?tematica=08

[3] TomTom. (2019) Traffic index results. [Online]. Available: https://www.tomtom.com/en_gb/traffic-index/ranking/

[4] Ministerio de Salud, "Sala situacional covid-19 Perú," 2020.

[5] Instituto Nacional de Estadística e Informática, "Un análisis de la eficiencia de la gestión municipal de residuos sólidos en el perú y sus determinantes," 2018.

[6] A. A. R. Alsaeedy and E. K. P. Chong, "Detecting regions at risk for spreading covid-19 using existing cellular wireless network functionalities," *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 1, pp. 187–189, 2020.

[7] M. Byers, M. Russell, and D. J. Lockey, "Clinical care in the "hot zone"," *Emergency Medicine Journal*, vol. 25, no. 2, pp. 108–112, 2008. [Online]. Available: https://emj.bmj.com/content/25/2/108

[8] A. J. Y. Tan, H. E. Lee, S. M. Ng, and H. S. Chua, "Smart iot-based waste monitoring system," in *2019 International UNIMAS STEM 12th Engineering Conference (EnCon)*, 2019, pp. 40–45.

[9] P. Nehete, D. Jangam, N. Barne, P. Bhoite, and S. Jadhav, "Garbage management using internet of things," in *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2018, pp. 1454–1458.

[10] K. Nirde, P. S. Mulay, and U. M. Chaskar, "Iot based solid waste management system for smart city," in *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2017, pp. 666–669.

[11] U. Draz, T. Ali, J. A. Khan, M. Majid, and S. Yasin, "A real-time smart dumpsters monitoring and garbage collection system," in *2017 Fifth International Conference on Aerospace Science Engineering (ICASE)*, 2017, pp. 1–8.

[12] M. A. Al Mamun, M. A. Hannan, and A. Hussain, "Real time solid waste bin monitoring system framework using wireless sensor network," in *2014 International Conference on Electronics, Information and Communications (ICEIC)*, 2014, pp. 1–2.

[13] S. Murugaanandam, V. Ganapathy, and R. Balaji, "Efficient iot based smart bin for clean environment," in *2018 International Conference on Communication and Signal Processing (ICCSP)*, 2018, pp. 0715–0720.

[14] A. S. Bharadwaj, R. Rego, and A. Chowdhury, "Iot based solid waste management system: A conceptual approach with an architectural solution as a smart city application," in *2016 IEEE Annual India Conference (INDICON)*, 2016, pp. 1–6.

[15] M. Gokhale, R. Wagh, P. Chaudhari, S. Khairnar, and S. Jadhav, "Iot based e-tracking system for waste management," in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, 2018, pp. 1–6.

[16] V. Sreelekshmi and J. J. Nair, "Dynamic vehicle routing for solid waste management," in *2017 IEEE Region 10 Symposium (TENSYMP)*, 2017, pp. 1–5.

[17] M. Dorigo, M. Birattari, and T. Stutzle, "Ant colony optimization," *IEEE Computational Intelligence Magazine*, vol. 1, no. 4, pp. 28–39, 2006.

[18] A. Gogna and A. Tayal, "Metaheuristics: review and application," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 25, no. 4, pp. 503–526, 2013. [Online]. Available: https://doi.org/10.1080/0952813X.2013.782347

[19] C. Lei, Z. Jiang, and Y. Ouyang, "A discrete-continuous hybrid approach to periodic routing of waste collection vehicles with recycling operations," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 12, pp. 5236–5245, 2020.

[20] Y. Hartono, F. M. Puspita, D. I. Permatasari, and B. Arisha, "Lingo-based on robust counterpart open capacitated vehicle routing problem (rc-ocvrp) model of waste transportation in palembang," in *2018 International Conference on Information and Communications Technology (ICOIACT)*, 2018, pp. 429–435.

[21] R. D. N. de Morais and C. Garrozi, "Multi-objective sequential route optimization applied in large-scale road networks with traffic congestion," in *Proceedings of the XIV Brazilian Symposium on Information Systems*, ser. SBSI'18. New York, NY, USA: Association for Computing Machinery, 2018. [Online]. Available: https://doi.org/10.1145/3229345.3229388

[22] D. G. Rossit, J. Toutouh, and S. Nesmachnow, "Exact and heuristic approaches for multi-objective garbage accumulation points location in real scenarios," *Waste Management*, vol. 105, pp. 467–481, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0956053X2030074X

[23] A. Anuntachai, A. Thammano, and O. Wongwirat, "Searching optimization route by using pareto solution with ant algorithm for mobile robot in rough terrain environment," in *2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, 2016, pp. 1–6.

[24] R. J. Kuo and F. E. Zulvia, "Hybrid genetic ant colony optimization algorithm for capacitated vehicle routing problem with fuzzy demand — a case study on garbage collection system," in *2017 4th International Conference on Industrial Engineering and Applications (ICIEA)*, 2017, pp. 244–248.

[25] S. K. Nambiar and S. M. Idicula, "A multi-agent vehicle routing system for garbage collection," in *2013 Fifth International Conference on Advanced Computing (ICoAC)*, 2013, pp. 72–76.

[26] S. Kapil, M. Chawla, and M. D. Ansari, "On k-means data clustering algorithm with genetic algorithm," in *2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, 2016, pp. 202–206.

[27] L. Xu, A. Apon, F. Villanustre, R. Dev, and A. Chala, "Massively scalable parallel kmeans on the hpcc systems platform," in *2019 4th International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS)*, vol. 4, 2019, pp. 1–8.

[28] M. Reed, A. Yiannakou, and R. Evering, "An ant colony algorithm for the multi-compartment vehicle routing problem," *Applied Soft Computing*, vol. 15, pp. 169–176, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494613003517

[29] T. N. Reddy and K. P. Supreethi, "Optimization of k-means algorithm: Ant colony optimization," in *2017 International Conference on Computing Methodologies and Communication (ICCMC)*, 2017, pp. 530–535.

[30] H. Asri, H. Mousannif, and H. Al Moatassime, "Reality mining and predictive analytics for building smart applications," *Journal of Big Data*, vol. 6, no. 1, p. 66, Jul. 2019.

[31] M. Antunes, D. Gomes, and R. L. Aguiar, "Knee/elbow estimation based on first derivative threshold," in *2018 IEEE Fourth International Conference on Big Data Computing Service and Applications (BigDataService)*, 2018, pp. 237–240.

[32] D. H. Dayanara, N. I. Arvitrida, and N. Siswanto, "Vehicle routing problem with split service, time window and intermediate facility for municipal solid waste collection in surabaya city with ant colony optimization algorithm," *IOP Conference Series: Materials Science and Engineering*, vol. 598, no. 1, p. 012020, aug 2019. [Online]. Available: https://doi.org/10.1088/1757-899x/598/1/012020

[33] Z. Ismail and S. L. Loh, "Ant colony optimization for solving solid waste collection scheduling problems," *Journal of Mathematics and Statistics*, vol. 5, pp. 199–205, 03 2009.

[34] V. V. Panicker and I. O. Mohammed, "Solving a heterogeneous fleet vehicle routing model - a practical approach," in *2018 ieee international conference on system, computation, automation and networking (icscan)*, 2018, pp. 1–5.

[35] R. Tavakkoli-Moghaddam, M. Gazanfari, M. Alinaghian, A. Salamatbakhsh, and N. Norouzi, "A new mathematical model for a competitive vehicle routing problem with time windows solved by simulated annealing," *Journal of Manufacturing Systems*, vol. 30, no. 2, pp. 83–92, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0278612511000380

[36] Wisconsin. (2019, 12) Wisconsin department of health services. [Online]. Available: https://www.dhs.wisconsin.gov/

[37] D. N. Kumar and M. J. Reddy, "Ant colony optimization for Multi-Purpose reservoir operation," *Water Resources Management*, vol. 20, no. 6, pp. 879–898, Dec. 2006.

[38] Municipalidad de San Borja, "Plan Urbano Distrital de San Borja 2020-2030," 2019.

[39] Australian Government. (2018) Wyndham smart bin daily fill level. [Online]. Available: https://data.gov.au/data/dataset/wyndham-smart-bin-fill-level-historical/

[40] J. Rodrigo. (2015, 09) Ultrasonic-hc-sr04. [Online]. Available: https://github.com/JRodrigoTech/Ultrasonic-HC-SR04

[41] E. Simões. (2019, 03) Ultrasonic. [Online]. Available: https://github.com/ErickSimoes/Ultrasonic

[42] T. Addabbo, A. Fort, A. Mecocci, M. Mugnaini, S. Parrino, A. Pozzebon, and V. Vignoli, "A lora-based iot sensor node for waste management based on a customized ultrasonic transceiver," in *2019 IEEE Sensors Applications Symposium (SAS)*, 2019, pp. 1–6.

[43] F. M. Millan, "Diseño e implementación de un sistema de medida de gases con arduino." pp. 17–18, 2016.

[44] G. Krocker. (2014, 12) Mq135. [Online]. Available: https://github.com/GeorgK/MQ135

[45] El Peruano, "Ordenanza que Aprueba la Estructura de Costos y Régimen Tributario de los Arbitrios de Limpieza Pública, Parques y Jardines y Seguridad Ciudadana para el Ejercicio 2019," 2019.

Empirical Analysis of Learning-based Malware Detection Methods using Image Visualization

Abdullah Sheneamer
Department of Computer Science
Jazan University
Jazan, Saudi Arabia

Essa Alhazmi
Department of Computer Science
Jazan University
Jazan, Saudi Arabia

James Henrydoss
Vision and Security Technology Lab
University of Colorado
Colorado Springs, USA

***Abstract*—Malware, a short name for malicious software is an emerging cyber threat. Various researchers have proposed ways to build advanced malware detectors that can mitigate threat actors and enable effective cybersecurity decisions in the past. Recent research implements malware detectors based on visualized images of malware executable files. In this framework, a malware binary is converted into an image, and by extracting image features and applying machine learning methods, the malware is identified based on image similarity. In this research work, we implement the Image visualization-based malware detection method and conduct an empirical analysis of various learners for selecting a candidate learning classifier that can provide better prediction performance. We evaluate our framework using the following malware datasets, Search And RetrieVAl of Malware (SARVAM), Xue-dataset, and Canadian Institutes for Cyber Security (CIC) datasets. Our experiments include the following learning algorithms, Linear Regression, Random Forest, K-Nearest Neighbor (KNN), Classification and Decision Tree (CART), Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), and deep learning-based Convolutional Neural Network (CNN). This image-visualization-based method proves to be effective in terms of prediction accuracy. Some conclusions emerge from our initial study and find that a Convolutional Neural Network (CNN) algorithm provides relatively better performance when used against SARvAM and various malware datasets. The CNN model achieved a high performance of F1-score and accuracy in the binary classification task reaching 95.70% and 99.50%, consecutively. The model in the multi-classification task achieved of 95.96% and 99.30% (F1-score and accuracy) for detecting malware types. We find that the KNN model outperforms other traditional classifiers.**

Keywords*—*Malware detection; malware analysis; deep learning; machine learning; malware features

I. Introduction

Malware attacks are continuously evolving, and the spread of malware is ubiquitous and unstoppable. Attackers spread the malware through the Internet, email, and social media for the primary purpose of harming computers with a fraudulent intention. Typically, malware is an umbrella name for a set of malicious software, including viruses, worms, Trojan horses, spyware, etc. and are created to cause extensive damage to either data or systems or gain unauthorized access to a system or network. Cisco's recent cyber security threat report explores how cybercriminals exploited by building coordinated multi-step attacks using the following four types of attacks: crypto mining, phishing, Trojans, and Ransomware. These attacks received ten times more queries than any other type of attack. Building advanced software defense methods to mitigate the spread of malware is an absolute necessity. Malware detection is a process of detecting the presence of malware on a host system or identifying whether a particular program file is malicious or not malicious, i.e., benign [1]. The malware detectors play a crucial role in building early warning systems to thwart any attacks and prevent hackers from using computer systems during the zero-day attack period.

In the past, various research works built a plethora of malware detection methods that address techniques to implement advanced malware detectors. Nevertheless, malware authors continuously innovate new ways to penetrate the defense mechanism. A vast amount of research work exists that formulates the signature-based, behavioral-based, and machine learning-based malware detection approaches [2], [3], [4], [5], [6]. In a signature-based detection, each file is analyzed, assigned a signature or hash (a unique alphanumeric way to identify malware), and then added to the signature database, where it's used for comparison in subsequent malware incidents. This technique identifies specific patterns in the application to determine whether the file is malicious by verifying against a known set of signatures for matching patterns. These patterns can be syntactic, e.g., the sequence of instructions, or semantic-based, e.g., control or data flow properties. One of the critical aspects of any malware detection system is identifying whether a file is malicious or benign. The signature-based detection is a simple and widely used method built as AV (Antivirus) and malware detectors. A behavioral-based technique implements a dynamic environment where malware binaries are executed in a sandbox machine to extract the run-time characteristics to identify the malware. In a static or code analysis-based malware detection technique, the malware can be classified by using the code structure of the malicious code, e.g., use of control Flow Graphs (CFG) to identify the malware. The static method is the process of analyzing malware/binary without executing it. Its main objective is to extract useful information from the malware code structure, and it helps us get an idea of the type of malware and what the malware can do.

Signature-based approaches allow security analysts to identify the malicious component quickly, and so they are widely used by several commercial anti-virus (AV) companies. However, this technique's primary disadvantage is that they require a trained security analyst to manually write appropriate signatures that can be used to detect each malware family and load them in a signature database for run-time access. Unfortunately, this manual effort is error-prone and as well as time-consuming. Also, storing the whole world of signature-learning is a daunting task [7], [8], [9], [10]. Besides, when the malware mutates, the original stored signatures of malware become obsolete and unable to successfully detect the newly mutated malware due to the polymorphic and metamorphic methods that implement code obfuscation [11].

In contrast to signature-based detection, the machine learning-based (ML) technique aims to address this limitation of manual intervention and introduce automation by implementing a set of ML-based malware classifiers. Machine learning algorithms in the classification of malware rely on features extracted from binary files or a disassembled assembly code by using either the static code analysis techniques or dynamic analysis techniques where malware behavioral characteristics are studied during run-time and at the point of execution. This information is helpful for future analysis as it will allow us to analyze the sample efficiently. These techniques extract various features from the malicious samples and use standard machine learning algorithms to learn a classifier that labels the sample as either benign or malicious. However, there are still some challenges such as processing a large amount of malware, learning high-dimensional vectors, high storage usage, and low scalability in learning. Traditional approaches to malware detection using automatic classification are facing some limitations. The first one concerns feature extraction: static approaches are hindered by code obfuscation techniques, while dynamic methods are time-consuming, and evasion techniques often impede the correct execution of the code. The second limitation regards the building of the prediction models: the completeness of a training dataset may degrade over time as the malware authors evolve new techniques or can not be sufficient for some malware families or instances [12]. In addition, many malware runs independent of the operating system and executes its malicious code even before loads, running malicious.

The rest of the paper is organized as follows. Section II discusses related work. Our proposed framework are introduced in Section III. In Section IV, the framework's experiments are described. Discussion are covered in great details in Section V. Finally, the paper is concluded in Section VI.

II. Related Work

This section includes the relevant research associated with this project work comprising visualizing malware binaries and Machine-Learning (ML) based malware detection approaches. Malware detection has been widely practiced in the past decade. The detection approaches depend primarily on the following techniques: static - that relies on analyzing the code structure, call flow graphs of malware, the dynamic technique that addresses the behavioral characteristics of malware at run time execution in a sandbox environment during zero-day attack period [13]. Nataraj et al. [14] implemented the first framework on visualizing and classifying malware using image processing techniques. In their work, malware binaries are visualized as gray-scale images and observed that malware from the same families exhibited similarity in texture and layout. The authors conducted experiments on over 9,458 samples from 25 families of malware using image visualization and texture analysis. Nataraj et al. [15] also extended the approach to have malware binaries are represented as signals or images, and signal processing-based features are used to characterize malware. Han et al. [16] implemented visualization of malware images and performed a similarity calculation between images of malware variants. In this new classification method, they proposed a new classifier by first converting malware into gray-scale images and then applied a histogram similarity measurement to study the similarity of gray-scale image entropy maps. Han et al. [17] proposed a new malware family classification method that converts malware binary files into images and entropy graphs. Xue et al. [18] built a homology-based malware analysis using an ensemble of learning methods. Xue used gray-scale images, RGB color images, opcode sequences, and system flow graph-based image visualization methods and used Convolutional Neural Networks (CNNs) as base learners to perform bagging ensemble learning that extract features from malware images. Liu et al. [19] proposed an automatic malware classification and a new malware detection scheme using a clustering-based machine learning method. They implemented a new malware detection using Opcode n-gram based gray-scale images and feature extraction with Shared Nearest Neighbor (SNN)-based clustering algorithm on discovering new malware. Fu et al. [20] proposed visualizing malware using color images and extracted global texture and color features from the images for classification. A series of unique byte sequences are extracted from code and data sections of malware and using simhash functions converted to local features. K-Nearest Neighbors, SVM, and Random Forest methods are used to classify the malware. Makandar et al. [21] implemented malware classification methods that apply image processing techniques that use image textures-based features extraction from visualized malware binaries. Multi-resolution and wavelets are used to build effective texture feature vectors using Gabor Wavelet, GIST, and Discrete Wavelet Transform. They proposed using Support Vector Machine (SVM) based multi-class malware image classification. Singh et al. [22] implemented a CNN-based deep-network in building visualization-based malware detection methods. The use of ensemble learning techniques in malware detection is not new. Zhang et al. [23] implemented malicious code detection using multiple classifiers fusion and is not strictly dependent on specific malicious code. Menahem et al. [24] improved the prediction performance of malware detectors by combining the results of the individual classifiers into one final result to achieve overall higher detection accuracy. Recent research by [25] implemented an ensemble classification scheme of using both binary and multi-class classification as part of implementing intrusion detection solutions. There exists incomplete knowledge of class examples present during the training time. Scheirer et al. [26] introduced the open-set-based recognition method for a computer vision problem. The open-set classifiers are expected to detect all unknown classes present during testing but are not known to the model during training time. Open-set recognition describes the scenario in such a way that new classes (unknown unknown classes) unseen in training appear in testing, and requires the classifiers not only to accurately classify the *known classes*, but also to effectively deal with the *unknown* ones. In a malware survey paper, Rudd et al. [13] identified the following six flawed assumptions to use an open-set based malware detection instead of a closed-set one: intrusions are closed-set, anomalies imply class labels, static models are sufficient, no feature space transformation is required, model interpretation is optional, and class distributions are Gaussian. Henrydoss et al. [27]

implemented an open-set-recognition-based intrusion detection scheme.

Han et al. implemented a feature extraction method of malware binary using texture fingerprints index structure. This method achieves the highest accuracy of 85.77 %. Han et al., proposed a malware visualization method that uses both the static and dynamic code analysis technique with RGB color images. Using RGB image Opcode is generated from the malware by executing it, converting that opcode sequence into images for classification. The accuracy of this method yielded 98.96%. Wang et al. [28] addressed the problem of using small training sets. Opcode sequence extracted from the malware binary file was converted to an image and then normalized by histograms. Dilated and eroded PCA is applied to extract features for SVM-enabled classification. Tobiyama et al. [29], Kolosonjaji B et al. [30], Zhao et al.[23] implemented deep learning techniques CNN to classify malware images and achieve a prediction accuracy of 96%, 85.6%, and 96.7% respectively. These experiments were performed on a few malware samples, and these techniques are not evaluated against larger datasets. Liu et al., Azia Makandar et al., Huang et al. implemented malware classification based-on image analysis using multiple features, e.g., use of binary files, opcode sequences, and API call sequences, and resulted in an accuracy of 98.9%, 98.8%, and 99.51%. These methods used an advanced feature set ranging from 4000 to 50000 feature sets. Huang et al. [31], provided the state of the art classification performance evaluated against a large dataset that contains a training set of 4.5 million and a testing set of 2 million samples.

III. Framework

A. Malware Analysis using Image Visualization

Traditional malware detection methods are based on static and dynamic analysis techniques. The static analysis technique involves analyzing the disassembled code of a malware file. The dynamic code analysis and run-time-based methods use malware behavioral characteristics at the point of execution in a sandbox test environment. We adopt an image-visualization-based technique to implement malware detectors in this work. Basically, instead of detecting the malware binary as is where we convert them into an image and use computer vision techniques to recognize the image and detect the appropriate malware family. Visualizing malware is a new type of malware detection method. Using image processing is a simple but very effective method [14] of malware detection. Nataraj et al. [14] propose the initial framework of malware analysis using an image visualization-based malware detection technique. This method converts a malware binary into an image, using image processing techniques, and extract image features, e.g., texture and layout. Then a machine-learning-based image detector is used to identify the malware family that is solely based on malware images that exhibit similar behaviors. Fig. 1 depicts a detailed architecture and framework of the image-visualization based malware detection method. Firstly, malware binaries are converted or visualized as gray-scale images. They observe that for many malware families, images belonging to the same family aims to have similar image layout and texture. This is based on the assumption that most of the malware authors use an existing malware and modify the source code to create new malware. It has been observed that the Image visualization-based malware detection methods [19], [20], [18] yield a comparable detection performance in comparison to the regular static and dynamic malware detection technique [32], [33].

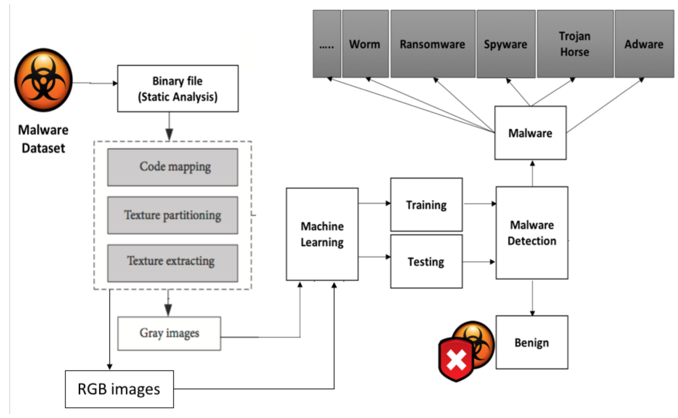


Fig. 1. Malware Detection Framework using Visualization Architecture.

The following covers the types of malware visualization methods: Gray-scale image visualization, RGB color image visualization-based methods that are widely used in the previous research works [19], [20], [18]. Also, using the malware binary file and other code analysis techniques like Opcode visualization and System Flow Graph (SFG) are also utilized [18]. As shown in Fig. 1 and Fig. 2, the system uses gray-scale and RGB color images of malware and computes the feature vector, i.e., fingerprint, to identify a malware binary. This fingerprint captures the structural, visual similarity between malware variants. The Opcode and SFG methods are out of scope for this work.

B. Gray-Scale Image Generation

In Fig. 2, a malware binary executable file is converted into a gray-scale image. A malware binary is converted to a vector of 8-bit unsigned integers and then organized into a two-dimensional (2D) array. This 2D data is visualized as a gray-scale image in the range [0,255] where 0 is black, and 255 denotes white. The image's width is fixed, and the height is allowed to vary according to the malware binary file size. Nataraj et al. recommend image widths for different file sizes based on empirical observations. We use their code base for converting and feature extraction for the Gray-scale image visualization prior to applying our proposed learning classifiers.

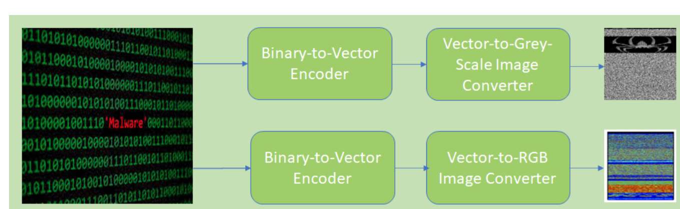


Fig. 2. Malware to Image Converter.

C. RGB-Color Image Generation

In Fig. 2, a malware binary executable file is converted into a RGB, color image. A malware binary is converted into a vector of 24 bits binary data (8-bit represents red color, 8-bit represents the green color, and 8-bit represents blue color) unsigned integers and then organized into a 2D array. This 2D data is visualized as a color image [0,255] where 0 is black, and 255 denotes white. The image's width is fixed, and the height varies according to the malware binary file size.

D. Malware Feature Extractor (MFE)

In this work, we conduct the malware detection approach using various machine learning algorithms on the features of the visualized image of the malware executable. We learn from the extensive research on signal and image processing techniques that implement compact signature extraction methods. The learning algorithms used in this work use image features extracted from these binaries. We consider techniques from the signal and image processing where compact structure extraction methods have been extensively studied. The image processing for content-based image retrieval has been extensively explored by Manjunath et al. [34], and scene classification by Olivia et al. [35] and Torralba et al. [36]. Manjunath et al.[34] propose image processing using texture information for browsing and retrieval of large image data. It uses the Gabor wavelet features for texture analysis and supports a comprehensive experimental evaluation. The Gabor features provide the best pattern retrieval accuracy [34]. We adopted a similar malware feature extraction method formulated using the GIST-based image features. The GIST method uses texture and spatial layout of an image [14], [37], [38]. Refer Fig.3 for the design details of the Malware Feature Extractor. Nataraj et al. founded the following feature extraction technique. Typically, a smaller resized or reshaped version of the image is used to compute the features [36]. Firstly, the binary executable file is converted to a discrete 1-dimensional signal by numerically coding every byte value as an 8-bit number that ranges from $0\breve{\ }255$. Then the signal is "reshaped" to a 2-dimensional gray-scale image with "d" being the width and "h" being the height of the reshaped image. During reshaping, the width "d'' and the height "h" are fixed depending on the number of bytes in the binary. The horizontally adjacent pixels in the image correspond to the adjacent bytes in the binary. The vertically adjacent pixels are associated with the bytes spaced by a multiple width of "d'' in the binary. Then the image is passed through various filters that capture both the short-range and long-range correlations in the image. The localized statistics are obtained by dividing the filtered images into non-overlapping sub-blocks from these filtered images and computing the average value on those blocks. This is called sub-block averaging. A compact signature is formed by concatenating the averages computed from all the filters. Typically, the features are extracted from the image's smaller "$resized''$ version.

The feature computation details are explained below. The image on which the feature needs to be extracted is defined using $(I(x, y))$. The GIST descriptor id computed by filtering the image through a bank of Gabor filters that are band-pass filters whose responses are Gaussian functions modulated with a complex sinusoid. The filter response $t(x, y)$ and the Fourier Transformed version $T(u, v)$ are defined by the following equations: Filter Response (FR) equation 1 and Fourier Transform of FR equation 2.

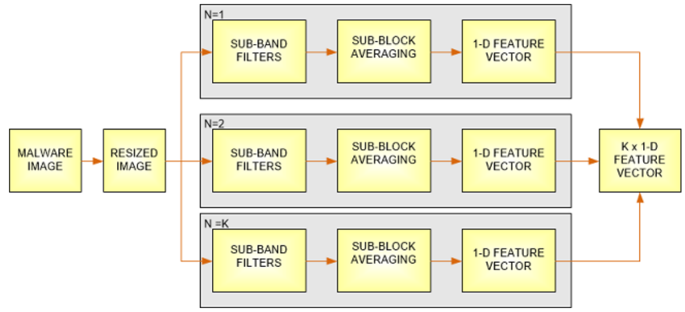


Fig. 3. Malware Feature Extractor.

$$t(x, y) = \frac{1}{(2\pi\sigma_x\sigma_y)} exp[\frac{1}{2}(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2})] \quad (1)$$

$$T(u, v) = exp[-\frac{1}{2}(\frac{(u-w)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2})] \quad (2)$$

A filter bank, i.e., a block of filters, is created by rotating orientation and scaling the basic filter response function $t(x, y)$, resulting in a set of similar filters. Let S be the number of scales and O be the number of orientations per scale in a decomposed multi-resolution image. An image is filtered using "k'' such filters to derive "k'' filtered images. To conduct our experiment, we select the number of filters, $k = 20$ with a number of scales $S = 3$ and in which the first two scales have 8 orientations (O=8), and the last one has four orientations, i.e., $O = 4$. Each filtered image is further divided into $B.B$ sub-blocks, and the average value of a sub-block is computed and stored as a vector of length "L'' where $L = B^2$. Using this above method, "k'' vectors of length "L'' are created per image. These vectors are then concatenated to form a $kL{-}dim$ feature vector called GIST. We choose a B value of 4 to obtain a 320-dimensional feature vector in our work. When computing GIST descriptors, there is a key pre-processing step involved, that is, to reduce the image size to a square of dimensions SxS. In our work we choose a value of $S = 64$. An optimal value of S to be used in the computation because larger the value of S increased the computational complexity. Because of the sub-band averaging, this higher $S - vale$ does not significantly affect and strengthen the signature.

E. Machine Learning Approaches

We have explored a few learning classifiers to evaluate a better performing algorithm for implementing the malware detection function using virtualized images as part of this work. This section describes the theory and implementation of the machine learning algorithms that we used in this study: Naïve Bayes Classifier (NBC), Support Vector Machines (SVMs), Random Forests, k-Nearest Neighbours (KNN), CART, and MLP. As a note, all of the ML classifiers were trained on normalized data.

F. Evaluation Methods

The evaluation of machine learning classifiers is critical when studying the learning models and their performance. To evaluate the performance of the classifier models, we have used similar evaluation measures that are adopted in most of the previous research experiments that involve learning-based malware detection using visualized images of malware binaries. It covers the prediction accuracy and F1-score under varying conditions of input parameters. Most of the time, we use classification accuracy to measure the performance of machine learning models, and we have also used confusion matrices to compare the prediction accuracy and failures.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

$$Precision(P) = \frac{TP}{TP + FP} \tag{4}$$

$$Recall(R) = \frac{TP}{TP + FN} \tag{5}$$

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{6}$$

In the above equations, TP, TN, FN, FN, and FP are true positives, true negatives, false positives, and false negatives. We use F1-score as the primary performance indicator to evaluate all the classifier models used in our experiments. F1 score is a single metric that combines both precision and recall. The precision or True Positive Rate (TPR) is a way to look at the accuracy of positive predictions performed by a classifier and can be defined as follows: precision = $TP/(TP + FP)$ where the True Positives (TP) is the number of true positives, i.e., correct prediction of a positive sample, and the False Positives (FP), i.e., the wrong prediction. But precision is used with another parameter called recall. Recall is defined by $TP/(TP + FN)$.

We have also used a confusion matrix table to study the performance of classifiers. The confusion matrix is a table with rows and columns that report false positives, false negatives, true positives, and true negatives. This allows a more detailed analysis than the mere proportion of correct classifications, i.e., prediction accuracy.

IV. Experiments

Our experiments include the following three step approach. Firstly, we convert the malware binaries into images. Secondly, we perform feature extraction using the GIST-enabled MFE. Finally, we perform the malware classification using the following machine learning methods: Linear Regression, Random Forest, KNN, CART, SVM, and MLP, CNN. We evaluate on the following three malware datasets; SARVAM [14] (25 malware families), Xue et al. [18] (10 malware families), and CIC [39] (six malware categories). These experiments leverage gray-scale image features and RGB Color image features to build models of different classifiers and evaluate their performances.

A. Datasets Summary

Fig. 4 shows malware families and categories distributions across all datasets. SARVAM dataset contains 9,339 instances broken into 25 malware families. The families distribution of malware are unbalanced with majority instances belonging to a family called *Allaple.A* as seen Fig. 4(a). In Xue et al. dataset, there are 10 malware families distributed in 5,314 instances. Fig. 4(b) presents malware family distribution with almost 20% of *Backdoor.Win32.Bifrose* as the top family. CIC is the smallest dataset which contains 439 instances and broken into six categories with a lot of families. This data shows that *Scareware* category includes the most instances among the families and *Botnet* includes the least instances as noticed in Fig. 4(c). Last, the benign dataset includes 1,024 instances collected from different sources and combined as one dataset. Finally, our experiments for the gray-scale images include all datasets while the RGB color images were just included for Xue and CIC datasets. We have not included the RGB color images as part of experiments and used only the gray-scale images due to unavailability of the original executable files of SARVAM dataset.

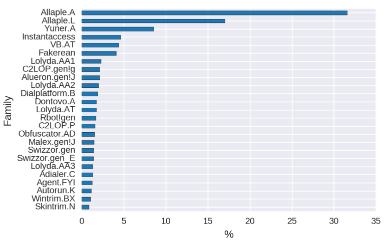
B. Models Setup

Our experiments are conducted in two phases. The first phase focuses on identifying malware from benign as binary classification tasks. The second phase involves a multi-class classification that identifies the individual malware family. We evaluate seven classification algorithms and compare between them based on four performance metrics stated above. Due to the limitation of datasets sizes, we used 10-fold cross-validation mechanism to train and test the classification algorithm. The second phase aims to detect the family or the category of malware as a multi-classification task using the most effective classification algorithm in the previous phase. This task is evaluated by confusion matrix in order to see how the actual unbalanced malware families and categories are truly and falsely predicted. Both phases were consecutively built to compare the results of gray-scale images and RGB color images.

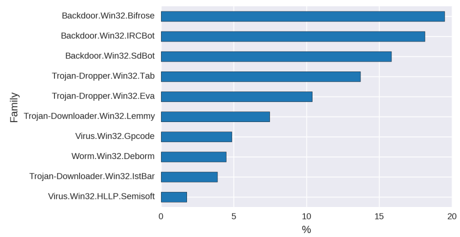
C. Binary Classification Performance

In this phase. we compare the performance results of the seven classifiers as described in Fig. 5 and Table I. This task aims to predict malware instances from benign instances. Each instance includes 320-dimensional feature vector with the class types. The total elements of the feature vector are broken into 960 (320x3) elements. We firstly measure the accuracy of each fold (using 10-fold cross-validation) in each classifier. We depicted the accuracy results for the folds through their quarterlies in the box plots as seen in Fig. 5. We found that K-Nearest Neighbors classifier outperform others in the accuracy results across all datasets, while support vector machines and Naïve Bayes classifiers perform closely low in most cases.

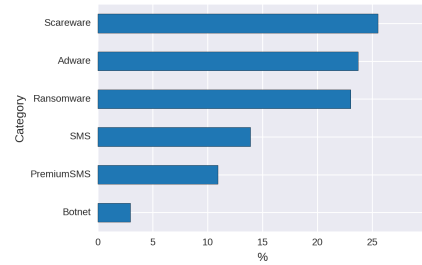
In a more nuanced view, the performance of binary classification for malware and benign can be described through four following metrics: average accuracy, precision, recall and F1-score as seen in Table I. Whilst recall denotes the ability to locate every relevant instance in classification datasets, precision denotes the number of data points a classifier classifies



(a) (a) Sarvam dataset

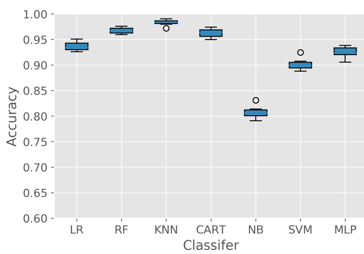


(b) (b) Xue dataset

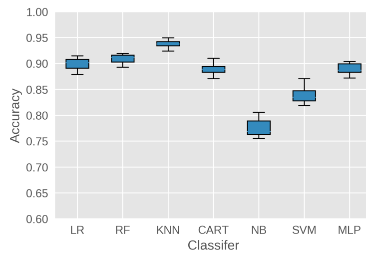


(c) (c) CIC dataset

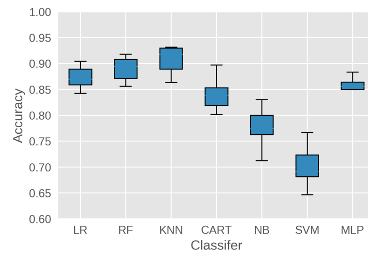
Fig. 4. Malware Families and Categories Distribution over All Datasets.



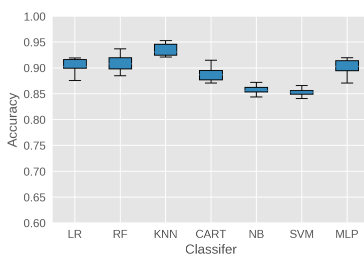
(a) (a) gray-scale: Sarvam



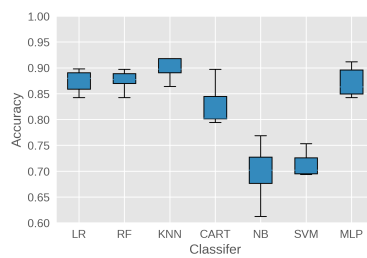
(b) (b) gray-scale: Xue



(c) (c) gray-scale: CIC



(d) (d) RGB Color: Xue



(e) (e) RGB Color: CIC

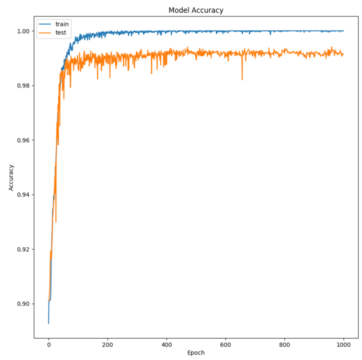
Fig. 5. Comparison of Seven Classifiers 10-fold CV Accuracy Results for Predicting Malware and Benign.

TABLE I. BINARY CLASSIFICATION TO PREDICT MALWARE AND BENIGN USING SEVEN CLASSIFIERS OF THREE DATASETS.

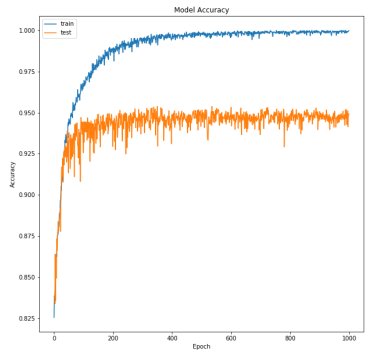
		Sarvam				Xue								CIC							
Classifiers	**Class**	**Avg.Acc.**	**Precision**	**Recall**	**F1-score**	**Avg.Acc.**		**Precision**		**Recall**		**F1-score**		**Avg.Acc.**		**Precision**		**Recall**		**F1-score**	
						Gray-Scale	**RGB Color**	**Gray-Scale**	**RGB Color**	**Gray-Scale**	**RGB Color**	**Gray-Scale**	**RGB Color**	**Gray-Scale**	**RGB Color**	**Gray-Scale**	**RGB Color**	**Gray-Scale**	**RGB Color**	**Gray-Scale**	**RGB Color**
LR	Benign	93.69%	87.30%	42.29%	56.97%	89.92%	90.50%	82.33%	81.98%	48.24%	52.44%	60.84%	63.97%	87.36%	87.83%	91.18%	91.71%	90.82%	90.72%	91.00%	91.21%
	Malware		94.01%	99.33%	96.59%			90.76%	91.43%	98.01%	97.78%	94.25%	94.50%			78.78%	78.89%	79.50%	80.87%	79.14%	79.87%
RF	Benign	96.95%	95.64%	72.75%	82.64%	91.02%	90.68%	85.04%	84.15%	52.73%	53.42%	65.10%	65.35%	89.27%	86.60%	93.66%	90.60%	90.82%	91.31%	92.22%	90.95%
	Malware		97.09%	99.64%	98.35%			91.51%	91.61%	98.21%	98.06%	94.74%	94.73%			80.00%	79.35%	85.65%	77.90%	82.73%	78.62%
KNN	Benign	98.27%	98.84%	83.50%	90.52%	93.83%	93.59%	**88.56%**	86.58%	71.09%	71.19%	78.87%	78.14%	91.05%	90.29%	**96.47%**	95.66%	90.63%	90.33%	93.45%	92.92%
	Malware		98.22%	99.89%	99.05%			94.63%	94.63%	98.23%	97.87%	96.40%	96.23%			80.84%	80.04%	92.26%	90.43%	86.17%	84.92%
CART	Benign	96.40%	84.17%	80.47%	82.28%	88.91%	88.10%	64.23%	63.65%	63.48%	61.91%	63.85%	62.77%	84.89%	83.73%	90.09%	87.81%	88.77%	87.21%	89.42%	87.51%
	Malware		97.87%	98.34%	98.10%			92.98%	92.70%	93.19%	93.19%	93.08%	92.94%			74.67%	70.63%	77.22%	71.75%	75.92%	71.19%
NB	Benign	80.66%	26.00%	51.86%	34.64%	77.20%	86.01%	35.25%	60.03%	48.44%	39.45%	40.81%	47.61%	77.51%	69.79%	96.77%	94.21%	70.12%	60.35%	81.31%	73.57%
	Malware		94.08%	83.82%	88.65%			89.29%	89.06%	82.86%	94.94%	85.95%	91.90%			57.56%	49.69%	**94.53%**	91.34%	71.55%	64.37%
SVM	Benign	90.12%	0.00%	0.00%	0.00%	83.84%	85.33%	0.00%	**94.34%**	0.00%	9.77%	0.00%	17.70%	70.00%	69.99%	69.99%	69.99%	**100.00%**	**100.00%**	82.35%	82.35%
	Malware		90.12%	**100.00%**	94.80%			83.84%	85.17%	**100.00%**	**99.89%**	91.21%	91.95%			0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
MLP	Benign	92.59%	86.43%	28.61%	42.99%	88.85%	89.93%	79.45%	82.46%	39.26%	49.12%	52.55%	61.57%	86.33%	85.58%	88.98%	91.12%	90.72%	90.14%	89.85%	90.62%
	Malware		92.71%	99.51%	95.99%			89.33%	90.90%	98.04%	97.99%	93.49%	94.31%			77.33%	77.56%	73.80%	79.50%	75.52%	78.52%
CNN	**Benign**	**99.50%**	98.20%	**96.70%**	**97.50%**	**94.83%**	**95.46%**	85.28%	83.84%	**82.25%**	**87.03%**	**83.73%**	**85.40%**	**91.17%**	**93.75%**	95.94%	**99.31%**	**90.83%**	91.14%	**93.31%**	**95.05%**
	Malware		**99.60%**	**99.80%**	**99.70%**			**96.60%**	**97.65%**	97.26%	96.98%	**96.93%**	**97.31%**			**82.58%**	**85.26%**	**91.88%**	**98.78%**	**86.98%**	**91.53%**

as relevant where it really is relevant. There is a compromise in these two evaluation metrics to maximize, when increasing the recall, the model decreases the precision. In case we want to find an optimum balance of recall and precision, we use the F1-score to combine measurements for both metrics. We applied these performance metrics per class to evaluate the instance type (belonging to either malware and benign).

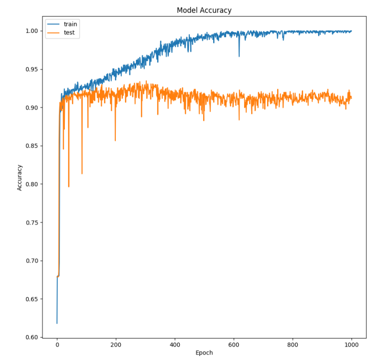
The K-Nearest Neighbor classifier (KNN) achieved in the average accuracy about 98.27%, 93.83%, and 91.05% for the gray-scale images of SARVAM, Xue, CIC datasets respectively, irrespective of the instance types, i.e., malware or bengin. It also achieved similar performance using the RGB color images of the two datasets, i.e., Xue and CIC. While the average accuracy metric cannot evaluate the performance of benign and malware separately, the performance of KNN vary across datasets. For the Xue dataset, F1-score on both gray-scale and RGB color images achieves closely better results in detecting malware (96.48% and 96.23%) than other classifiers except CNN classifier. While KNN F1-score for both gray-



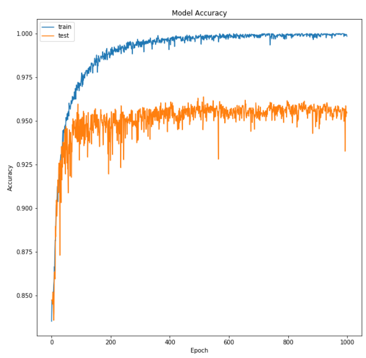
(a) Gray-scale: Sarvam dataset



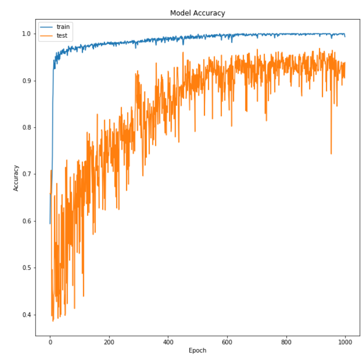
(b) Gray-scale: Xue dataset



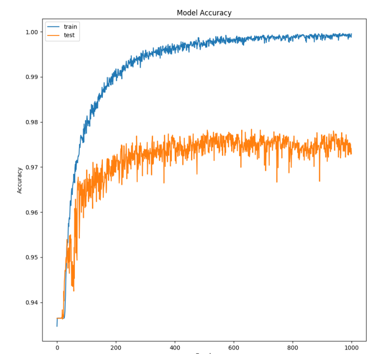
(c) Gray-scale: CIC dataset



(d) RGB Color: Xue dataset



(e) RGB Color: CIC dataset



(f) Gray-scale: All datasets

Fig. 6. Binary-classification CNN Accuracy Models based on Epochs.

scale and RGB color images in detecting benign achieves better results than others (78.87% and 78.14%), it preforms low compared to SARVAM dataset. Conversely, KNN model in CIC dataset can be accurately detecting more benign instances than malware instances for both gray-scale and RGB color images as clearly seen in the F1-score results for both benign (93,45% and 92.92%) and malware (86.17% and 84.92%).

While the tree-based classifiers especially Random Forest (RF) perform well in this classification domain in SARVAM dataset, Decision Tree (CART) results underperform the logistic regression (LR) in the remaining datasets. The performance of these classifiers across datasets follows the same pattern of KNN when comparing F1-score results for detecting malware and benign. In other words, the classifiers' performance for detecting more instances of malware than benign were seen in SARVAM and Xue datasets. In contrast, the results in CIC dataset show that classifiers can detect more benign instances than malware.

The Multilayer Perceptron (MLP) classifier performance results are seen in the median position and they can be improved over all datasets. They are close to the performance results in the logistic regression classifier.

Support Vector Machine classifier (SVM) usually fails to detect the minority class. This means it can learn only one class that reflects the majority. For example, benign instances are considered of the minority class in SARVAM and Xue datasets, and the F1-score results are either zero or below 10, and vice versa. In addition, Naïve Bayes classifier achieves the lowest results in all metrics over all datasets.

The Convolutional Neural Network classifier (CNN) achieved in the average accuracy about 99.50%, 94.83%, and 91.17% for the gray-scale images of SARVAM, Xue, CIC datasets respectively, irrespective of the instance types, i.e., malware or bengin. It also achieves similar performances using the RGB color images of the two datasets, i.e., Xue and CIC. Specifically, CNN F1-score (where there is a balance of precision and recall) in SARVAM dataset achieves higher results for malware instances than benign instances because of their falling off in their recall scores. This means that some instances of benign are predicted to be malware. In contrast, the high precision scores reflect that classifier correctly predicts the majority instances of each type, whether benign or malware. For the Xue dataset, F1-score on both gray-scale and RGB color images achieves closely better results in detecting malware (96.93% and 97.31%) than all other classifiers and F1-score for both gray-scale and RGB color images in detecting benign achieves better results than others (83.73% and 85.40%). In contrast, CNN model of CIC dataset can be accurately detecting more benign instances than malware instances for both gray-scale and RGB color images as clearly seen in the F1-score results for both benign (93.31%

and 95.05%) and malware (86.98% and 91.53%). Finally, the results show that the CNN classifier is the best algorithm to be used in the second phase of classification task. Binary-classification CNN accuracy models based on 1000 Epochs are shown in Fig. 6(a), (b), (c), (d), (e), and (f).

D. Multi-classification Performance

After evaluating binary classification tasks in the first phase, we selected the best classification algorithm that performs well in all metrics among all datasets to be used in this phase. Thus, we aim in this phase to measure the performance of the KNN classification task on identifying the individual malware family or category. Due to the variation in naming malware families on each datasets, we separately evaluated the model performance for each dataset. We chose the confusion matrix as a performance metric in order to see how the actual unbalanced malware families and categories are truly and falsely predicted. It also depicts where the model is confusing classes and mislabeling one as another.

The performance results of the KNN models for malware families are reflected in a confusion matrix as shown in Fig. 7, where the ordinate and abscissa are the number of the malware family or category. The abscissa indicates the actual malware family and the ordinate indicates the predicted malware family. The color gray patches in the figure indicate the similarity between the predicted instances and the actual instances under specific family or category. According to the ribbon on the right, the more dark the color is to the top, the higher the similarity is, and less dark it is to the bottom, the lower the similarity is.

In SARVAM dataset (available only for gray-scale images), the KNN model achieves a large probability of true positives over the majority of malware families depicted in the dark color of the right diagonal in Fig. 7(a). In the other hand, there are few families misclassified. For example, the figure shows that lot of instances of "*Swissor.gen*" malware are falsely predicted as "*Swissor.gen E*" malware (42%) and vice versa (34%). It seems the features vector of both malware families showing same patterns and have identical sequences of codes. Similarly, many "*C2LOP.P*" malware instances are misclassified by the model as "*C2LOP.gen!g*". This indicates that the model is biased toward the malware family that contains more instances and have identical sequences of codes with the other family. In other words, when a false prediction occurs, the predicted family is likely to belong to the same family series.

While the Xue dataset for the gray-scale images and the RGB color images are evaluated separately, the performance results of true positives and false positives are compatibly approximate among malware families as seen in Fig. 7(b) and Fig. 7(d). The KNN model for detecting malware family for this dataset achieves lower performance in the confusion matrix than for the SARVAM dataset. In addition, non malware families exceed 89% of the true positives of the total instances. For the CIC dataset, the KNN model also does not achieve a good results (Fig. 7(c) and Fig. 7(e)) compared to the models built for the other datasets.

We also train and test a model for multi-classification using CNN. For the SARVAM dataset, accuracy on gray-scale images achieves closely better results in detecting malware 99.30% than all other classifiers as shown in Fig. 8(a) and Fig. 9(a). For the Xue dataset, accuracy on both gray-scale and RGB color images achieves better results in detecting malware (87.79% and 88.09%) than all other classifiers as shown in Fig. 8(b) and (d) and Fig. 9(b) and (d) . For the CIC dataset, accuracy on both gray-scale and RGB color images achieves better results in detecting malware (78.51% and 85.93%) than all other classifiers as shown in Fig. 8(c) and (e) and Fig. 9(c) and (e)

TABLE II. MODEL GENERALIZATION: TESTING THE MODEL ON UNSEEN DATA TO PREDICT MALWARE AND BENIGN USING KNN AND CNN

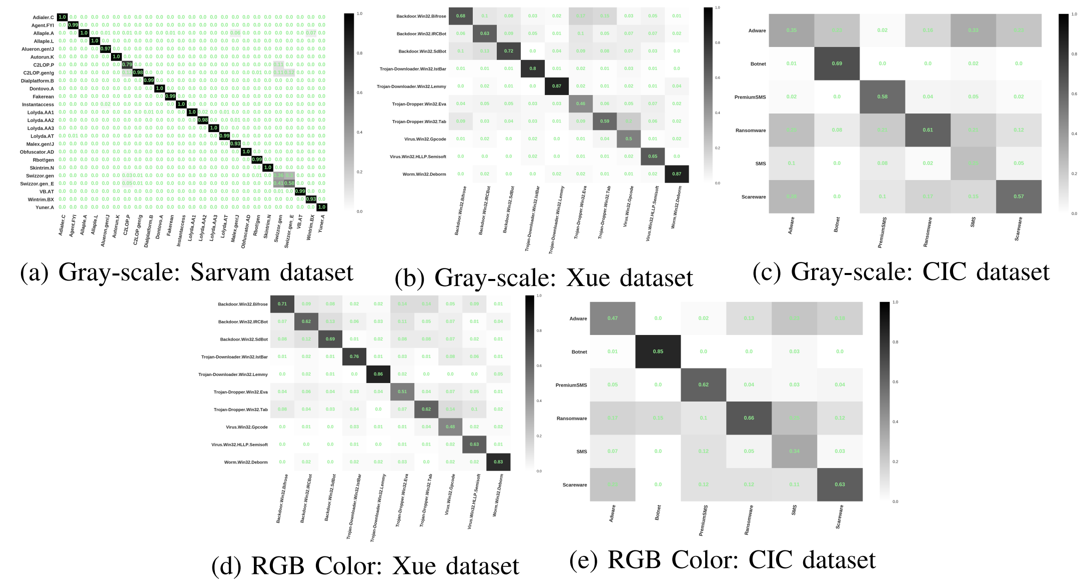
Accuracy	Classes	Precision	Recall	F1-Score
97.10% (KNN)	Benign	89.66%	61.54%	72.98%
	Malware	97.44%	99.52%	98.47%
97.56% (CNN)	Benign	82.4%	77.81%	80.18%
	Malware	98.5%	98.9%	98.7%

E. Deep Learning Experiments using CNN

We also conducted deep learning experiments on visualized malware images using the Convolutional Neural Networks (CNN). After we generate grayscale and RGB images from Binary malware or benign executable files. we construct the CNN which has 24 layers (excluding the input layer), including 8 convolutional layers, 5 pooling layers, 6 dropout layers, 3 full-connection layers, and an output layer. All the convolutional layers use a 3×3 convolution kernel with a step size of 1; the number of convolution kernels in the eight layers are 8, 16, 32, 32, 64, 64, 128, and 256. Because the size of the feature map does not change when the feature map passes through a convolutional layer, a 1-pixel edge fill is performed on each input feature map in the convolution layer. We use all max pooling layers with a 2×2 sliding window and a step size of 2 as shown in Fig. 10. Because the last fully-connected layer of the CNN requires that the input feature maps should be the same size, the general CNN network structure needs to preprocess the image to unify the image size. We use a dropout regularization layer with 0.25 and 0.5 after each pair of convolutional and pooling layers to prevent CNN network overfitting. We also use ReLU and softmax activation functions for multi malware classification and sigmoid for binary malware classification.

F. Multi-class Open-Set Recognition Performance

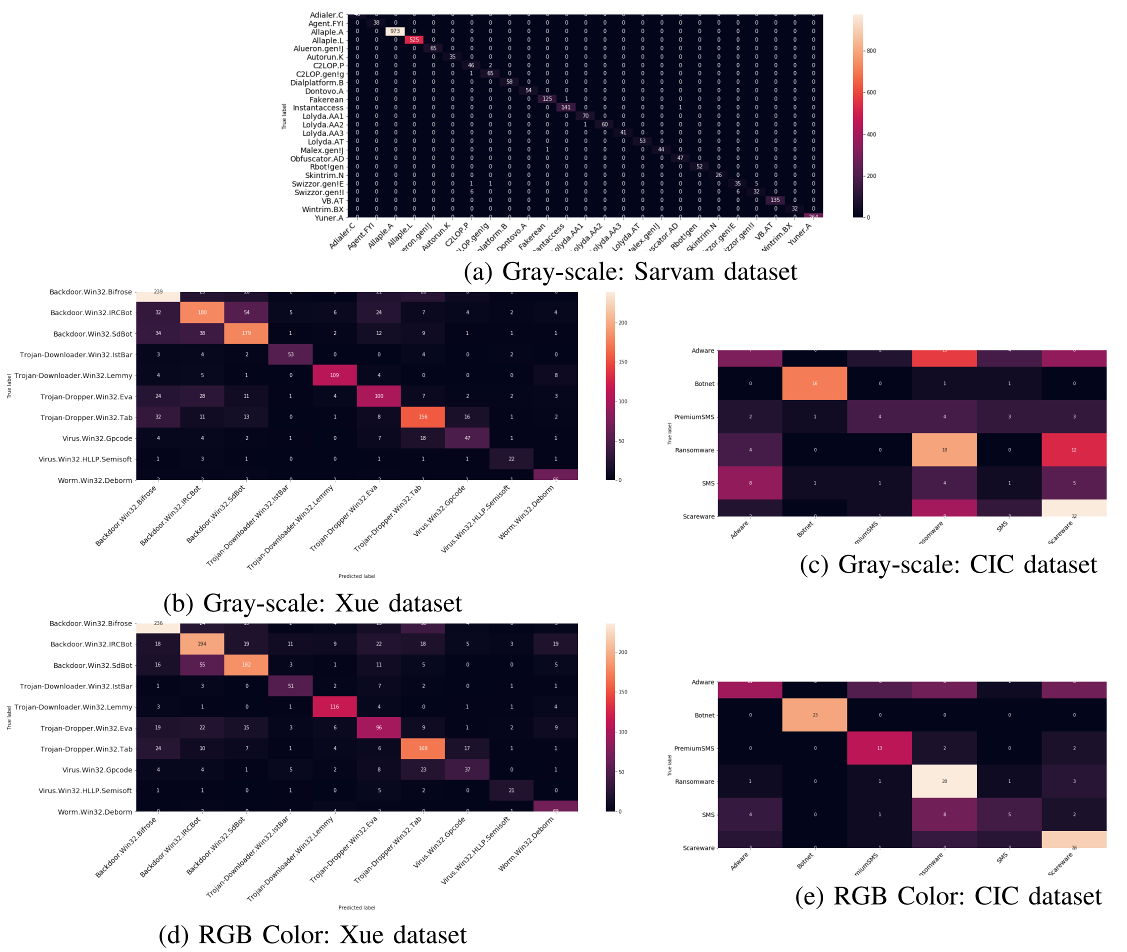
Most of the malware detectors used today fall under the category of closed-set assumption. In a closed-set operation, the data for training and testing are drawn from the same label space and from the same distribution. A large database containing malware signatures are used and signature vectors received are compared against the database leading to a binary classification scheme of malicious or benign files. If there is a new malware, i.e., the training and testing distributions are different, all these methods from closed-set detection will fail. One possible approach that can be extended from our empirical study is to use the Open-set recognition based approach to implement malware detection technique using visualized malware images.



(a) Gray-scale: Sarvam dataset (b) Gray-scale: Xue dataset (c) Gray-scale: CIC dataset

(d) RGB Color: Xue dataset (e) RGB Color: CIC dataset

Fig. 7. Multi-classification Confusion Matrix of the KNN Classifier to Detect Malware Family and Category over Three Datasets.

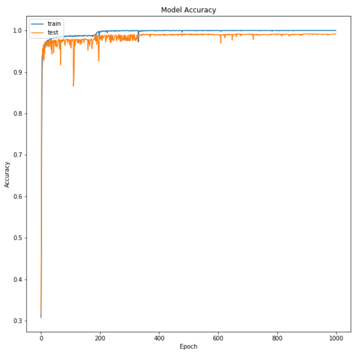


(a) Gray-scale: Sarvam dataset

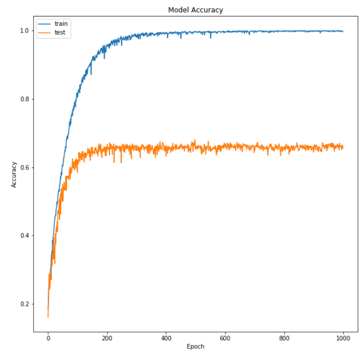
(b) Gray-scale: Xue dataset (c) Gray-scale: CIC dataset

(d) RGB Color: Xue dataset (e) RGB Color: CIC dataset

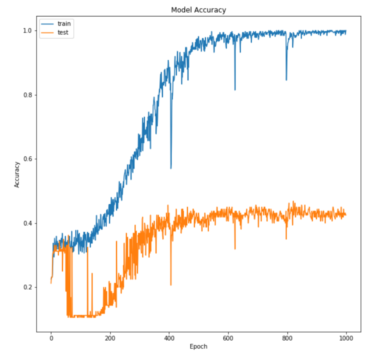
Fig. 8. Multi-classification Confusion Matrix of the CNN Classifier to Detect Malware Family and Category over Three Datasets.



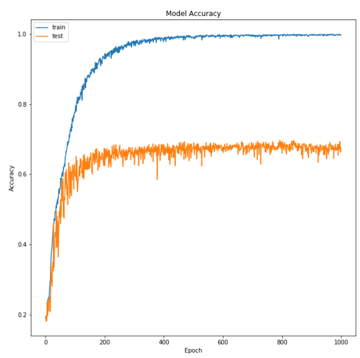
(a) Gray-scale: Sarvam dataset



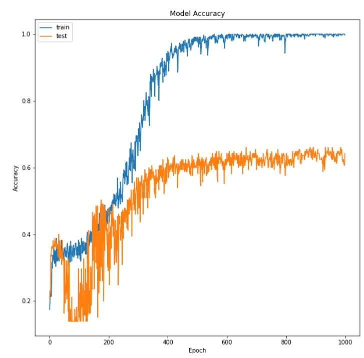
(b) Gray-scale: Xue dataset



(c) Gray-scale: CIC dataset



(d) RGB Color: Xue dataset



(e) RGB Color: CIC dataset

Fig. 9. Multi-classification CNN Accuracy Models based on Epochs.

G. Model Generalization

Unlike other works, we built a generalized KNN and CNN models using 66.67% of the three datasets for the binary classification task. Then, we tested the model on the remaining combined datasets (33.33%) as unseen data. Our splitting mechanism applies the stratified sampling method to avoid the bias toward a majority class. This approach is more solid for evaluating the performance of the model. Moreover, combining datasets from independent sources will examine whether the model can be generalized for any cases. Table II shows the performance of the models tested on unseen instances of benign (338) and malware (4,981). In general, the models achieves above 97% of the accuracy where it is better detecting malware than benign. To investigate the model in depth, let us assume that malware cases are positive and benign cases are negative. The models were tested on 4,981 malware cases and 338 benign cases as unseen cases. The model of KNN predicted correctly more than 4,000 malware instances as true positive cases while there are 24 malware instances predicted as benign. In the other side, the model of KNN predicted truly 208 benign instances while the remaining are predicted as malware. The model of CNN predicted correctly more than 4,000 malware instances as true positive cases while there are 55 malware instances predicted as benign. In the other side, the model of CNN predicted truly 263 benign instances while the remaining was predicted as malware.

V. Discussion

In this section, we discuss the experimental results and their broader implications. Addressing the malware problem is an ongoing research area. We studied the impact of using both the gray-scale and RGB color images in the malware visualization method. We implemented binary and multi-class classification using both gray-scale imaging and RGB-color images. In our study, we find that converting the binary files to either gray-scale images or RGB color images does not impact the performance results. Hence, focusing on a new approach in feature generation and extraction is required. Using new visualization techniques like Speeded Up Robust Features(SURF), Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), and Scale Invariant Feature Transform (SIFT) might better support and enhance the malware detection model [40] but subject to evaluation. In this work, we have included only the GIST-based Image visualization technique for malware detection. Any other visualization and non-visualization-based discussion are out of our work's scope. Nataraj et al. [14], [37], [15]'s original setup included 9,458 samples from 25 malware families with gray-scale image conversion and GIST-based feature extraction. Their approach achieved a classification accuracy score of 98 %. While our results align with the performance results, we extended the framework to include data not only from SARVAM [14] (25 malware families), but also from Xue et al. [18] (10 malware families), and CIC [39]

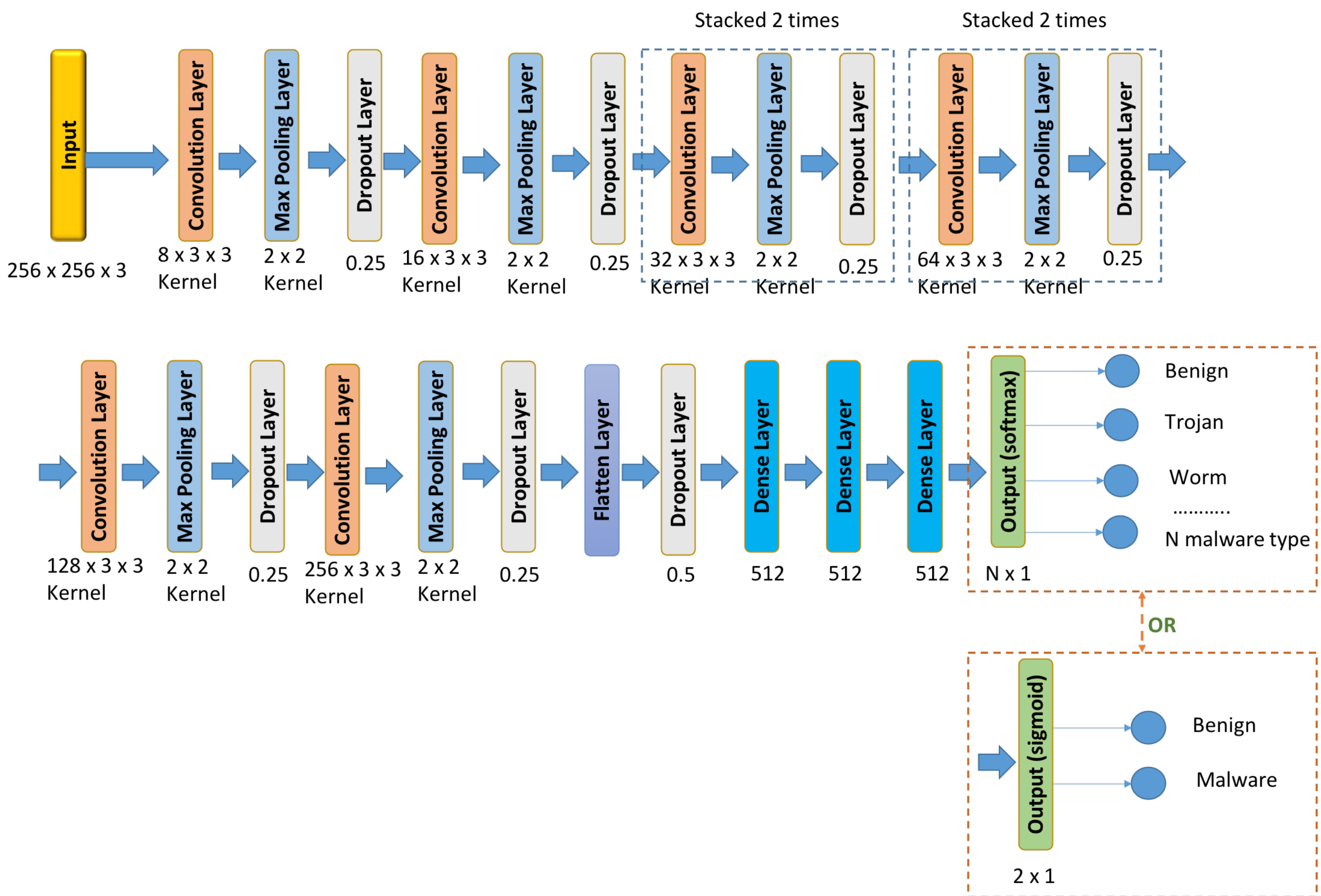


Fig. 10. Convolutional Neural Networks Architecture for Malware Classification.

(6 malware families). In addition, we evaluated the proposed design using various machine learning algorithms i.e. Linear Regression, Random Forest, KNN, CART, SVM. We also have evaluated using CNN-based deep learning technique and MLP. In our setup, we consider the idea of using both the binary and multi-class classifiers that provide handlers to extend the design to an open-set and ensemble-based learners.

Our approach is producing a comparable performance to the above-mentioned classifiers. Even though the datasets used in these classifiers are SARVAM and other datasets, our statistical T-TEST on F1-scores are not statistically significantly different.

VI. Conclusion

In this work, we implement the computer visualization-based technique to build ML-based malware detection. We employ a GIST-based approach to extract malware image features from both the gray-scale and color images of malware binary samples. To study the prediction performance, we empirically analyze various machine learning algorithms. Our experimental study includes the following learning algorithms: linear regression, random forest, k-NN, CART, SVM, and MLP, and a CNN-based deep learning model and selects a candidate learning classifier that can yield better prediction performances. We evaluate our approach using the following malware datasets SARVAM, PhD-thesis, and CIC. In comparison to traditional malware detectors, the visualization-based approach provides a significant performance enhancement. Our study observes that the CNN-based deep-learning model yields significantly better performance when tested against the various malware datasets listed above. Malware authors constantly innovate new methods for implementing sophisticated attacks that make the malware detection as an active research area. Our approach provides a path forward to implement an innovative malware detector. Nevertheless, this research work needs to be verified against various new malware datasets to be more effective. In the future, we would like to extend our research to benchmark datasets and as well conduct a large-scale evaluation.

References

[1] A. Souri and R. Hosseini, "A state-of-the-art survey of malware detection approaches using data mining techniques," Human-centric Computing and Information Sciences, vol. 8, no. 1, p. 3, 2018.

[2] M. Fredrikson, S. Jha, M. Christodorescu, R. Sailer, and X. Yan, "Synthesizing near-optimal malware specifications from suspicious behaviors," in 2010 IEEE Symposium on Security and Privacy. IEEE, 2010, pp. 45–60.

[3] X. Hu, T. Chiueh, and K. G. Shin, "Large-scale malware indexing using function-call graphs," in Proceedings of the 16th ACM conference on Computer and communications security, 2009, pp. 611–620.

[4] I. Santos, J. Devesa, F. Brezo, J. Nieves, and P. G. Bringas, "Opem: A static-dynamic approach for machine-learning-based malware detection," in International Joint Conference CISIS'12-ICEUTE 12-SOCO 12 Special Sessions. Springer, 2013, pp. 271–280.

[5] H. Sayadi, N. Patel, S. M. PD, A. Sasan, S. Rafatirad, and H. Homayoun, "Ensemble learning for effective run-time hardware-based malware detection: A comprehensive analysis and classification," in 2018 55th ACM/ESDA/IEEE Design Automation Conference (DAC). IEEE, 2018, pp. 1–6.

[6] H. Sayadi, H. M. Makrani, O. Randive, S. M. PD, S. Rafatirad, and H. Homayoun, "Customized machine learning-based hardware-assisted malware detection in embedded devices," in 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE). IEEE, 2018, pp. 1685–1688.

[7] Y. Feng, S. Anand, I. Dillig, and A. Aiken, "Apposcopy: Semantics-based detection of android malware through static analysis," in Proceedings of the 22nd ACM SIGSOFT International Symposium on Foundations of Software Engineering, 2014, pp. 576–587.

[8] K. Griffin, S. Schneider, X. Hu, and T. Chiueh, "Automatic generation of string signatures for malware detection," in International workshop on recent advances in intrusion detection. Springer, 2009, pp. 101–120.

[9] D. Venugopal and G. Hu, "Efficient signature based malware detection on mobile devices," Mobile Information Systems, vol. 4, no. 1, pp. 33–49, 2008.

[10] I. Santos, F. Brezo, J. Nieves, Y. K. Penya, B. Sanz, C. Laorden, and P. G. Bringas, "Idea: Opcode-sequence-based malware detection," in International Symposium on Engineering Secure Software and Systems. Springer, 2010, pp. 35–43.

[11] J. Scott, "Signature based malware detection is dead," Institute for Critical Infrastructure Technology, 2017.

[12] A. Pinhero, M. Anupama, P. Vinod, C. Visaggio, N. Aneesh, S. Abhijith, and S. AnanthaKrishnan, "Malware detection employed by visualization and deep neural network," Computers & Security, p. 102247, 2021.

[13] E. M. Rudd, A. Rozsa, M. Günther, and T. E. Boult, "A survey of stealth malware attacks, mitigation measures, and steps toward autonomous open world solutions," IEEE Communications Surveys & Tutorials, vol. 19, no. 2, pp. 1145–1172, 2016.

[14] L. Nataraj, S. Karthikeyan, G. Jacob, and B. S. Manjunath, "Malware images: visualization and automatic classification," in Proceedings of the 8th international symposium on visualization for cyber security, 2011, pp. 1–7.

[15] L. Nataraj and B. Manjunath, "Spam: Signal processing to analyze malware," IEEE Signal Processing Magazine, vol. 33, no. 2, pp. 105–117, 2016.

[16] K. Han, B. Kang, and E. G. Im, "Malware analysis using visualized image matrices," The Scientific World Journal, vol. 2014, 2014.

[17] K. S. Han, J. H. Lim, B. Kang, and E. G. Im, "Malware analysis using visualized images and entropy graphs," International Journal of Information Security, vol. 14, no. 1, pp. 1–14, 2015.

[18] D. Xue, J. Li, W. Wu, Q. Tian, and J. Wang, "Homology analysis of malware based on ensemble learning and multifeatures," PloS one, vol. 14, no. 8, p. e0211373, 2019.

[19] L. Liu, B. Wang, B. Yu, and Q. Zhong, "Automatic malware classification and new malware detection using machine learning," Frontiers of Information Technology & Electronic Engineering, vol. 18, no. 9, pp. 1336–1347, 2017.

[20] J. Fu, J. Xue, Y. Wang, Z. Liu, and C. Shan, "Malware visualization for fine-grained classification," IEEE Access, vol. 6, pp. 14 510–14 523, 2018.

[21] A. Makandar and A. Patrot, "Malware class recognition using image processing techniques," in 2017 International Conference on Data Management, Analytics and Innovation (ICDMAI). IEEE, 2017, pp. 76–80.

[22] A. Singh, A. Handa, N. Kumar, and S. K. Shukla, "Malware classification using image representation," in International Symposium on Cyber Security Cryptography and Machine Learning. Springer, 2019, pp. 75–92.

[23] B. Zhang, J. Yin, J. Hao, D. Zhang, and S. Wang, "Malicious codes detection based on ensemble learning," in International conference on autonomic and trusted computing. Springer, 2007, pp. 468–477.

[24] E. Menahem, A. Shabtai, L. Rokach, and Y. Elovici, "Improving malware detection by applying multi-inducer ensemble," Computational Statistics & Data Analysis, vol. 53, no. 4, pp. 1483–1494, 2009.

[25] C. Iwendi, S. Khan, J. H. Anajemba, M. Mittal, M. Alenezi, and M. Alazab, "The use of ensemble models for multiple class and binary class classification for improving intrusion detection systems," Sensors, vol. 20, no. 9, p. 2559, 2020.

[26] W. J. Scheirer, A. de Rezende Rocha, A. Sapkota, and T. E. Boult, "Toward open set recognition," IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 7, pp. 1757–1772, 2012.

[27] J. Henrydoss, S. Cruz, E. M. Rudd, M. Gunther, and T. E. Boult, "Incremental open set intrusion recognition using extreme value machine," in 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, 2017, pp. 1089–1093.

[28] T. Wang and N. Xu, "Malware variants detection based on opcode image recognition in small training set," in 2017 IEEE 2nd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA). IEEE, 2017, pp. 328–332.

[29] S. Tobiyama, Y. Yamaguchi, H. Shimada, T. Ikuse, and T. Yagi, "Malware detection with deep neural network using process behavior," in 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC), vol. 2. IEEE, 2016, pp. 577–582.

[30] B. Kolosnjaji, A. Zarras, G. Webster, and C. Eckert, "Deep learning for classification of malware system call sequences," in Australasian Joint Conference on Artificial Intelligence. Springer, 2016, pp. 137–149.

[31] W. Huang and J. W. Stokes, "Mtnet: a multi-task neural network for dynamic malware classification," in International conference on detection of intrusions and malware, and vulnerability assessment. Springer, 2016, pp. 399–418.

[32] M. Amin, T. A. Tanveer, M. Tehseen, M. Khan, F. A. Khan, and S. Anwar, "Static malware detection and attribution in android byte-code through an end-to-end deep system," Future Generation Computer Systems, vol. 102, pp. 112–126, 2020.

[33] H. Cai, "Assessing and improving malware detection sustainability through app evolution studies," ACM Transactions on Software Engineering and Methodology (TOSEM), vol. 29, no. 2, pp. 1–28, 2020.

[34] B. S. Manjunath and W. Ma, "Texture features for browsing and retrieval of image data," IEEE Transactions on pattern analysis and machine intelligence, vol. 18, no. 8, pp. 837–842, 1996.

[35] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," International journal of computer vision, vol. 42, no. 3, pp. 145–175, 2001.

[36] A. Torralba, K. P. Murphy, W. T. Freeman, M. A. Rubin et al., "Context-based vision system for place and object recognition." in ICCV, vol. 3, 2003, pp. 273–280.

[37] L. Nataraj, V. Yegneswaran, P. Porras, and J. Zhang, "A comparative assessment of malware classification using binary texture analysis and dynamic analysis," in Proceedings of the 4th ACM Workshop on Security and Artificial Intelligence, 2011, pp. 21–30.

[38] M. Douze, H. Jégou, H. Sandhawalia, L. Amsaleg, and C. Schmid, "Evaluation of gist descriptors for web-scale image search," in Proceedings of the ACM International Conference on Image and Video Retrieval, 2009, pp. 1–8.

[39] A. H. Lashkari, A. F. A. Kadir, L. Taheri, and A. A. Ghorbani, "Toward developing a systematic approach to generate benchmark android malware datasets and classification," in 2018 International Carnahan Conference on Security Technology (ICCST). IEEE, 2018, pp. 1–7.

[40] N. Raj and V. Niar, "Comparison study of algorithms used for feature extraction in facial recognition," Int. J. Comput. Sci. Inf. Technol, vol. 8, no. 2, pp. 163–166, 2017.

Eye-movement Analysis and Prediction using Deep Learning Techniques and Kalman Filter

Sameer Rafee[1], Xu Yun[2], Zhang Jian Xin[3],
School of Mechanical Engineering & Automation
Zhejiang Sci-Tech University
Hangzhou, China

Zaid Yemeni[4]
Collage of Internet of Things (IoT) Engineering
Hohai University
Changzhou, China

***Abstract*—Eye movement analysis has gained significant attention from the eye-tracking research community, particularly for real-time applications. Eye movement prediction is predominantly required for the improvement of sensor lag. The previously introduced eye-movement approaches focused on classifying eye movements into two categories: saccades and non-saccades. Although these approaches are practical and relatively simple, they confuse fixations and smooth pursuit by putting them up within the non-saccadic category. Moreover, Eye movement analysis has been integrated into different applications, including psychology, neuroscience, human attention analysis, industrial engineering, marketing, advertising, etc. This paper introduces a low-cost eye-movement analysis system using Convolutional Neural Network (CCN) techniques and the Kalman filter to estimate and analyze eye position. The experiment results reveal that the proposed system can accurately classify and predict eye movements and detect pupil position in frames, notwithstanding the face tracking and detection. Additionally, the obtained results revealed that the overall performance of the proposed system is more efficient and effective comparing to Recurrent Neural Network (RNN).**

Keywords*—Eye Movement Classification; Eye Movement Prediction; Convolutional Neural Network (CNN); Recurrent Neural Network (RNN)

I. Introduction

For a few decades, eye movement analysis has attained significant attention in the Human-Computer Interaction community. Various researchers have proposed using different technologies and algorithms to perform an automatic analysis for the position and direction of human eyes, helping different applications perform some analytics. The research community of eye movement analysis is growing owing to its capabilities of facilitating various tasks. One of these most useful is to detect someone's interest by determining the eye movement position and pupil size at a certain time. The analysis of eye movement has been commonly used in a variety of applications and different research areas, including visual systems [1], [2], [3], neuroscience and psychology [4], healthcare and psycholinguistics [5], user experience and interaction [6], consumer research, professional performance, and marketing [7], clinical research, economy, and education [8], software engineering, and product design [9], virtual reality [10], and transportation [11].

One of the main significant features of this research is to provide a method that can be exploited to accurately and objectively record and analyze human visual behavior. It would be incredible to request people glance over the supermarket aisles to remember how many minutes they looked at every item or what advertisements they noticed most [12], [13]. Analyzing human eye movements enables researchers to study the movements of a participant's eyes during various activities, which provides insightfulness into the cognitive processes and understanding the common behavior of human behavior. Besides, it can show different things, including the methods of social interaction and learning patterns. It also helps researchers to screen typical nervous growth and perceptual disabilities. Furthermore, analyzing eye movements helps people clarify their thoughts [14], [15].

Eye movement analysis is referred to as the process of determining either the gaze point or the eye movement relative to the head position. In other words, Eye movement analysis classifies eye positions and eye movements by providing informative details about such movements. Eye movement analysis systems are widely utilized in various research fields, such as psychology, visual system, marketing, psycholinguistics, product design, and human-computer interaction. They are also ever-increasingly exploited for assistive and rehabilitative applications. Besides, along with eye movement analysis, it is required to estimate a noise-reduced signal's open eye movement position. Therefore, signal smoothing is obligatory to minimize the eye tracker's limited accuracy, and prediction is desirable to reduce the sampling lag of tracking systems. The eye movement prediction is meaningfully connected to distributed systems where instantaneous data of eye movements are required to be propagated over the network—the best example of this is Collaborative Virtual Environment (CVE) [10], [16].

There are numerous fields in which eye movement analysis could be applied extensively. Here to mention a few:

- Scientific Research: Studying visual behavior helps research obtain significant insightfulness into nervous growth, learning patterns, and cognitive impairments or diseases, such as people with Parkinson, Alzheimer, schizophrenia, brain injury, depression, autism, etc. In the same way, eye movement analysis can identify dyslexia and other learning or reading difficulties [17][18].
- Market Research: Eye movement analysis provides detailed and impartial information about the behavior of consumers and helps carry out decision-making processes. It enables brand owners and market researchers to detect consumers' behavior when observing and

choosing a product. It enables them to figure out what the most products naturally attracted the customers and the most product they ignored. Unlike questionnaires or surveys, eye movement analysis provides detailed, trustworthy behavior critical for branding, advertising, and packaging [16].

- Human Performance: Eye movement analysis and prediction can provide good insight into the ways of performing tasks and implementing processes implemented. It is used to identify risks, streamline training, operational inefficiencies, time efficiencies, and productivity improvement. Likewise, in sports and coaching, eye movement analysis can be used to improve performance by identifying skills and setting up strategies. It gives a unique insight into actions and tasks carried out hurriedly and often unconscious [19].

This paper introduces an eye-movement analysis and prediction system using a combination of the Kalman Filter and a convolutional neural network Network (CCN) to analyze the eye movement signal's instantaneous classification and predict forthcoming eye movement positions. Furthermore, using a numerical interpolation of the eye movement positions, the Kalman filter is utilized to provide a more useful interaction between the proposed system and the end-user. The contributions of this paper are as follows:

- Developing a real-time eye-movement analysis system using deep learning techniques, such as Convolutional Neural Network (CNN) and Kalman filter, which provides an accurate estimation for eye movements.
- Developing a cost-efficient eye-movement analysis and prediction system, which uses easy powerful deep learning-based calibration technique to predict eye movement and predict the eye movements based on the output of the deep learning technique.
- Designing hardware modules which are inexpensive based on a web camera and Arduino board.
- Conducting a quantitative evaluation of the performance for the implemented eye-movement analysis and prediction system.

The remainder of this paper is organized as follows: Section II presents state-of-the-art approaches related to the proposed system. Section III introduces the architecture of the proposed system. Section IV elaborates the evaluation of the proposed system and discusses the obtained results. Finally, Section VI concludes this work.

II. Related Work

Video/image-based eye movement analysis is classified into eye detection, eye position interpretation, and eye-tracking. It is critical to determine the eye model parameters from the image data in eye detection and tracking. The identification of the eye model can be made based on the eye region intensity distribution, the shape of the iris and eye, the pose of the head, etc. The authors in [20] and [21] present an excellent review of state-of-the-art image/video-based eye movement detection and tracking methods. As the authors stated, the eye detection models are categorized into the following classes:

- Eye-feature-based model;
- Eye-shape-based model;
- Eye-appearance-based model;
- Hybrid model.

A. *Eye-shape-based Models*

The approaches based on this model detect the eye by contouring the pupil and iris and the eyelids' shape. Various eye-movement tracking applications use an elliptical model to prototype the eye, iris, and pupil's external shape. In these models, the model-fitting and voting-based methods are employed to calculate the elliptical model parameters. The authors in [22] exploited a voting-based method to contour the iris as an ellipse. They then applied horizontal scanning for the eye region to detect the iris center. At the horizontal line center, the pupil is located. The authors in [23] used ellipse to model the shape of the iris. Then, eye detection is carried out using spatio-temporal information. Because of the area darkness of the iris, the direction of the vertical and horizontal gradients will be outward from the iris center. Accordingly, a proposed voting-based method is used to find the iris center. At each edge point, an extrapolated line is delineated in the opposite direction of the gradient lines. The iris center is determined by the area where the largest number of lines are passing. In [24], the authors proposed a thresholding and contouring method to measure the radius of the iris. Other authors in [25] also used an ellipse to model the iris. Their model is built using RANSAC optimization and Expectation Maximization. In [26], the authors proposed an alternative ellipse fitting algorithm. They suppose that the contour points of the iris lay on an ellipse, and the contour points are applied to determine the ellipse parameters by a least-square fitting method. To obtain the sub-pixel level accuracy, the authors also introduced a sub-pixel edge detection approach. In [27] eye shape-based algorithm is proposed. The proposed algorithm utilizes a weakly supervised eye landmarks detection algorithm along with object detection and recurrent learning modules. The proposed algorithm can augment training data effectively and our specific format data consist of supervised and weakly supervised samples.

B. *Feature-based Models*

The main goal of feature-based models is to detect local features, such as the pupil, limbus, etc., for localizing and modeling the eye parameters [20]. The authors in [28] tried to determine the region between the eyes rather than locating the local features. Their method is known as the "between-the-eyes" region. The idea behind this method is that the region between the eyes is brighter than the two sides. As both eyes are located on this region's sides, it contains the forehead and nose. Therefore, it is easier to be tracked, and both eyes can be detected by localizing both sides. In [29], the authors used the color of the skin to determine the face region. After that, in the detected skin area, both eyes are localized using four Gabor wavelets (linear filtering) to localize the eyes and non-linear filtering to detect the corners of both eyes. A feature-based gaze patterns recognition method is proposed in [30]. The proposed method used an eye tracker to collect data. Furthermore, Long Short-Term Memory (LSTM) technique is employed in this work for eye movements recognition.

C. Appearance-based Models

These models are image template-based models. However, these models are single image-based templates and limited by the head pose changes and eye movements. If the head pose changes, these models fail. Also, these models are limited if there is a change in rotations and scales. Therefore, they need many different training datasets of eye areas to collect various eye orientations and states, and head poses in various illumination environments. A classifier is then modeled based on the pixel information of the given training datasets. Several authors introduced modified versions of featured-based models presented in [29] and [31] for eye detection and feature extractions. In [32] a human-computer interaction model is proposed based on medical stuff eye movement. The Appearance-Based model is used such that the medical staff used their eye movements to control the robot during the operations time.

D. Hybrid Models

Hybrid models are based on the three models introduced above. The authors in [33] and [34] use the appearance-based model as the combination of appearance and shape-based models by [20]. They also combined expectation-maximization for accurate estimation of the head poes and particle filtering for tracking the iris movements [21]. They used the distribution of face color and eye region for eye detection and tracking. In [35], the authors used image color information for face region detection. A hybrid eye movement recognition method is proposed in [36]. The proposed method aims to solve the inaccurate positioning problem of the initial position of the shape model in the process of eyelid matching by using machine learning. In addition, this method developed the algorithm by combining the AK-EYE model based on the combination of the ASM algorithm and Kalman filtering to create a local feature model for each feature point.

Similarly, the authors in [28] and [37] employed the same approach. After the face region detection, the detected face region is partitioned into smaller regions to find the eyes' region using thresholding and edge information. The template matching is then used for eye tracking. After the skin region detection, the detection of the face region is modeled as a face mirror-symmetry and ellipse. Besides, infrared illumination is also used for eye tracking, using more than one camera to capture IR images. If the light source is near the camera's optical axis, the light reflection from the pupil is used to detect the eye region. Accordingly, the pupil will be a brighter area in the image. Conversely, if the light source is far from the camera's optical axis, the pupil will appear as a darker area.

Different techniques for eye movement locating and measuring were proposed. Eye movement analysis is the process of locating the eye and measuring the eye movements. There are various methods and techniques used for eye-movement analysis. Thees techniques are divided into four classes:

- Scleral Search Coil (SSC)
- Infrared Oculography (IOG)
- Electrooculography (EOG)
- Video Oculography (VOG)

E. Scleral Search Coil Technique (SSC)

As stated in [38], the SSC and Videooculography (VOG) techniques are primarily used for 3D eye-movement measurement. The authors proposed a method aiming to compare the accuracy of SSCS and VOG. Accordingly, the obtained results showed that VOG showed better accuracy than SSCS. Another Scleral Search Coil System was presented in [39]. The main goal is to investigate the effects of the Scleral Search Coils placement on the eyes saccades' kinematics. To that end, saccades of the human eye were recorded with an infrared video system while wearing coils. The results were compared with results while no coils were worn. According to [40], the SSC technique is suitable for animal studies, where coil lenses can be surgically implanted, leaving no peripheral to interfere with or damage. However, the SSC technique has several limitations. It can only be worn for a short period (approximately 30 minutes), and the eye cornea is required to be anesthetized. Even though it usually causes mild discomfort throughout and after the experimental sessions. The mounted annulus results in corneal dryness and vision blurring. The accuracy can be degraded by the annulus slippage on the eyes.

F. Infrared Occulography (IOG)

This technique relies on the strength of infrared light reflected sclera to detect and localize the eye position using light beams generated by a pair of glasses. In this technique, a reference point, known as a glint or corneal reflection, is involved using an infrared light source when solving the sensitivity issue of head movements. It has been commonly used by various techniques, such as low pass filtering [41], K-means [42], Euclidean distance [43], random sample consensus [44], etc. The authors in [45] proposed a method based on a combination of IOG and Electromyography (EMG) techniques. The obtained result demonstrated that there is a robust linear correlation between time courses and amplitudes of EMG-and-IOG-recorded startle.

G. Electrooculography (EOG)

As stated in [46], electrooculography (EOG) is an electrophysiologic test measuring the existing resting electrical potential between the Bruch's membrane and cornea. The mean transepithelial voltage of bovine Retinal pigment epithelium is 6 millivolts (mV). According to [47], the detection of saccades within raw mobile electrooculography (EOG) data includes complex algorithms processing data collected when seated static tasks. For dynamic tasks, the data processing is relatively infrequent and sophisticated, especially in the elders or people who have Parkinson's disease. EOG is an applicable and low-priced technique used in the field of human-computer interaction. This technique uses sensors attached to the eyes' surrounding area to obtain an electric field during eye rotation by measuring skin fluctuations. The eye movements are recorded unconnectedly employing electrodes.

Nevertheless, the recorded signal can be subjected to alteration without eye movements. Even if EOG is practicable, it is not widely used. To put it another way, the EOG is limited to medical applications and laboratories. Various approaches have employed this technique to perform different tasks [46][47]. The authors in [47] introduced a new method by employing a differential electrooculography (EOG) signal using a fixation curve (DOSbFC) to eliminate baseline noise and drift based on a new electrode positioning scheme. The authors experimented with EOG eyeglasses and a new detection protocol for long-term step-by-step eye movement detection. The proposed DOSbFC computes the difference values of accumulated EOG signals between the initial eye movement and fixation time. It allows long-term eye movement detection with high accuracy and only needs a single calibration.

H. Video Oculography (VOG)

The primary methods for three-dimensional eye movement measurement are the scleral search coil system (SSCS) and Video Oculography (VOG) [48]. The authors proposed a method to evaluate the accuracy of SSCS and VOG. The obtained results revealed that VOG recorded better accuracy than SSCS. Another scleral search coil system was presented in [49]. The proposed method investigates the extent to which scleral search coils' placement onto the eyes influences saccades' kinematics. The saccades were recorded using an infrared video system. Simultaneously, The coils were worn, and the main-sequence properties were compared with recordings without coils mounted on the eyes. According to [50], the VOG technique is the most widely used technique for currently eye-movement tracking systems.

III. Proposed System Architecture

This section presents the general architecture and detailed components of the proposed system, which focuses on the analysis, classification, and prediction of eye movements. For the classification and prediction of eye movements, eye detection is the first step to be performed. The eye-detection step requires the head pose estimation for proper eye detection. Then, it is followed by finding the pupil to determine the iris center. The eye movement classification uses the positional eye data for the separation of eye movements: saccades, fixations, and smooth pursuit, which is a highly challenging task. Finally, the future eye movement prediction is carried out based on the classification of eye movements and head pose estimation.

A. System Components

The main task of this paper is to construct a system for eye movement analysis and prediction. The system's practical aspect includes incorporating all the components given in detail in the following sections. Thus, the system comprises the following components, as shown in Fig. 1. As illustrated by Fig. 1, the proposed system comprises the following components:

- Hardware components
- Software components.

B. Hardware Components

The hardware component is one of the essential parts of the proposed system. It is considered as the data frame source. The hardware component consists of sub-components: webcam and Arduino Uno board. In this section, the details of each sub-component are articulated. Fig. 2. shows a general connection view of the hardware components and the system. While Fig. 3 represents the real-world implementation of the system described in Fig. 2.

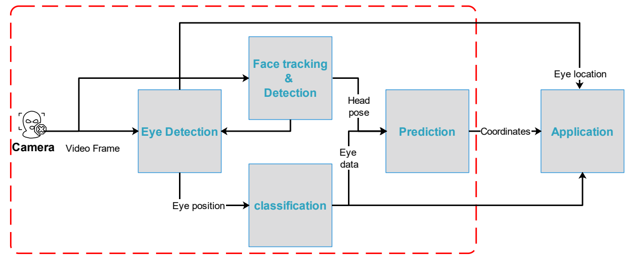


Fig. 1. The Proposed System's Architecture.

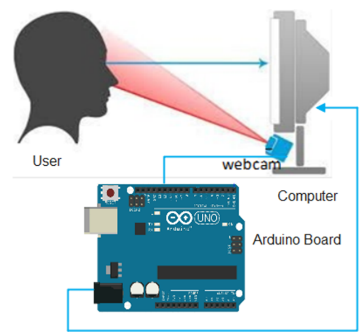


Fig. 2. Hardware Connection.

C. Software Components

The software module is responsible for performing different tasks and consists of the following sub-modules:

- Eye Detection,
- Face Tracking & detection,
- Eye Movement classification, and
- Prediction.

The detailed procedure of each module is given below.

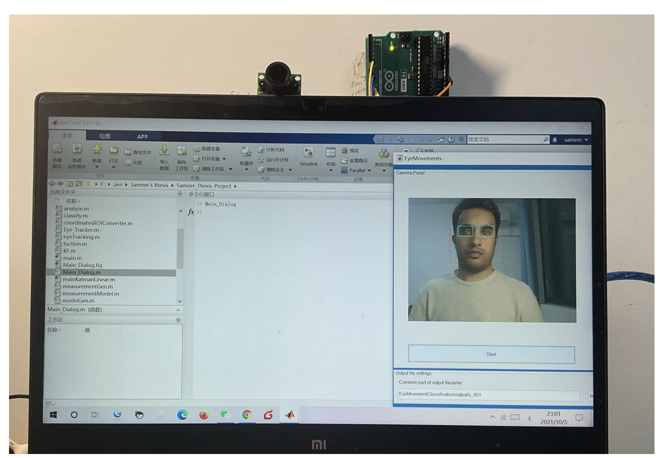


Fig. 3. Software Components.

1) Eye Detection: This component's main task is to detect the eyes on the face and determine the pupils' position. For eye detection, face region detection is the first task to be performed based on the head pose provided by the Head Pose Estimation module. The next step is to detect the eyes region and then detect the iris region.

a) Eye Region Detection: After face region detection, a feature-based method is applied sequentially to determine the regions of the left eye and the right eye. The gradient is calculated for each video frame based on the detected face box to accomplish this objective. Then, a horizontal projection is applied to calculate video frame gradient. In this application, the eye regions are located in the upper face part, characterized by the changeful value compared to the other face parts. After determining the horizontal position and the face dimensions, a vertical projection is applied to each part of the horizontal position. Owing to the area's brightness between the left and right eyes, the peak of the vertical projection is considered as the face center. Based on the horizontal and vertical projections, two peaks are located as the left and right boundaries of the video frame, and the face width is calculated based on the two lines of the vertical projection. Based on the face segmentations resulted from both vertical and horizontal projections, the widths and heights of both eye regions can be estimated. And then, the regions of interest (ROIs) can be calculated.

b) Pupil Centers Localization: To localize the pupils centers, an eye template must be created to match the eye size on the template and the size of the actual eyes on the video frame. Assume that the two eye regions' created template is denoted as T[i, j], and V[I,j] represents the size of the eyes on the video frame, which is required to detected. To calculate V[i.j], the template can be placed on the video frame by calculating the template intensity values and the corresponding points of the image. Since the template does not absolutely match the video frame, the dissimilarity between the template intensity values with the corresponding intensity values of the video frame is calculated, as follows:

$$\max_{[i,j]\in TR}|I-V|, \sum_{[i,j]\in TR}|I-V|, or \sum_{[i,j]\in TR}(I-V)^2 \tag{1}$$

where TR is the template region.

If there is a match between the template and the video frame. The dissimilarity measure is calculated by Equation (2).

$$\sum_{[i,j]\in TR}(V-T)^2 = \sum_{[i,j]\in TR}V^2 + \sum_{[i,j]\in TR}T^2 - 2\sum_{[i,j]\in TR}(V\times T)^2 \tag{2}$$

Reasonably, this calculation for determining the template instances is to create a general template that can be used for localizing the pupils' centers of the other video frames. This is to decrease the computational overhead of repeating the calculation of pupils' centers each time. If we have a template of MxN dimensions, Equation (3) is used, which is known as cross-correlation.

$$M[i,j] = \sum_{a=1}^{M}\sum_{b=1}^{N}T[a,b]\times V[i+a,j+b] \tag{3}$$

where a and b are the offset between the video frame and the template.

The equations mentioned above are used to calculate the cross-correlation when V and T are fixed. However, in the case of T is fixed, while V is varying, M will depend on V. Therefore, the normalization of the cross-correlation is carried out to solve this problem, as shown in Equation (4) and (5).

$$corr[i,j] = \sum_{a=1}^{M}\sum_{b=1}^{N}T[a,b]\times V[i+a,j+b] \tag{4}$$

$$m[i,j] = \frac{corr_{TV}[i,j]}{\sqrt{\left\{\sum_{a=1}^{M}\sum_{b=1}^{N}V^2[i+a,j+b]\right\}}} \tag{5}$$

2) Eye Movement Classification: Eye movements are classified into three categories: fixation, saccades, or smooth pursuit. The eye regions obtained from the previous stage are used in a multiclass classification for predicting the eye movement on the next stage.

a) Classification Threshold Setting: To classify the eye movements, the proposed system relies on velocity thresholds of eye gaze. It adopts an adaptive threshold for the saccade classification. The saccades are initialized by a user-defined threshold of the time series velocity to a certain value. Based on the specified initial value (TV1), the new threshold is adaptively determined by calculating the variance of sub-thresholds (SV), as given by Equation (6).

$$TV_n = \overline{SV}_{n-1} + St.dev \times \sqrt{\frac{\sum\left(SV_{n-1}-\overline{SV}_{n-1}\right)^2}{N-1}} \tag{6}$$

where St.dev represents the number of standard deviations higher than the average velocity.

This step is repeated until a threshold velocity is stabilized, as given by Equation (7).

$$|TV_n - TV_{n-1}| \prec 1^\circ/\sec \tag{7}$$

The previous steps are carried out to determine the saccades classification. Regarding pursuit classification thresholding, if the velocity exceeds the minimum velocity threshold (pursuit threshold), it is classified as a pursuit. Otherwise, it is classified as a fixation.

b) Convolutional Neural Network (CNN): CNN is known as a feed-forward neural network used for different machine learning purposes. Although one of the drawbacks of CNN is the huge time needed for training, its output accuracy compared to other deep learning models is better. Our proposed system adopts a CNN model by specifying three convolution stages as shown in Fig. 4:

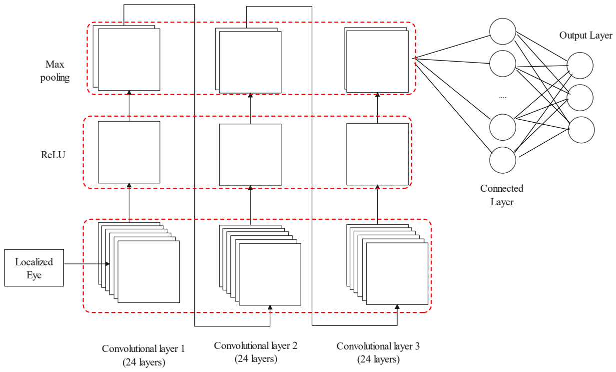


Fig. 4. The CNN Network Architecture.

- The input stage: the video frame is set to have a dimension of 42×50. The first layer utilizes 24 filters, and each filter is of 7×7 dimensions.
- A rectified linear unit (ReLU) stage: In this stage, the ReLU layer is added, and a non-linearity to the activations is introduced, which is mathematically represented as given by Equation (8).

$$f\left(A\right) = max\left(0, \mathrm{A}\right) \tag{8}$$

where A is the input, and f(A) is the output of the ReLU layer.

- A max-pooling stage: In this stage, A max-pooling is followed the ReLU layer. It carries out a spatial sub-sampling of video frames. In this stage 2×2 dimensional max-pooling layers have been used for the reduction of the spatial resolution of the video frames to half.

Moreover, two layers consisting of input, ReLU, and max-pooling layers are added after the previous stages with 5×5 and 3×3 dimensions, respectively. Then, a connected layer receives the output from the previous activation layer. In our case, the output layer consists of three nodes representing the eye movements: saccade, fixation, and smooth pursuit. The softmax loss function is used to measure the error, and the Cross entropy loss function is used as a loss function, which is reduced during the training phase. Equation (9) gives a mathematical representation of the Cross entropy loss function (L).

$$L = -[t\log(f(A)) + (1-t)\log(1-f(A)) \tag{9}$$

where A is the classification vector, t is the truth value taking a value of 0 or 1, and $f(A)$ is the softmax probability of i^{th} class

3) Face Tracking and Detection: This section is dedicated to elaborating on our solution for face tracking and detection. Fig. 5 presents an illustrative demonstration of the proposed solution for face tracking and detection.

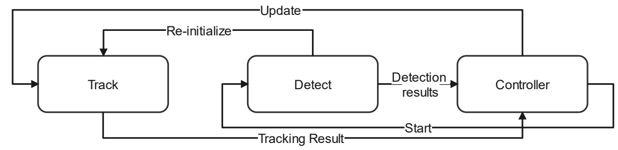


Fig. 5. Face Tracking and Detection Flowchart.

a) Face Tracking: Our proposed solution utilized a kernelized correlation filter (KCF) introduced in [51] for face tracking. Although there is a drawback of KCF when the face is not in the camera's view, it fails and hardly recovered. This challenge is solved using the detection unit, whose results are used to re-initialize the KCF algorithm and update the model with a specific learning rate. The procedure of KCF is summarized in the following steps:

- KCF uses a window that is 2.5 times larger than the given face to cycle shift to obtain training samples. Then, the extracted samples are labeled using Gaussian distribution and are saved in the filter model.
- KCF uses windows, which are 2.5 times larger than the best-obtained result to cycle shift to obtain the samples in the next frames. The similarity is calculated between the target in the KCF model and the samples. Then, the KCF is updated using the obtained result.

b) Face detection: For face detection, we adopt the lightweight multi-task cascade Convolution Neural Network. The following phases briefly describe the overall steps of face detection.

- Phase one: In this phase, a proposal Network is exploited to get the windows obtained from the face tracking unit and the related vector of their bounding boxes. Then. The regression vectors of bounding boxes are used to carry out the calibration of the candidates. The highly overlapping candidates are merged using non-maximum suppression.
- Phase two: The candidates obtained from phase one are used as an input to Refined Network to perform false candidate rejection. The non-rejected candidates with their boxes are calibrated using non-maximum suppression.
- Phase three: In this phase, the outputs of phase two are fed to O-net to obtain more accurate results. In this phase, we ignored the landmarks produced by O-net since we only need bounding surrounding the face to be used by the eye detection module.

The control unit controls the starting point of the face detection unit and updates the face tracking unit.

4) Eye Movement Prediction: This component's main task is to predict eye movement based on the output of head pose estimation and eye data of the classification component. This component works in conjunction with Head pose estimation, whose main task is to detect and estimate the pose direction in order to detect the eye position and direction. Eye movement prediction can reliably track both eyes and detect the eye features in each video frame. However, by accurately detecting the number of head movements in each video frame, the user's eye movement estimation can be carried out accurately, as it is difficult for the users to keep their heads still.

In our proposed system, Kalman filter is modeled by two-state vectors to predict eye movements, as given by Equations (10) and (11), and the transitional state matrix is modeled as given by Equation (12).

$$x_n = \begin{bmatrix} \omega_x(n) \\ \upsilon_x(n) \end{bmatrix} \tag{10}$$

$$y_n = \begin{bmatrix} \omega_y(n) \\ \upsilon_y(n) \end{bmatrix} \tag{11}$$

$$A_n = \begin{bmatrix} 1 & \Delta s \\ 0 & 1 \end{bmatrix} \tag{12}$$

where $\omega_x(n)$ and $\omega_y(n)$ are the eyes' horizontal and vertical positions, $\upsilon_x(n)$ and $\upsilon_y(n)$ represent the horizontal and vertical velocities at time instance n, and Δs represents the sampling interval. The identity matrix for both state vectors is set as given by Equation (13).

$$H_n = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{13}$$

To represent the noise resulted from Arduino board or the camera itself, The standard deviation of measurement noise $R_n = \delta_v^2 = 1°$. Suppose there is a corruption in the eye signal position(the eye noise resulted from eye sub-movements, such as drifting). In that case, the estimation of system noise is done by the following covariance matrix, as given by Equation ((14)).

$$Q_n = \begin{bmatrix} \sigma_w^2 & 0 \\ 0 & \sigma_w^2 \end{bmatrix} \tag{14}$$

where σ_w^2 represents the variance of the system noise.

For eye movements predictions, the following is carried out:

- Saccades prediction: The prediction of saccadic movement is carried out using the method presented in [52]. To detect saccades, the differences between the actual eye velocity and the predicted eye velocity are measured. Two-State Kalman Filter is used to measure the predicted eye-velocity, and a chi-square test is employed to measure the difference between the actual eye velocity and the Kalman-filter-based predicted eye velocity, as follows:

$$V^2 = \sum_{i=1}^{p} \frac{(\hat{v}_2^-(i) - (\frac{z_i - z_{i-1}}{\Delta t}))}{\sigma^2} \tag{15}$$

where $\hat{v}_2^-(i)$ is the predicted velocity and z_i is the measured coordinate of eye position. Δt is the sampling interval and σ^2 is the standard deviation of the actual eye velocity during Δt. We adopted the proposed method in [53] for determining the saccade amplitude using the chi-square value. The proposed method was based on the following formula:

$$Sac\_amp = -0.000024x^6 + 0.0536x^4 + 1.5 \tag{16}$$

$$Sac\_dur = (2.2 \times Sac\_amp\_21)/1000 \tag{17}$$

Once Sac_amp is calculated, the saccade duration (in seconds) can be calculated by Equation (17).

- Fixation prediction: fixation analysis is carried on by updating eye positions. The fixation is predicted when the eye velocity threshold is exceeding $0.5°$ per second for a minimum of 100 ms.
- Smooth prediction: this eye movement is predicted when the eye position does not belong to fixation or saccade eye movements, and the velocity is not exceeding $140°$ per second.

IV. System Implementation & Evaluation

When examining the proposed system, the realistic implementation environment is essential because it is exploratory by nature. This section is devoted to describing the implementation of the proposed system. It clarifies the environment in which the proposed system has been implemented and presents the prototype's functionalities.

A. Implementation Programming Environment

MatLab has been used for the proposed system implementation. This is because most of the proposed system computation is based on arrays (lists), which is simple to be implemented in Matlab. The version used for implementation was MatLab R2018b, which works on Windows 10 platform.

B. System Evaluation

For evaluating the proposed system, many experiments have been carried out to investigate the solution's success in different scenarios. Since the main objective of the proposed system is to classify and predict the movements, it has been evaluated in terms of eye movement classification and prediction. Different evaluation metrics have been employed, and an actual training dataset has been used. The following subsections are devoted to the description of the training dataset and evaluation metrics.

1) Training Dataset Description: The GazeCom dataset is the largest available dataset [54]. The dataset includes 18 recorded videos, and each video has 47 subjects. Far from the screen by 40 cm, the subjects were placed looking at the screen. The size of the screen was 40x30 cm with a resolution of 1280x720. The visual angle covered by the stimulus was 48x27 degrees, and each degree corresponds to 26.7 pixels. The recordings contain only monocular data; the resulted mean validation error was 0.62 degrees. The ARFF format was used to store the recordings due to the access simplicity by programming languages, such as Python and MatLab. The ARFF files contain different data fields, the time in microseconds, x and y positions, the scoring, and combined scoring. The eye movement classes contained in the dataset are unknown, saccades, fixations, smooth pursuit, and noise numbered from 0 to 4, respectively. Other fields presented in the dataset are the direction, velocity, and signal acceleration. The whole dataset has been used to train the Eye Movement Classification module, which runs based on the CNN and RNN models. For testing, we used the data frame recorded by the system camera described above.

2) Evaluation Metrics: The following metrics are used to evaluate the performance of the proposed system in terms of eye movement classification and prediction. It is essential to specify quantitative and qualitative scores for the performance assessment of the proposed system and its counterparts in terms of eye movement classification, which can evaluate the proposed system in terms of eye movement prediction. The following evaluation metrics are selected.

- Pursuits quantitative score (PQnS):
The PQnS metric is used to determine the number of detected eye smooth pursuit, while the smooth pursuit of the stimuli is given. It is calculated by comparing the eye position (x_e, y_e, t) with the corresponding coordinate of smooth pursuit of stimuli (x_s, y_s, t). If the eye position is classified as smooth pursuit, then the smooth pursuit detection counter increases. Equation (18) shows the mathematical representation of PQnS) [51][55].

$$\text{PQnS} = \frac{\text{Saccade\_detection\_counter}}{\text{stimulus\_saccade\_points}} \times 100 \quad (18)$$

The ideal PQnS is calculated by Equation (19).

$$\text{Ideal\_PQnS} = \frac{n \times s_t + \sum_{j=1}^{n} D_{cor\_sac\_dur_j}}{\sum_{j=1}^{n} D_{stim\_pur\_dur_i}} \quad (19)$$

where n donates the pursuits of stimulus, $D_{stim\_pur\_dur_i}$ represents pursuit duration stimulus i, $D_{cor\_sac\_dur_j}$ is the corrective saccade duration and s_t is the pursuit latency.

- Fixation quantitative score (FQnS):
This metric measures the number of detected fixation behavior to the number of presented fixations of stimuli. FQnS is calculated by sampling the fixation stimulus using a similar frequency of eye position. Then, the stimulus fixation coordinate (x_s, y_s) is compared with the recorded eye position coordinates (x_e, y_e). The mathematical representation of FQnS is given by Equation (20) [51][55].

$$\text{FQnS} = \frac{\text{fixation\_detection\_counter}}{\text{stimulus\_stimuli\_points}} \times 100 \quad (20)$$

Where fixation_detection_counter denote the number of fixation points, when stimulus_stimuli_points the number of fixation stimuli is given.

- Saccade quantitative score (SQnS):
The main idea behind SQnS is to measure the number of the detected saccades when the stimuli saccadic behavior is given. The stimuli metric is calculated by considering each jump in the fixation target as a stimuli saccade, and distances between targets are added to the $total\_stimuli\_saccade_amplitude$. Correspondingly, $total\_detected\_saccade\_amplitude$ is the total of the absolute values of the detected saccade amplitudes. The mathematical representation of SQnS is given by Equation (21) [51], [55].

$$SQnS = \frac{total\_\det ected\_saccade}{total\_stimuli\_saccade} \times 100 \quad (21)$$

- Fixation qualitative score (FQlS):
FQlS measures how the detected fixation is approximate to the given stimuli. Therefore, it provides the detected fixation's positional accuracy. Like FQnS, FQIS is calculated for fixation coordinates (x_s, y_s) of the given stimuli compared to the coordinates of eye position. if such coordinates are considered as a fixation, the distance between the centroid of the detected fixation coordinates (x_e, y_e) and given fixation coordinates is calculated. Equation (22) gives the mathematical representation of FQIS [51], [55].

$$FQlS = \frac{\sum_{i=1}^{n} fixation\_dis\tan ce_i}{N} \quad (22)$$

$$fixation\_dis\tan ce_i = \sqrt{(x_s^i - x_c^i) + (y_s^i - y_c^i)} \quad (23)$$

where N is the number of stimuli position coordinates. The Ideal FQlS is 0°, which indicates the absolute accuracy of the system equipment.

- Misclassified fixation score (MisFix):
MisFix is calculated by separately measuring $SP\_fixation\_points$, the number of fixation coordinates in eye position classified as smooth pursuit, and the total number of fixation coordinates in stimuli ($total\_stimuli\_fixation\_points$). Equation (24) gives the mathematical representation of MisFix [51][55].

$$MisFix = \frac{sp\_fixation\_\text{points}}{total\_stimuli\_fixation\_\text{points}} \quad (24)$$

The Ideal MisFix is 0%, meaning that no smooth pursuit was classified during the fixational stimulus.

- Pursuit qualitative scores (PQIS):
PQlS measures how the detected smooth pursuit is approximate to smooth pursuit in the given stimulus. PQlS is measured by $PQlS\_P$ for position and the $PQlS\_V$ for velocity accuracy. The $PQlS\_P$ and $PQlS\_V$ are calculated in a similar way to FQnS. The mathematical representations of $PQlS\_P$ and $PQlS\_V$ are given in Equations (25) and (26) [51], [55].

$$PQIS\_P = \frac{\sum_{i=1}^{N} pursuit\_dis\tan ce_i}{N} \quad (25)$$

$$PQIS\_V = \frac{\sum_{i=1}^{N} pursuit\_speed\_difference_i}{N} \quad (26)$$

$$pursuit\_dis\tan ce_i = \sqrt{(x_s^i - x_c^i) + (y_s^i - y_c^i)^2} \quad (27)$$

$$pursuit\_speed\_difference_i = |\upsilon_s^i - \upsilon_c^i| \quad (28)$$

where N is the number of points of stimuli position. The evaluation ideal scores of this paper is shown in Table I.

TABLE I. THE EVALUATION IDEAL BEHAVIOR SCORES

SQnS	FQnS	PQnS	MisFix	FQIS	PQIS_P	PQIS_V
100	81.599	52.04	7.1	0	0	0

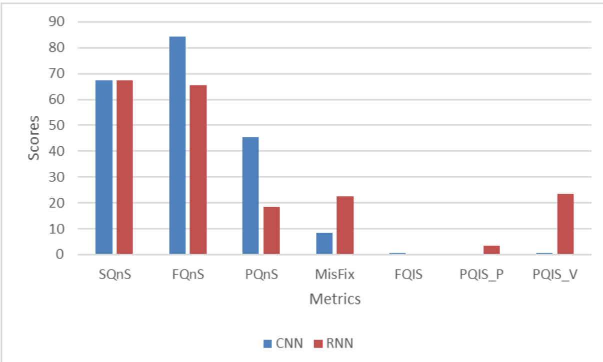


Fig. 6. The Behavioral Classification Scores of the Proposed System and RNN with Sampling Rate 30 Hz and Saccadic Threshold 75 Deg.

3) Classification Evaluation: In this section, the evaluation, in terms of eye movement classification, of the proposed system employing CNN against the RNN model is carried out based on the evaluation metrics given above. To evaluate the performance of the proposed system, the data frame was sampled using different frequencies (30 Hz, 100 Hz, 500 Hz, and 1000Hz).

The behavioral scores introduced in 4.2.2 were utilized to evaluate the effectiveness of the proposed system and its counterpart (RNN) in eye movement classifications.

Figs 6-9 display the obtained classification behavioral scores of both the CNN adopted in our proposed system and the RNN model, and the saccadic threshold is 75 deg. According to the obtained results for sampling rate 30 Hz, the CNN and RNN models achieve almost the same results for SQnS. However, the CNN model outperforms the RNN model in FQnS, PQnS, MisFix, FQIS, $PQIS\_P$, and $PQIS\_V$,

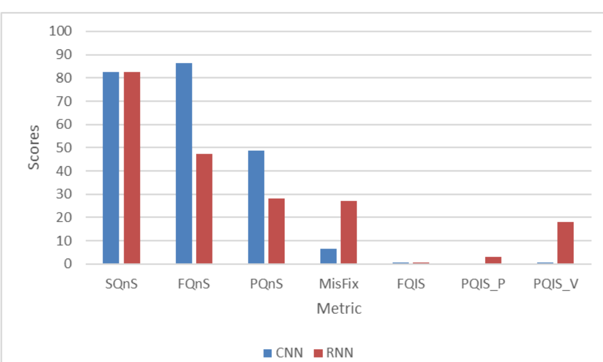


Fig. 7. The Behavioral Classification Scores of the Proposed System and RNN with Sampling Rate 100 Hz and Saccadic Threshold 75 Deg.

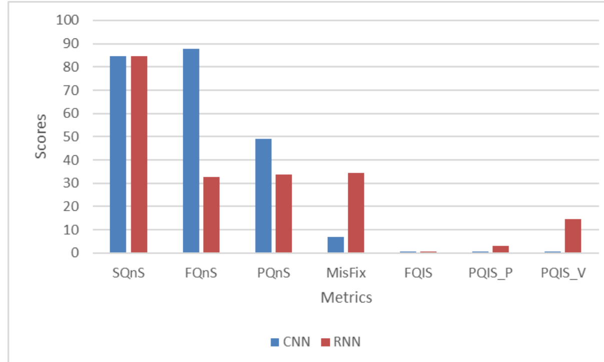


Fig. 8. The Behavioral Classification Scores of the Proposed System and RNN with Sampling Rate 500 Hz and Saccadic Threshold 75 Deg.

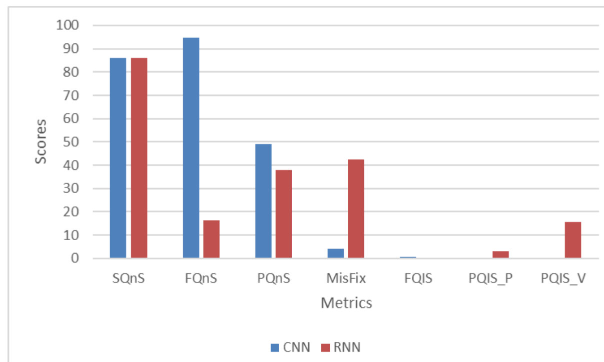


Fig. 9. The Behavioral Classification Scores of the Proposed System and RNN with Sampling Rate 1000 Hz and Saccadic Threshold 75 Deg.

respectively. For 100 Hz, it is noticeable the CNN model achieves better results than the RNN model, but are the close scores for saccades classification metric (SQnS). Similarly, for sampling rates (500 Hz and 1000 Hz), the CNN model shows better results than the RNN model. The RNN model shows an increase in MisFix as the value of the sampling rate increases. Besides, $PQIS\_P$ and $PQIS\_V$ show instability in measuring the pursuit position and velocity. It is worth mentioning that the CNN behavioral classification show proximity to ideal scores as the sampling rate frequency increases.

Moreover, the proposed system against the RNN model is evaluated in terms of eye movement classification with decreasing 10% of the saccadic threshold of the previous evaluation. As mentioned above, the evaluation is carried with different sampling rates, and the saccadic threshold is 65 degrees. We selected this slight change in the saccade threshold to determine the changes of eye movement classification that can take place with different classification thresholds of saccade movement. Fig. 10 to 13 depict the obtained classification behavioral scores of both the CNN adopted in our proposed system and the RNN model.

However, compared to the RNN model, the CNN model

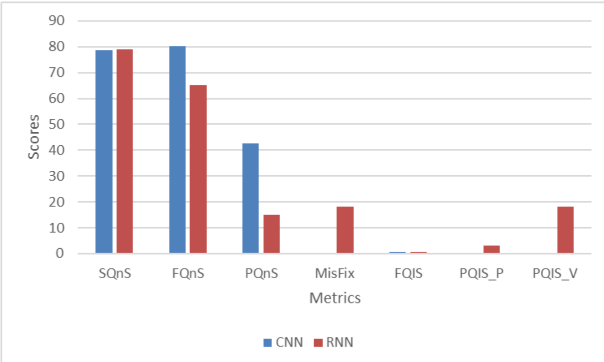


Fig. 10. The Behavioral Classification Scores of the Proposed System and RNN with Sampling Rate 30 Hz and Saccadic Threshold 65 Deg.

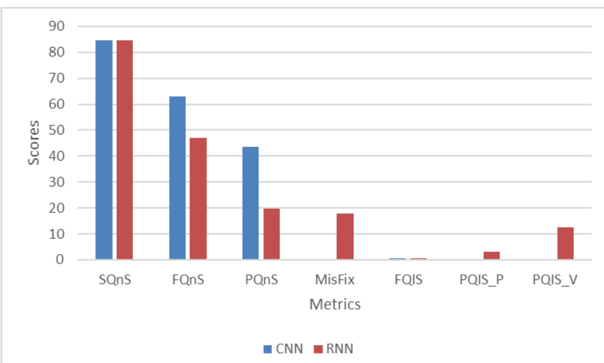


Fig. 11. The Behavioral Classification Scores of the Proposed System and RNN with Sampling Rate 100 Hz and Saccadic Threshold 65 Deg.

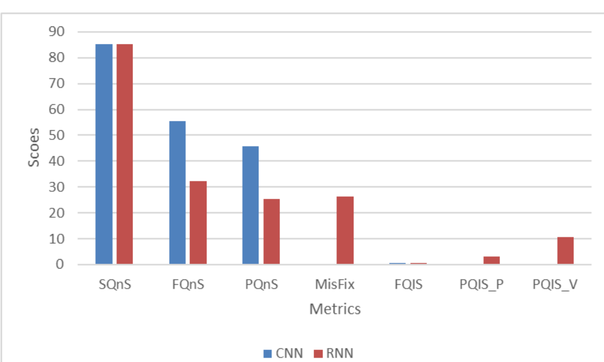


Fig. 12. The Behavioral Classification Scores of the Proposed System and RNN with Sampling Rate 500 Hz and Saccadic Threshold 65 Deg.

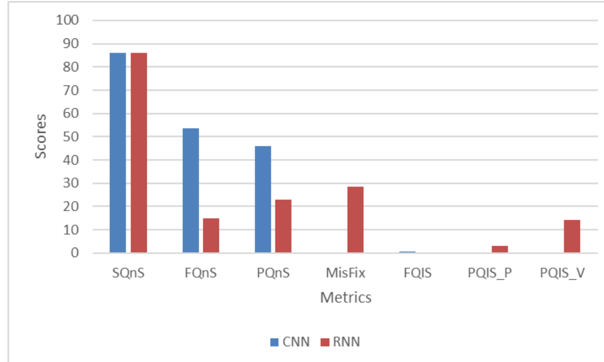


Fig. 13. The Behavioral Classification Scores of the Proposed System and RNN with Sampling Rate 1000 Hz and Saccadic Threshold 65 Deg.

achieves better results as indicated by the behavioral scores. Both models achieve almost the same SQnS scores for the given sampling frequencies. The other behavioral scores (FQnS, PQnS, FQIS, $PQIS\_P$, $PQIS\_V$) show that the achieved scores decrease as the sampling frequencies increase. This leads us to conclude that the saccadic threshold has effects on the behavioral classification scores as it decreases. It can noticed that the CNN model outperforms the RNN model in achieving more reliable results. Table II gives the behavioral classification scores of the proposed system and RNN with various sampling rates, and the saccadic threshold is 75 deg.

Compared with the results listed in Table II, the eye movement classification performance of both the CNN and RNN models is notably decreased with the saccadic threshold decrease.

C. Prediction Evaluation

The proposed system utilized the Kalman filter for eye-movement prediction with two states (velocity and position). The Kalman filter is applied to the recorded eye movement to predict the eye velocity and position. Then, the Chi-square test uses the measured and predicted position and velocity values to classify each positional value as fixation, saccade, and smooth pursuit. The eye movement prediction module of the proposed system is evaluated using two scenarios.

The first scenario was carried with a saccadic threshold of 75 degrees. The second scenario was carried out with a saccadic threshold of 65 degrees. Both scenarios were implemented with various sampling rate frequencies (30 Hz, 100 Hz, 500 Hz, and 1000 Hz). Figs 14-17 illustrate the behavioral scores of both scenarios.

As it can be seen, the different thresholds influence the Kalman filter prediction scores.The SQnS score has shown no changes in both scenarios for sampling frequencies (30 Hz, 100 Hz, and 500 Hz). As the sampling frequency increased to 1000Hz, the results have shown remarkable changes. For FQnS, it can be noticed that capability of the Kalman filter in classifying fixation is influenced by the increase in the sampling frequencies. This indicates that the accuracy of the

TABLE II. THE BEHAVIORAL CLASSIFICATION SCORES OF THE PROPOSED SYSTEM AND RNN WITH VARIOUS SAMPLING RATES AND SACCADIC THRESHOLD 75 DEG

sampling rates		SQnS	FQnS	PQnS	MisFix	FQIS	PQIS_P	PQIS_V
30 Hz	CNN	67.2945	84.3947	45.32	8.42	0.7092	0.2	0.5
	RNN	67.3928	65.4482	18.5856	22.4984	0.1772	3.3521	23.5821
100Hz		SQnS	FQnS	PQnS	MisFix	FQIS	PQIS_P	PQIS_V
	CNN	82.495	86.4211	48.57	6.5	0.578	0.34	0.5
	RNN	82.6959	47.4095	28.0949	27.1789	0.5466	3.0109	17.9955
500Hz		SQnS	FQnS	PQnS	MisFix	FQIS	PQIS_P	PQIS_V
	CNN	84.5859	87.7445	48.93	7.02	0.5543	0.7	0.5
	RNN	84.6112	32.7983	33.8999	34.3412	0.5508	3.1152	14.4503
1000Hz		SQnS	FQnS	PQnS	MisFix	FQIS	PQIS_P	PQIS_V
	CNN	85.9343	94.6117	49.02	4.2	0.4984	0	0
	RNN	85.9343	16.2093	38.0532	42.3631	0.4505	3.1789	15.5107

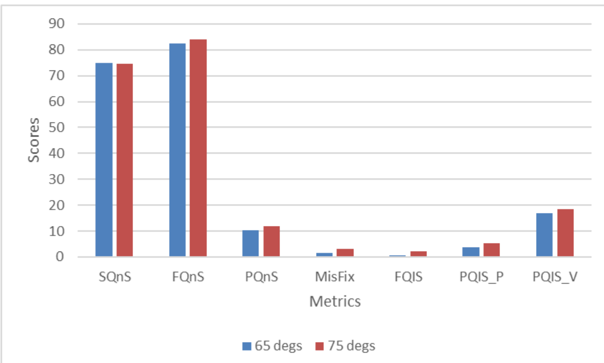


Fig. 14. The Behavioral Classification Scores with a Sampling Rate of 30 Hz and Saccadic Threshold 75 and 65 Degs.

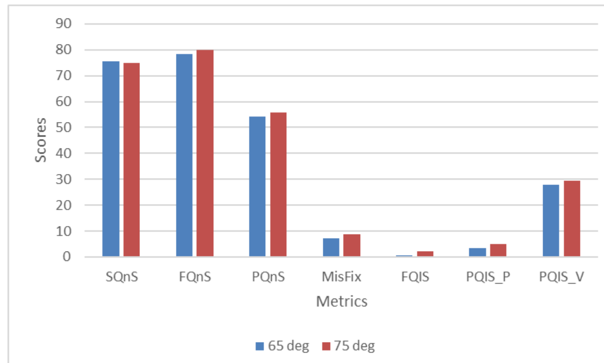


Fig. 16. The Behavioral Classification Scores with a Sampling Rate of 500 Hz and Saccadic Threshold 75 and 65 Degs.

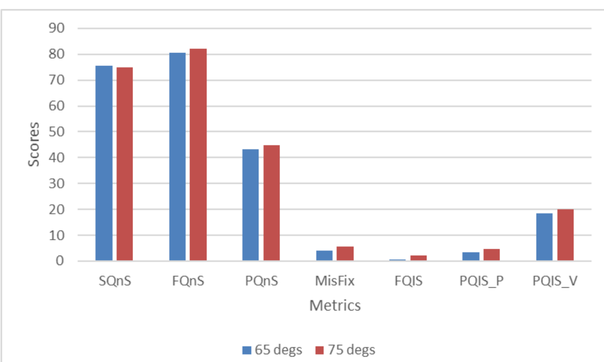


Fig. 15. The Behavioral Classification Scores with a Sampling Rate of 100 Hz and Saccadic Threshold 75 and 65 Degs.

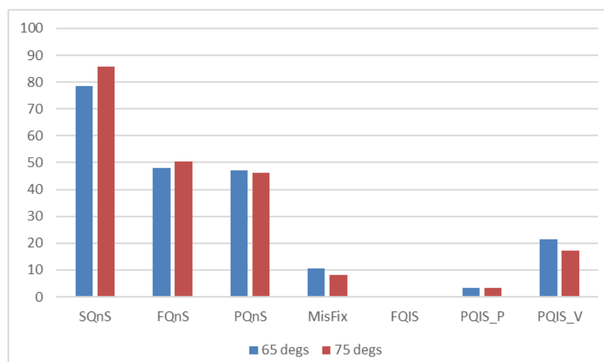


Fig. 17. The Behavioral Classification Scores with a Sampling Rate of 1000 Hz and Saccadic Threshold 75 and 65 Degs.

Kalman filter in fixation classification decreased as the sampling frequencies increase. Unlike FQnS, PQnS is increased in both scenarios. This indicates that the Kalman filter accuracy in pursuit classification as the sampling frequencies increase. However, the Kalman filter shows high fixation misclassification, as shown by MisFix scores. Moreover, the Kalman filter shows a deviation in predicting pursuit positional and velocity, as indicated by $PQIS\_P$ and $PQIS\_V$ scores, in both scenarios. This can happen in a case of a very high position detected before a very low position.

V. CONCLUSION AND FUTURE WORK

A. Conclusion

This paper proposed a system for eye movement analysis and prediction. The system proposed in this paper is a comprehensive system for three-type eye movement classification and prediction data consistency with a various and wide range of components. It combines two essential components: hardware and software. Each component encompasses different subcomponents or modules, and each module has different functions. The design of the proposed system relies on software designing architecture. We exploited one software architecture, namely, module architecture. As a result, the proposed system's design is highly flexible, and components can be accessed and used individually or collectively. Based on the obtained results, the proposed system accomplished good success in eye-movement classification and prediction.

B. Future work

Eye movement classification and prediction is still an active research area. According to the work achieved in this paper, some research directions could be proposed.

- Empirical studies could cover a wide range of deep learning techniques to test and evaluate the behavior of the proposed system in eye-movement classification. Moreover, different lightweight techniques can be implemented to perform eye movement prediction.
- A recommended future research be done by employing different techniques for face tracking and detection to detect the face with different angles.
- Further recommended studies can be introduced to eye detection and localization. Besides, improvement of the developed prototype can be conducted. This includes the optimization of source Code and improvements of code speed, the implementation of the developed prototype in an implementation environment that support parallelization.

References

[1] K. Harezlak and P. Kasprowski, "Application of eye tracking in medicine: A survey, research issues and challenges," *Computerized Medical Imaging and Graphics*, vol. 65, pp. 176–190, 2018.

[2] N. Hristozova, P. Ozimek, and J. P. Siebert, "Efficient egocentric visual perception combining eye-tracking, a software retina and deep learning," *arXiv preprint arXiv:1809.01633*, 2018.

[3] B. H. Ulutas, N. F. Özkan, and R. Michalski, "Application of hidden markov models to eye tracking data analysis of visual quality inspection operations," *Central European Journal of Operations Research*, pp. 1–17, 2019.

[4] M. Cutumisu, K.-L. Turgeon, T. Saiyera, S. Chuong, L. M. González Esparza, R. MacDonald, and V. Kokhan, "Eye tracking the feedback assigned to undergraduate students in a digital assessment game," *Frontiers in psychology*, vol. 10, p. 1931, 2019.

[5] Z. CHEN, H. FU, W.-L. LO, and Z. CHI, "Artificial intelligence in medical applications," *Journal of Healthcare Engineering*, vol. 2018, p. 7692198, 2018.

[6] K. Lukander, "A short review and primer on eye tracking in human computer interaction applications," *arXiv preprint arXiv:1609.07342*, 2016.

[7] Y. Hang, X. Yi, and C. Xianglan, "Eye-tracking studies in visual marketing: Review and prospects," *Foreign Economics & Management*, vol. 40, no. 12, pp. 98–108, 2018.

[8] T. Colliot and E. Jamet, "Understanding the effects of a teacher video on learning from a multimedia document: an eye-tracking study," *Educational Technology Research and Development*, vol. 66, no. 6, pp. 1415–1433, 2018.

[9] U. Obaidellah, M. Al Haek, and P. C.-H. Cheng, "A survey on the usage of eye-tracking in computer programming," *ACM Computing Surveys (CSUR)*, vol. 51, no. 1, pp. 1–58, 2018.

[10] R. B. Noland, M. D. Weiner, D. Gao, M. P. Cook, and A. Nelessen, "Eye-tracking technology, visual preference surveys, and urban design: preliminary evidence of an effective methodology," *Journal of Urbanism: International Research on Placemaking and Urban Sustainability*, vol. 10, no. 1, pp. 98–110, 2017.

[11] V. Clay, P. König, and S. Koenig, "Eye tracking in virtual reality," *Journal of Eye Movement Research*, vol. 12, no. 1, 2019.

[12] A. George, "Image based eye gaze tracking and its applications," *arXiv preprint arXiv:1907.04325*, 2019.

[13] R. G. Lupu and F. Ungureanu, "A survey of eye tracking methods and applications," *Buletinul Institutului Politehnic din Iasi, Automatic Control and Computer Science Section*, vol. 3, no. 1, pp. 72–86, 2013.

[14] H. Azami, S. E. Arnold, S. Sanei, Z. Chang, G. Sapiro, J. Escudero, and A. S. Gupta, "Multiscale fluctuation-based dispersion entropy and its applications to neurological diseases," *IEEE Access*, vol. 7, pp. 68718–68733, 2019.

[15] Y. Terao, H. Fukuda, and O. Hikosaka, "What do eye movements tell us about patients with neurological disorders?—an introduction to saccade recording in the clinical setting—," *Proceedings of the Japan Academy, Series B*, vol. 93, no. 10, pp. 772–801, 2017.

[16] K. Hercegfi, A. Komlódi, M. Köles, and S. Tóvölgyi, "Eye-tracking-based wizard-of-oz usability evaluation of an emotional display agent integrated to a virtual environment," *Acta Polytechnica Hungarica*, vol. 16, no. 2, pp. 145–162, 2019.

[17] P. Beach and J. McConnel, "Eye tracking methodology for studying teacher learning: A review of the research," *International Journal of Research & Method in Education*, vol. 42, no. 5, pp. 485–501, 2019.

[18] K. Harezlak and P. Kasprowski, "Application of eye tracking in medicine: A survey, research issues and challenges," *Computerized Medical Imaging and Graphics*, vol. 65, pp. 176–190, 2018.

[19] P. Bhatia, A. Khosla, and G. Singh, "A review on eye tracking technology," in *Interdisciplinary Approaches to Altering Neurodevelopmental Disorders*, pp. 107–130, IGI Global, 2020.

[20] D. W. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 3, pp. 478–500, 2009.

[21] D. W. Hansen and A. E. Pece, "Eye tracking in the wild," *Computer Vision and Image Understanding*, vol. 98, no. 1, pp. 155–181, 2005.

[22] K.-N. Kim and R. Ramakrishna, "Vision-based eye-gaze tracking for human computer interface," in *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028)*, vol. 2, pp. 324–329, IEEE, 1999.

[23] R. Kothari and J. L. Mitchell, "Detection of eye locations in unconstrained visual images," in *Proceedings of 3rd IEEE International Conference on Image Processing*, vol. 3, pp. 519–522, IEEE, 1996.

[24] A. Peréz, M. L. Córdoba, A. Garcia, R. Méndez, M. Munoz, J. L. Pedraza, and F. Sanchez, "A precise eye-gaze detection and tracking system," 2003.

[25] Q. Ji, "Special issue on eye detection and tracking," *Computer vision and image understanding (Print)*, vol. 98, no. 1, 2005.

[26] J. Zhu and J. Yang, "Subpixel eye gaze tracking," in *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*, pp. 131–136, IEEE, 2002.

[27] B. Huang, R. Chen, Q. Zhou, and W. Xu, "Eye landmarks detection via weakly supervised learning," *Pattern Recognition*, vol. 98, p. 107076, 2020.

[28] R. Stiefelhagen, J. Yang, and A. Waibel, "Tracking eyes and monitoring eye gaze," in *Proc. Workshop on Perceptual User Interfaces*, pp. 98–100, 1997.

[29] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, vol. 1, pp. I–I, Ieee, 2001.

[30] S. Yeamkuan and K. Chamnongthai, "Fixational feature-based gaze pattern recognition using long short-term memory," in *2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 1–4, IEEE, 2020.

[31] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[32] P. Li, X. Hou, X. Duan, H. Yip, G. Song, and Y. Liu, "Appearance-based gaze estimator for natural interaction control of surgical robots," *IEEE Access*, vol. 7, pp. 25095–25110, 2019.

[33] D. W. Hansen, D. J. MacKay, J. P. Hansen, and M. Nielsen, "Eye tracking off the shelf," in *Proceedings of the 2004 symposium on Eye tracking research & applications*, pp. 58–58, 2004.

[34] D. W. Hansen, J. P. Hansen, M. Nielsen, A. S. Johansen, and M. B. Stegmann, "Eye typing using markov and active appearance models," in *Sixth IEEE Workshop on Applications of Computer Vision, 2002.(WACV 2002). Proceedings.*, pp. 132–136, IEEE, 2002.

[35] W.-B. Horng, C.-Y. Chen, Y. Chang, and C.-H. Fan, "Driver fatigue detection based on eye tracking and dynamic template matching," in *IEEE International Conference on Networking, Sensing and Control, 2004*, vol. 1, pp. 7–12, IEEE, 2004.

[36] S. Zhao, J. Luo, and S. Wei, "A hybrid eye movement feature recognition of classroom students based on machine learning," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 2, pp. 2803–2813, 2021.

[37] S. Tsekeridou and I. Pitas, "Facial feature extraction in frontal views using biometric analogies," in *9th European Signal Processing Conference (EUSIPCO 1998)*, pp. 1–4, IEEE, 1998.

[38] N. Tsumura, C. Endo, H. Haneishi, and Y. Miyake, "Image compression and decompression based on gazing area," in *Human Vision and Electronic Imaging*, vol. 2657, pp. 361–367, International Society for Optics and Photonics, 1996.

[39] K. Eibenberger, B. Eibenberger, D. C. Roberts, T. Haslwanter, and J. P. Carey, "A novel and inexpensive digital system for eye movement recordings using magnetic scleral search coils," *Medical & biological engineering & computing*, vol. 54, no. 2-3, pp. 421–430, 2016.

[40] P. J. Murphy, A. L. Duncan, A. J. Glennie, and P. C. Knox, "The effect of scleral search coil lens wear on the eye," *British journal of ophthalmology*, vol. 85, no. 3, pp. 332–335, 2001.

[41] M. Frens and J. Van der Geest, "Scleral search coils influence saccade dynamics," *Journal of neurophysiology*, vol. 88, no. 2, pp. 692–698, 2002.

[42] T. Imai, K. Sekine, K. Hattori, N. Takeda, I. Koizuka, K. Nakamae, K. Miura, H. Fujioka, and T. Kubo, "Comparing the accuracy of video-oculography and the scleral search coil system in human eye movement analysis," *Auris Nasus Larynx*, vol. 32, no. 1, pp. 3–9, 2005.

[43] T. Santini, W. Fuhl, D. Geisler, and E. Kasneci, "Eyerectoo: Open-source software for real-time pervasive head-mounted eye tracking.," in *VISIGRAPP (6: VISAPP)*, pp. 96–101, 2017.

[44] M. Pavec, J. Navratil, R. Soukup, and A. Hamacek, "A bowtie antenna prepared by aerosol jet and embroidering technology," in *2018 41st International Spring Seminar on Electronics Technology (ISSE)*, pp. 1–4, IEEE, 2018.

[45] C. R. Picanço and F. Tonneau, "A low-cost platform for eye-tracking research: Using pupil© in behavior analysis," *Journal of the experimental analysis of behavior*, vol. 110, no. 2, pp. 157–170, 2018.

[46] S. Anders, N. Weiskopf, D. Lule, and N. Birbaumer, "Infrared oculography—validation of a new method to monitor startle eyeblink amplitudes during fmri," *Neuroimage*, vol. 22, no. 2, pp. 767–770, 2004.

[47] H. Kimmig, M. W. Greenlee, F. Huethe, and T. Mergner, "Mr-eyetracker: a new method for eye movement recording in functional magnetic resonance imaging," *Experimental Brain Research*, vol. 126, no. 3, pp. 443–449, 1999.

[48] M. Khamis, F. Alt, and A. Bulling, "The past, present, and future of gaze-enabled handheld mobile devices: Survey and lessons learned," in *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pp. 1–17, 2018.

[49] D. R. van Renswoude, M. E. Raijmakers, A. Koornneef, S. P. Johnson, S. Hunnius, and I. Visser, "Gazepath: An eye-tracking analysis tool that accounts for individual differences and data quality," *Behavior research methods*, vol. 50, no. 2, pp. 834–852, 2018.

[50] S. Goni, J. Echeto, A. Villanueva, and R. Cabeza, "Robust algorithm for pupil-glint vector detection in a video-oculography eyetracking system," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, vol. 4, pp. 941–944, IEEE, 2004.

[51] O. V. Komogortsev and A. Karpov, "Automated classification and scoring of smooth pursuit eye movements in the presence of fixations and saccades," *Behavior research methods*, vol. 45, no. 1, pp. 203–215, 2013.

[52] D. Sauter, B. Martin, N. Di Renzo, and C. Vomscheid, "Analysis of eye tracking movements using innovations generated by a kalman filter," *Medical and biological Engineering and Computing*, vol. 29, no. 1, pp. 63–69, 1991.

[53] O. V. Komogortsev and J. I. Khan, "Kalman filtering in the design of eye-gaze-guided computer interfaces," in *International Conference on Human-Computer Interaction*, pp. 679–689, Springer, 2007.

[54] M. Dorr, T. Martinetz, K. R. Gegenfurtner, and E. Barth, "Variability of eye movements when viewing dynamic natural scenes," *Journal of vision*, vol. 10, no. 10, pp. 28–28, 2010.

[55] R. Leigh and D. Zee, "The neurology of eye movements vol. 90: Oxford university press," 2015.

Independent Channel Residual Convolutional Network for Gunshot Detection

Jakub Bajzik[1], Jiri Prinosil[2], Roman Jarina[3], Jiri Mekyska[4]
Dept. of Mechatronics and Electronics, University of Zilina
Zilina 010 26, Slovakia[1]
Dept. of Telecommunications, Brno University of Technology
601 90 Brno, Czech Republic[2,4]
Dept. of Multimedia and Information and Communication Technology
University of Zilina, Zilina 010 26, Slovakia[3]

***Abstract*—The main purpose of this work is to propose a robust approach for dangerous sound events detection (e.g. gunshots) to improve recent surveillance systems. Despite the fact that the detection and classification of different sound events has a long history in signal processing, the analysis of environmental sounds is still challenging. The most recent works aim to prefer the time-frequency 2-D representation of sound as input to feed convolutional neural networks. This paper includes an analysis of known architectures as well as a newly proposed Independent Channel Residual Convolutional Network architecture based on standard residual blocks. Our approach consists of processing three different types of features in the individual channels. The UrbanSound8k and the Free Firearm Sound Library audio datasets are used for training and testing data generation, achieving a 98 % F1 score. The model was also evaluated in the wild using manually annotated movie audio track, achieving a 44 % F1 score, which is not too high but still better than other state-of-the-art techniques.**

***Keywords*—Acoustic signal processing; gunshot detection systems; audio signal analysis; machine learning; deep learning; residual networks**

I. Introduction

In the field of signal processing, the audio data analysis takes an extensive part, which is constantly studied. Many machine learning-based algorithms were proposed for solving tasks such as classification, segmentation, and denoising. In many cases, the methods are adapted for a specific type of sound, mainly speech and music, which take an extensive part in the research. On the other hand, the environmental sounds are unstructured, and it is challenging to generalize their nature. Many environmental sound analysis applications, ranging from urban monitoring [1] to IoT [2] and surveillance system [3], [4], [5], have been developed within the past years.

However, in recent years it has become a topical task to use known techniques to classify environmental sounds like explosion, gunshot, siren, car alarm, baby crying, window breakage and other events associated with potential danger [6]. Usage of learning algorithms tends to increase personal safety. The possible implementations are in-home or industrial protection systems, in cars to alert deaf or poorly hearing drivers of the siren, in homes to alert a parent of a crying child, and in a wide range of assistive devices, especially for deaf people. Recent modern surveillance systems for risk prevention purposes focuses mainly on the analysis of video signals from cameras using advanced computer vision techniques [7], [8], [9]. However, the analysis of audio signals has considerable potential in these systems as well. Especially gunshot detection technologies have been increasingly adopted by law enforcement agencies for mapping the spatial and temporal patterns of gun violence [10]. The general problem of gun detection technologies is a high rate of false alarms resulting in the waste of police resources when responding to those false alerts [11]. From a practical point of view, it is necessary to minimize the amount of false-positive predictions. Therefore, when setting the operating point of the system in practical applications, the false-positive rate must be taken into account.

The main objective of our work is to propose a method for detecting danger-related audio events (gunshots) that achieve high specificity in real conditions. We also aim to explore several types of feature spaces and neural architectures and discuss, what kind of setup is the most suitable for such an application.

The standard approach for sound analysis is to collect a single vector of features. The often used classifiers are Support Vector Machines (SVM), Deep Neural Networks (DNN), or multilayer perceptrons. When processing environmental sounds, the uniform structure can not be expected as in speech or music, where the signal contains a harmonic structure or repetitions. The features that perform well in specific applications may be inadequate for sounds with other nature and vice versa.

In our work we are using several 2-dimensional sound representations such as spectrograms as well as the standard 1-D approach and analyse the performance in the gunshot detection application. In addition to the frequently used spectrogram [12], [13], [14], also new visualizations and advanced methods for feature processing are used. The main contributions of our work are:

- Exploring the suitability of several state-of-the-art convolutional networks based approaches for gunshot detection.
- Proposing the convolutional model for boosting the performance on 2-dimensional independent feature spaces.

The signal is transformed into three independent audio feature sets forming an "RGB image" that is suitable for processing by common 2-D convolutional networks used for image processing. Unlike image processing, where color channels are highly correlated and processed together after the first DNN layer, the three audio feature sets are processed independently by the first two DNN blocks. Our proposed architecture is based on standard residual units. The important task is not only increasing the number of true predictions but also reducing the number of false-positive gunshot predictions.

II. Related Work

Over the years several works dealing with gunshots detection from audio signal have been published. Most of them are based on the extraction of handcrafted acoustic features and the use of machine learning techniques for the task of classification. A combination of 7 Linear Predictive Coding (LPC) coefficients and 13 Mel Frequency Cepstral Coefficients (MFCC) with SVM classifier is used in work [15] providing 8 % false alarms rate on a custom dataset. The author of [16] extended the previous feature set by Linear Predictive Coding Cepstral (LPCC) and auto-correlation coefficients reaching 82 % accuracy and 70 % precision on a combination of public available datasets. The same author then evaluated the effect of individual features on the accuracy of the classification task [17], considering the features with the best score to be the first five coefficients of the 24$^{\text{th}}$ order LPC. A large set of various acoustic features with Hidden Markov Model (HMM) and Viterbi decoder is used in EAR-TUKE system [18] for detecting gunshots and glass breaking events with 98 % accuracy in records with Signal to Noise Ratio (SNR) $\geq$ 20 dB. In case of microphones arrays it is possible to use a two stage methodology comprising of a Blind System Identification and Deconvolution (BSID) stage followed by a SVM-based classification [19] for gunshot detection in a noisy urban environment. In [20] a method of classifying impulsive sounds based on a Weighted Majority Voting (WMV) strategy is described. In [21] Convolutional Neural Network (CNN) with temporal and spectral features is used for gunshot sound categories classification (pistol, rifle and shotgun of different calibres) reaching over 90 % accuracy. Another CNN approach detects gunshots with 99 % accuracy and low false alarm rate using the ResNet architecture [22]. Most of the above approaches work with database recordings that contain a low level of environmental noise. In the case of real applications, this condition can hardly be met. For this reason, dealing with the detection and classification of environmental noisy sounds is important.

The datasets of environmental sounds are made mainly for learning algorithms that perform Environmental Sound Classification (ESC) task. One of the widely used datasets for ESC is the UrbanSound8K [23], further described in Section III-E.

The signal representation of audio is related to the architecture of the learning algorithm and learning objective. The standard process that follows the approaches from speech and music analysis is to collect a single vector of features. Short-term or long-term features may not always generalize the unstructured nature of environmental sounds.

The direct solution of the feature extraction problem is to build a model that operates on the raw audio signals directly. The 1-D CNNs can handle the internal representation of the input signal, which allows end-to-end usage. The important advantage is that there is no need to transform or pre-process the data, and such a model can adapt to a variety of audio signals. In the studies [24], [25], the first end-to-end ESC architecture called EnvNet was proposed in versions 1 and 2. Another 1-D architecture was proposed in [26], where the audio signal was processed at different time scales. The study [27] presents an end-to-end 1-D convolutional network that has fewer parameters compared to dense 2-D convolutional neural networks and does not require a large amount of training data. It reaches 87 % mean accuracy on the UrbanSound8K dataset with random weights initialization and 89 % with weights initialization by Gammatone filter bank coefficients [28] synthesizing an impulse response from nerve cells in the auditory fiber [29].

The 2-D CNN models operate on the pre-computed feature representations obtained by a fixed process of extraction. The study [13] is the first which deals with ESR using CNN trained on mel-scaled spectrograms. Such an approach is extended in study [12], where different augmentation methods are used. In work [14], the authors presented the model ESResNet based on STFT, that outperforms recently known approaches with ESC datasets achieving 82% accuracy when trained from scratch and 85% accuracy with ImageNet weights initalization on UrbanSound8K dataset. Since the Piczak's work [13], the research trend in environmental sound analysis seems to be the usage of 2-D feature matrices for feeding 2-D CNNs [30].

III. Materials and Methods

A. Signal Model and Problem Formulation

Following the recent studies [12], [13], [14], [31], [32], [33], [34], the most suitable setup for environmental sound recognition employs a 2-D convolutional neural network fed by a Time-Frequency representation of the audio signal (further mentioned as audio features). In order to be processed by a 2-D convolutional network, these audio features need to be converted into a suitable uniform 2-D representation. Based on this 2-D representation, it is then necessary to choose an optimal architecture of the convolutional network for the classification task. The whole workflow from audio samples to predictions is depicted in Fig. 1.

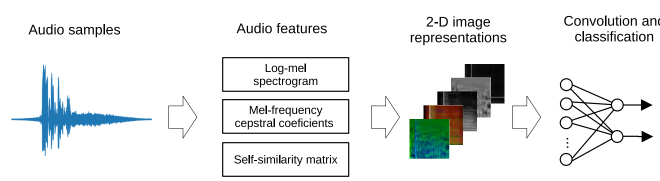


Fig. 1. Feature Extraction and Classification Workflow.

B. Audio Features

In the case of audio signal processing, there is a large number of various audio features. The Log-Mel Spectrograms (LM Spec) and Mel Frequency Cepstral Coefficients (MFCC) are among the most commonly used audio features. In our work, we additionally include the Self-Similarity Matrix (SSM), which is frequently used to analyze the global structure of musical works, to the audio feature list. The hyperparameters for feature extraction are detailed in Table I.

1) Log-Mel Spectrogram: The spectrogram is the most commonly used audio signal visualization. It shows the frequency spectrum change over time. A Short Time Fourier Transform (STFT) is used to convert the signal from time to frequency domain. Additional mel-frequency scale transform 1 is applied to embrace the psychoacoustic knowledge.

$$f_{mel} = 2595 \cdot log_{10}\left(1 + \frac{f_{Hz}}{700}\right) \tag{1}$$

2) Mel Frequency Cepstral Coefficient: MFCC reflects the non-linear and masking psychoacoustic characteristics of human hearing. MFCC coefficients are obtained by multiplying the signal spectrum by a mel-scale distributed filter bank, logarithm and Discrete Cosine Transform (DCT).

3) Self-Similarity Matrix: SSM is the measure of self-similarity of the signal based on distances. We use a self-similarity matrix to display signal correlation. To visualize the self-similarity, we use a matrix S defined by Equation 2. The matrix dimensions $N \times N$ depend on the number of signal samples. For reducing the computational complexity, a self-similarity matrix S is computed on downsampled envelope $s = (s_1, s_2, s_3, ..., s_N)$ of input audio signal, obtained using the Hilbert transform. As a measure of similarity we are using the absolute distance.

$$S(i, j) = |s_i - s_j| \quad i, j = 1, ..., N \tag{2}$$

The vertical and horizontal axes represent the time sequence. The matrix is symmetrical by the main diagonal where the similarity is maximal.

TABLE I. HYPERPARAMETERS FOR FEATURE EXTRACTION

Features	FFT length	Banks	Window
Log-mel spectrogram	2048	256	Hamming
MFCC	2048	20	Hamming
Self-similarity	-	-	-

C. 2-D Feature Representation

Audio features are extracted as 2-D matrices and aligned for convolutional neural network input. Since most 2D convolutional network architectures were primarily designed for image processing, they expect 3 sets of 2-D feature matrices at the input (an analogy to RGB channels of images).

1) Band-Splitted Spectrogram (BS Spec): In our experiments, we are using the log-mel spectrogram split to three frequency bands (high, middle, low) aligned with the RGB color channels (each band as one color channel). The band cutting frequencies depends the on maximal frequency $f_{max} = \frac{f_s}{2}$ given by the sampling rate (Fig. 2).

The same principle was used in study [14], where authors explain the usage of the band-splitted spectrogram for avoiding redundancy. Other solutions are replicating the spectrogram or passing zeros.

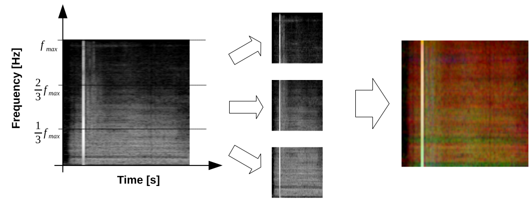
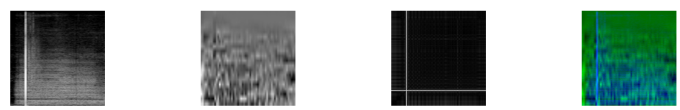


Fig. 2. Band-split Log-mel Spectrogram and Resulting RGB Image.

2) Independent Feature Spaces (IFS): The combination of the log-mel spectrogram, MFCC and self-similarity matrix represents the independent feature spaces. The similar method was used in study [32]. We assume that the MFCC and SSM will help to classify non-impulsive background sounds. The harmonicity of the gunshot signal is low, so the SSM is almost empty, while the background noise results in a visible grid.

Three feature matrices are overlapped in matched time positions. It means, that the x-axis resolutions are approximately the same for all matrices. However, on the y-axis we have different dimensions when using spectrogram (frequency), MFCC (mel banks) and SSM (time) (Fig. 3).



(a) (b) (c) (d)

Fig. 3. Independent Feature Spaces as RGB Image Channels. (a) Red Channel, Log-mel Spectrogram. (b) Green Channel, MFCCs. (c) Blue Channel, Self-similarity Matrix. (d) Resulting RGB Image.

D. Convolutional Neural Networks Architectures

The most widely used convolutional neural network models for 2-D feature space classification are based on the residual network architecture (ResNet). However, there is also an approach that uses only a 1-dimensional convolutional network fed directly by a raw audio signal to classify environmental sounds. In addition, we include a custom approach based on residual networks where individual channels are processed independently.

1) Residual Networks: Following recent studies [14], the residual models perform well on environmental sound classification. The residual network was designed as a network in a network, which means that the lower layer's inputs are connected to the outputs of the two higher layers. The example of the standard residual block is shown in Fig. 4.

The skip connections defined as

$$y = F(x) + x \tag{3}$$

are also called shortcut connections. The function F represents the convolution operations. The shortcut connections help to eliminate the problem of vanishing gradient in deep neural networks. The authors of [35] designed ResNets with a different number of layers, specifically 18, 34, 50, 101 and 152.

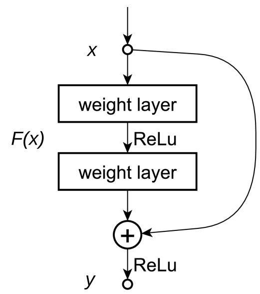


Fig. 4. Standard Residual Block [35].

2) End-to-End Classification using a 1-D CNN: In this approach, a 1-D convolutional (1-D CNN) neural network learns low-level and high-level information directly from the audio signal waveform. Since the size of the input data to the amount of data is in imbalance, it is not recommended to use too deep convolutional network architectures to avoid significant overfitting. The study [27] presents the optimal architecture with respect to the sampling frequency of the audio signal at different audio signal lengths. For the sampling frequency of 16 kHz considered further in this paper, it has been shown that the best score is achieved at the signal length of 1 second. The corresponding CNN architecture is shown in Table II consisting of 4 Convolutional Layers (CL), 2 Pooling Layers (PL) and 2 Fully Connected layers (FC). The Rectified Linear Unit (ReLU) activation function is used for all layers, except for the output layer where the softmax activation function is used with the output size equal to the number of classes beeing classified.

TABLE II. Architecture of 1-D CNN for 16 kHz Sampling Rate and Audio Length of 1 Second [27]

	CL1	**PL1**	**CL2**	**PL2**	**CL3**	**CL4**	**FC1**	**FC2**
Dimension	7969	996	483	60	23	8	128	64
Filters count	16	16	32	32	64	128	-	-
Filters size	64	8	32	8	16	8	-	-
Stride size	2	8	2	8	2	2	-	-

3) Independent Channel Residual Convolutional Network: We propose the Independent Channel Residual Convolutional Network (ICRCN), where the input RGB image is divided to the 2-D matrices in individual channels. The feature matrices share the dimension of the x-axis (time) but not the y-axis (frequency, mel banks, time). The system that combines different visual representations may suffer, when the features are combined as one input image. Therefore, we build the residual convolutional network, that processes different audio visualizations separately. The whole model architecture is shown in Table III.

The model input is a three channel RGB image. The separate channels contain residual blocks, where the number of filters is 32. The feature dimensions merging is made after the second residual block. From this point, the features are processed as in standard residual convolutional networks. The last convolutional block consists of 512 filters and it is followed by the classification layer. The proposed architecture is built up from standard residual blocks, as described in III-D1. The proposed architecture is compared to standard residual networks ResNet50 and 1-D CNN in Table IV.

TABLE III. Proposed Independent Channel Residual Convolutional Network

Output size	**ICRCN blocks**		
(224, 224, 3)	Input RGB image		
(112, 112, 32)	7×7, 32	7×7, 32	7×7, 32
(56, 56, 32)	3x3, 32 3x3, 32 ×2	3x3, 32 3x3, 32 ×2	3x3, 32 3x3, 32 ×2
(28, 28, 64)	3x3, 64 3x3, 64 ×2	3x3, 64 3x3, 64 ×2	3x3, 64 3x3, 64 ×2
(28, 28, 192)	Concatenation		
(14, 14, 256)	3x3, 256 3x3, 256 ×2		
(7, 7, 512)	3x3, 512 3x3, 512 ×2		
(512)	Global average pooling		
(2)	Dense 2 + softmax		

TABLE IV. Comparison of Architectures Complexity

Architecture	**Number of trainable parameters**
1-D CNN	256k
ResNet50	23.5M
ICRCN	11M

E. Datasets

One of the most widely known datasets for environmental sound classification is the UrbanSound8K [23]. In this work we build our own gunshot detection dataset as a combination of the UrbanSound8k and The Free Firearm Sound Library [36].

- **UrbanSound8k** - 8732 tracks of 10 classes (air conditioner, car horn, children playing, dog barking, drilling, engine idling, gunshot, jackhammer, siren, street music), with varying sampling frequency file to file.
- **The Free Firearm Sound Library** - 2200 tracks of gunshots in a noise free environment including handguns (pistols, revolvers, semi-automatic pistols), rifles (lever-action, semi-automatic, fully automatic, machine guns, etc.) and shotguns. Recordings are in lossless wav format with sampling frequency 44.1 kHz.

The examples from The Free Firearm Sound Library were combined with gunshots from the UrbanSound dataset. The other sounds from UrbanSound were used as negative examples (random background). Since UrbanSound records are not equal in length, the window size of our examples floats in the range from 1 s to 4 s.

In the field of supervised machine learning, the performance of algorithms strongly depends on the quality of the dataset. To some extent, it is possible to simulate a big dataset using augmentation methods. However, this approach will never be as good as the expansion of the dataset by real examples. The following augmentation techniques are applied to positive examples (gunshots) while training.

- **Random background mixing** - The gunshot were randomly mixed with background noise in a random SNR from 0 dB to 20 dB.
- **Random time shift** - The onset positions of gunshots within the processing window were chosen randomly from 0 to 0.8 % of the window length.
- **Random Gaussian noise addition** - SNR from 60 dB to 100 dB.

F. Evaluation Metrics

The softmax output activation function directly indicates the probability of the example belonging to a certain class. The softmax activation also guarantees that the sum of probabilities over classes (background and gunshot) is one. The examples are classified as positive or negative according to higher or lower activation of output neurons. The confusion matrix can be constructed using predicted and true labels. Thereafter, the performance is evaluated via known metrics.

For practical implementation of the system for dangerous sound detection (gunshots, explosion, ...) it is necessary to minimize the amount of false-positive predictions. In this case, the high accuracy value may be a little bit confusing and thus we have to choose a metric for relevant evaluation. Therefore, in our work we use the following evaluation metrics:

- **Accuracy**: It reflects the overall algorithm performance, so it also takes into account the true negative predictions (TN). The accuracy is high when the number of true positive (TP) and true negative predictions is large. False positives (FP) and false negatives (FN) are prediction errors.

$$Accuracy = \frac{TP + TN}{TN + TP + FN + FP} \tag{4}$$

- **Sensitivity**: The relative amount of true positive predictions against all positive examples. It is often called *recall* or True Positive Rate (TPR).

$$Sensitivity = \frac{TP}{TP + FN} \tag{5}$$

- **Specificity**: The relative amount of true negative predictions against all negative examples.

$$Specificity = \frac{TN}{FP + TN} \tag{6}$$

- **False-Positive Rate (FPR)**: Reflects the relative amount of false positives against all negative examples.

$$FPR = \frac{FP}{FP + TN} \tag{7}$$

- **F1 score**: The balance value between sensitivity and precision, where precision is the amount of true positives against all positive predictions. This means that the F1 score is not distorted by a large number of true negative predictions and may be considered as decisive, when testing data are unbalanced.

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{8}$$

- **Detection Error Tradeoff** - DET curve visualizes the false positive rate vs. the false negative rate. Standardly, the axes are scaled non-linearly. Unlike the ROC curve, the DET curves are more linear and situated in most of the plot area.
- **Equal Error Rate** - EER is defined as the point in DET or ROC where the errors are equal. The lower EER value reflects better performance of the system.

IV. RESULTS

All the networks were trained as long as validation loss was decreasing. Early stop was applied and only the best model was saved. Adam optimizer was used while training the models and the categorical cross-entropy loss was computed in each batch of 16 examples. Hardware used is NVIDIA GeForce GTX 1650, Intel Core i9-10900X CPU 3.7 GHz, RAM 64 GB. Fig. 5 shows the training and validation history of our ICRCN model when using 16 kHz sampling frequency.

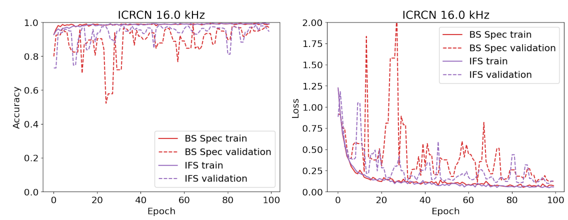


Fig. 5. Training and Validation History of ICRCN Model.

We use Tensorflow and Keras for building and training the models. For augmenting the audio we use the Audiomentations library. The Sci-kit learn is used for evaluation using described metrics.

A. 2-D Convolutional Networks

The training audio examples were generated in three sampling frequencies to test the effect of sampling on system performance. The overall F1 score and FPR is evaluated in Fig. 6. As the test subset is balanced, the F1 score and overall accuracy metrics are very similar. As seen, the high F1 score value does not necessarily guarantee good performance in the reduction of false-positives.

When looking at the F1 score of individual models, the difference between 44.1 kHz and 16 kHz is not significant. In several cases, the score drops significantly at the 8 kHz sampling. The significant drop of F1 score is seen in case of ICRCN trained on band-split spectrograms. Table V, Table VI, and Table VII show also the change in sensitivity and specificity over all combinations.

TABLE V. TESTING RESULTS FOR SAMPLING FREQUENCY 8 KHZ

Model	Features	F1 [%]	Sensitivity [%]	Specificity [%]
ICRCN	IFS	93.1	94.8	91.3
	BS Spec	85.1	74.0	100.0
ResNet50	IFS	95.3	98.4	91.9
	LM Spec	96.0	93.2	99.0
	BS Spec	**96.9**	97.1	96.7
	MFCC	84.8	91.5	75.8
	SSM	96.0	94.2	98.1

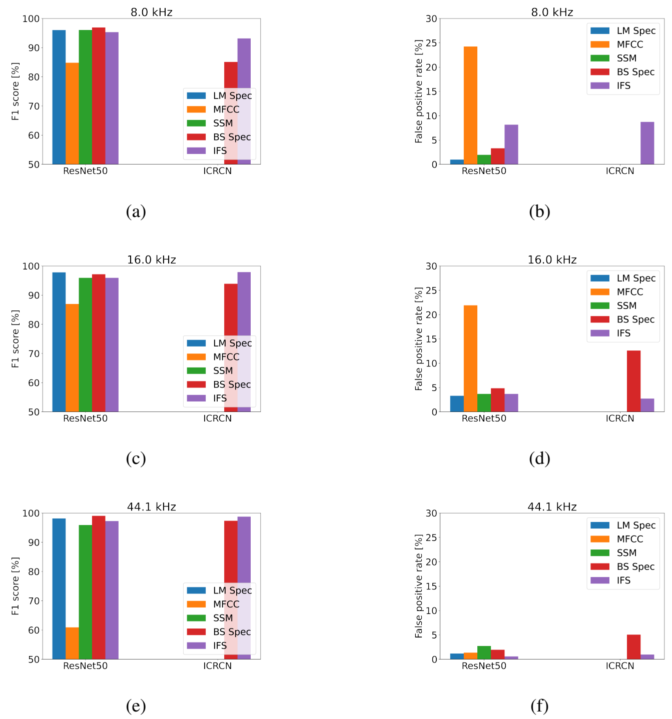


Fig. 6. Results for Different Sampling Frequencies. (a) F1 Score, 8 kHz. (b) FPR, 8 kHz. (c) F1 Score, 16 kHz. (d) FPR, 16 kHz. (e) F1 Score, 44.1 kHz. (f) FPR, 44.1 kHz.

TABLE VI. TESTING RESULTS FOR SAMPLING FREQUENCY 16 KHZ

Model	Features	F1 [%]	Sensitivity [%]	Specificity [%]
ICRCN	IFS	**97.9**	98.4	97.3
	BS Spec	93.9	99.6	87.4
ResNet50	IFS	95.9	95.5	96.3
	LM Spec	97.8	98.8	96.7
	BS Spec	97.1	99.0	95.2
	MFCC	87.0	93.8	78.1
	SSM	95.9	95.5	96.3

TABLE VII. TESTING RESULTS FOR SAMPLING FREQUENCY 44.1 KHZ

Model	Features	F1 [%]	Sensitivity [%]	Specificity [%]
ICRCN	IFS	98.7	98.4	99.0
	BS Spec	97.3	99.6	95.0
ResNet50	IFS	97.2	95.2	99.4
	LM Spec	98.1	97.5	98.8
	BS Spec	**99.0**	100.0	98.1
	MFCC	60.9	44.4	98.6
	SSM	95.9	94.6	97.3

As the softmax output activation function is used, the output gives the probabilities of the example belonging to a certain class. Neverthelss, the operating point can be moved by changing the threshold of positive predictions. Choosing the operating point is strongly related to the application of the developed system and there are different ways how to set up the operating point. The useful visualizations are the DET curves (Fig. 7), which tend to be more linear and highlight the differences in the operating region more clearly than the ROC curves. The EER value represents the operating point, where the system performs equal in terms of the false-positive rate and false-negative rate. The false-positive rate can be interpreted as the false alarm probability and false-negative rate is the probability of missing the positive detection.

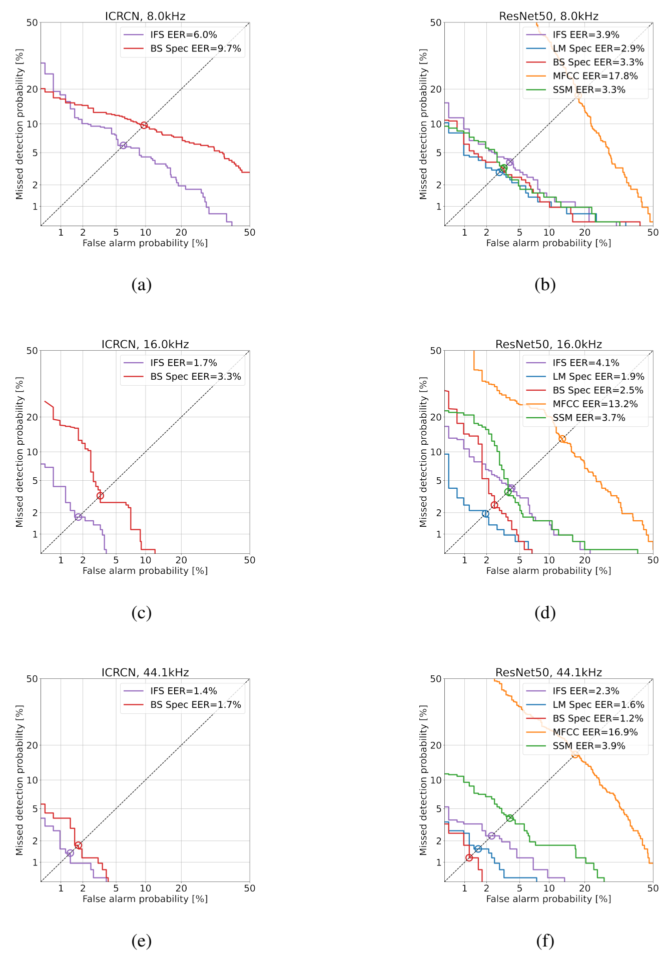


Fig. 7. Detection Error Tradeoff Curves for Different Models and Sampling Frequencies. (a) ICRCN, 8 kHz. (b) ResNet50, 8 kHz. (c) ICRCN, 16 kHz. (d) ResNet50, 16 kHz. (e) ICRCN, 44.1 kHz. (f) ResNet50, 44.1 kHz.

As seen in Table 7, the EER evaluation slightly correlates with the overall F1 score of the system. Though there is no evident regress over the parameters, we can highlight several findings from the results. Using the original sampling frequency when the data are clear and detailed, all architectures model the data well. The more trainable parameters seem to be useful as the system losses the information by audio downsampling. At 16 kHz sampling, the performance of architectures is closely comparable. The proposed ICRCN performs best with IFS feature combination. According to study [32], the single SSM performs worst on environmental sounds classification. Our results showed, that for gunshot detection, using the SSM only may leads to better scores than single MFCC, which surprisingly performs worst.

B. 1-D Convolutional Network

In this part of work, we compare the 1-D CNN model to the best 2-D models. For this comparison we consider only the 16 kHz sampling frequency. As the best candidates from the previous test we choose two model-features combinations, namely ICRCN-IFS and ResNet50-LM Spec. The comparison is shown in Table VIII. The 1-D CNN model was trained on the same dataset as 2-D models.

TABLE VIII. The Results for 1-D CNN Model in Comparison to the Best 2-D Models

Model	Features	F1 [%]	Sensitivity [%]	Specificity [%]
ICRCN	IFS	97.9	98.4	97.3
ResNet50	LM Spec	97.8	98.8	96.7
1-D CNN	-	98.1	98.2	98.1

C. Evaluation of Performance in the Wild

The results on artificially generated test subsets may not always meet the real performance in the wild. Therefore, we evaluate the models on a real audio track from action movie to simulate the real world conditions. On this stage we use only 16 kHz sampling frequency as the most suitable based on previous results. The audio track from the movie John Wick (2014) was manually annotated. We use a 4 seconds window length with 2 seconds overlap and third-order median filtering of the CNN output. Each segment is labeled as positive if there is an occurrence of a gunshot. The numbers of positive and negative segments is very unbalanced, therefore we use the F1 score as the evaluation metric. The evaluation results are shown in Table IX.

TABLE IX. Evaluation on Real Data in the Wild

Model	Features	Test	Evaluation		
		F1[%]	F1[%]	Sensitivity[%]	Specificity[%]
ICRCN	BS Spec	93.9	14.5	98.2	10.5
	IFS	97.9	**44.3**	59.2	91.6
ResNet50	LM Spec	97.8	26.4	89.0	62.5
	MFCC	87.0	16.5	47.2	67.0
	SSM	95.9	5.5	5.5	92.7
	BS Spec	97.1	17.6	98.2	29.2
	IFS	95.9	16.2	95.4	24.3
1-D CNN	-	98.1	4.2	7.8	96.5

As shown in Table IX, the performance on real data drops significantly. However, the evaluation test scores correlate in a relative way. The test score difference is slight in most cases, while the evaluation score difference is noticeable. We can see, that single SSM outperforms the MFCC features in test results, but fails in evaluation, where the sensitivity and specificity are very unbalanced. The results show, that the proposed ICRCN network architecture was able to model the in a data robust way and outperforms the standard residual network ResNet50 even using less trainable parameters. In the case of 1-D CNN, the bad final score was probably caused by the low number of parameters of the convolutional neural network. Thus the network was not able to generalize the extracted features sufficiently to deal with the high variability of real audio data.

The previous experiment mainly focused on the ability of the proposed algorithm to detect gunshot sounds in the simulated real recording. However, in the case of surveillance systems, in addition to accuracy, a low frequency of false alarms is also required. This value cannot be derived from the above experiment because the soundtracks of the movies are sound-exposed. Thus, in this case, recordings from the QUT-NOISE database [37], which is the only suitable publicly available database containing continuous recordings of city sounds, were used to determine the false error rate. Since the total recording time is only a few hours, this database was supplemented with custom recordings from a busy city street. Within these recordings, the proposed algorithm did not detect a single false gunshot event.

D. Processing Time

In Table X, we showed the mean processing time for each of the features used in the experiments. The generation of three feature matrices takes twice as much time than generation of a single spectrogram. This must be taken into account if the detector is implemented on the system with limited processing capacity.

TABLE X. Feature Extraction Time

Sampling frequency [kHz]	Features	Time [ms]
8	Spec	5.02
	MFCC	2.68
	SSM	2.31
16	Spec	6.23
	MFCC	3.62
	SSM	3.71
44.1	Spec	10.53
	MFCC	6.9
	SSM	10.58

When comparing processing time between multiple sampling frequencies, the good compromise is to prefer the 16 kHz sampling. It offers the comparable performance to original sampling while the feature extraction time is almost halved.

V. Discussion

When comparing the residual architectures and features using testing data, there is not a significant best setup. The evaluation on real data showed the benefit of independent features processing, where the difference in score is observable. The standard ResNet50 performs well with a single spectrogram. The feature combination works best with the proposed architecture with independent channels. Therefore, we prefer using the proposed ICRCN architecture and IFS for gunshot detection. Taking the feature extraction times into account, we prefer using the 16 kHz sampling.

We assume, that the ICRCN model fed by IFS matrices is the most suitable for gunshot detection systems, outperforming the other state-of-the-art approaches (in terms of F1 score). The combination of spectrogram, MFCC, and SSM was used in study [32], with conclusion that it does not improve the accuracy of environmental sounds classification. We assume, that the feature combination can be beneficial for gunshot detection applications and has a potential to lower the false-positive rate. The results showed, that there is a benefit of splitting the convolutional channels. The real performance of the system relates on the operating point. In some applications the specificity may be a more important score than sensitivity, or vice versa. We visualized the DET curves, as they can help to fit the requirements of the system for particular applications.

VI. Conclusion

In this paper, we analyzed the usage of several convolutional architectures for the gunshot detection task. We proposed a new architecture and compared its performance to standard ResNet50 and 1-D architectures. The effect of different features on the resulting performance was tested using several Time-Frequency audio representations. For training, validation and testing we collected the gunshots and random backgrounds audio clips from public datasets. In the training phase, the standard augmentation methods were used. Finally, we simulated the real world conditions by evaluation of a real audio track from the action movie.

We achieved the goal of our work by proposing a system for gunshot audio events detection. Our ICRCN approach is able to operate in noisy environments with high specificity (slightly over 90 %), by maintaining fair sensitivity (almost 60 %). The Detection Error Tradeoff analysis showed, that the real performance of the system strongly depends on the error tolerance and requirements. The operating point should be selected for specific applications. On our test data, the missed detection probability is about 10 % when the false alarm probability is as minimal as possible. We expect the similar results also for explosion detection. Such a systems can be implemented in a complex application together with a smoke or fire detector. Due to the fact that the proposed approach has a relatively low computational time, it can be easily integrated into existing surveillance systems without the need to invest in expensive computing servers to operate in real time.

Acknowledgment

This work was supported by the Slovak Grant Agency KEGA under contract no. KEGA 008ZU-4/2021.

References

[1] J. P. Bello, C. Silva, O. Nov, R. L. DuBois, A. Arora, J. Salamon, C. Mydlarz, and H. Doraiswamy, "SONYC: A System for the Monitoring, Analysis and Mitigation of Urban Noise Pollution," *arXiv:1805.00889 [cs, eess]*, May 2018, arXiv: 1805.00889.

[2] S. C. Tan and A. Abd Manaf, "Characterization of internet of things (iot) powered-acoustics sensor for indoor surveillance sound classification," in *2021 IEEE International Conference on Sensors and Nanotechnology (SENNANO)*. IEEE, 2021, pp. 109–112.

[3] F. Graf and M. Gruber, "Rapid Incident Detection in Tunnels through Acoustic Monitoring – Operating Experiences in Austrian Road Tunnels," Graz, Jan. 2018, p. 8.

[4] C. Watkins, L. Green Mazerolle, D. Rogan, and J. Frank, "Technological approaches to controlling random gunfire: Results of a gunshot detection system field test," *Policing: An International Journal of Police Strategies & Management*, vol. 25, no. 2, pp. 345–370, Jun. 2002.

[5] G. Valenzise, L. Gerosa, M. Tagliasacchi, F. Antonacci, and A. Sarti, "Scream and gunshot detection and localization for audio-surveillance systems," Oct. 2007, pp. 21–26.

[6] M. Sigmund and M. Hrabina, "Efficient feature set developed for acoustic gunshot detection in open space," *Elektronika ir Elektrotechnika*, vol. 27, no. 4, pp. 62–68, 2021.

[7] H. Luo, J. Liu, W. Fang, P. E. Love, Q. Yu, and Z. Lu, "Real-time smart video surveillance to manage safety: A case study of a transport mega-project," *Advanced Engineering Informatics*, vol. 45, p. 101100, 2020.

[8] N. Khalid, M. Gochoo, A. Jalal, and K. Kim, "Modeling two-person segmentation and locomotion for stereoscopic action identification: A sustainable video surveillance system," *Sustainability*, vol. 13, no. 2, 2021.

[9] P. K.-Y. Wong, H. Luo, M. Wang, P. H. Leung, and J. C. Cheng, "Recognition of pedestrian trajectories and attributes with computer vision and deep learning techniques," *Advanced Engineering Informatics*, vol. 49, p. 101356, 2021.

[10] W. Renda and C. H. Zhang, "Comparative analysis of firearm discharge recorded by gunshot detection technology and calls for service in louisville, kentucky," *ISPRS International Journal of Geo-Information*, vol. 8, no. 6, p. 275, 2019.

[11] J. H. Ratcliffe, M. Lattanzio, G. Kikuchi, and K. Thomas, "A partially randomized field experiment on the effect of an acoustic gunshot detection system on police incident reports," *Journal of Experimental Criminology*, vol. 15, no. 1, pp. 67–76, 2019.

[12] J. Salamon and J. P. Bello, "Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification," *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 279–283, Mar. 2017, arXiv: 1608.04363.

[13] K. J. Piczak, "Environmental sound classification with convolutional neural networks," in *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*. Boston, MA, USA: IEEE, Sep. 2015, pp. 1–6.

[14] A. Guzhov, F. Raue, J. Hees, and A. Dengel, "ESResNet: Environmental Sound Classification Based on Visual Domain Models," *arXiv:2004.07301 [cs, eess]*, Apr. 2020, arXiv: 2004.07301.

[15] T. Ahmed, M. Uppal, and A. Muhammad, "Improving efficiency and reliability of gunshot detection systems," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 513–517.

[16] M. Hrabina and M. Sigmund, "Gunshot recognition using low level features in the time domain," in *2018 28th International Conference Radioelektronika (RADIOELEKTRONIKA)*. IEEE, 2018, pp. 1–5.

[17] M. Hrabina, "Analysis of linear predictive coefficients for gunshot detection based on neural networks," in *2017 IEEE 26th International Symposium on Industrial Electronics (ISIE)*. IEEE, 2017, pp. 1961–1965.

[18] M. Lojka, M. Pleva, E. Kiktová, J. Juhár, and A. Čižmár, "Efficient acoustic detector of gunshots and glass breaking," *Multimedia Tools and Applications*, vol. 75, no. 17, pp. 10 441–10 469, 2016.

[19] A. A. Shiekh, M. Tahir, and M. Uppal, "Accurate gunshot detection in urban environments using blind deconvolution," in *2017 International Multi-topic Conference (INMIC)*. IEEE, 2017, pp. 1–4.

[20] A. Suliman, B. Omarov, and Z. Dosbayev, "Detection of impulsive sounds in stream of audio signals," in *2020 8th International Conference on Information Technology and Multimedia (ICIMU)*. IEEE, 2020, pp. 283–287.

[21] S. Raponi, I. Ali, and G. Oligeri, "Sound of guns: digital forensics of gun audio samples meets artificial intelligence," *arXiv preprint arXiv:2004.07948*, 2020.

[22] J. Bajzik, J. Prinosil, and D. Koniar, "Gunshot detection using convolutional neural networks," in *2020 24th International Conference Electronics*. IEEE, 2020, pp. 1–5.

[23] J. Salamon, C. Jacoby, and J. P. Bello, "A Dataset and Taxonomy for Urban Sound Research," in *Proceedings of the ACM International Conference on Multimedia - MM '14*. Orlando, Florida, USA: ACM Press, 2014, pp. 1041–1044.

[24] Y. Tokozume and T. Harada, "Learning environmental sounds with end-to-end convolutional neural network," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. New Orleans, LA: IEEE, Mar. 2017, pp. 2721–2725.

[25] Y. Tokozume, Y. Ushiku, and T. Harada, "Learning from Between-class Examples for Deep Sound Recognition," *arXiv:1711.10282 [cs, eess, stat]*, Feb. 2018, arXiv: 1711.10282.

[26] B. Zhu, K. Xu, D. Wang, L. Zhang, B. Li, and Y. Peng, "Environmental Sound Classification Based on Multi-temporal Resolution Convolutional Neural Network Combining with Multi-level Features," *arXiv:1805.09752 [cs, eess]*, Jun. 2018, arXiv: 1805.09752.

[27] S. Abdoli, P. Cardinal, and A. L. Koerich, "End-to-End Environmental Sound Classification using a 1D Convolutional Neural Network," *arXiv:1904.08990 [cs, stat]*, Apr. 2019, arXiv: 1904.08990.

[28] A. G. Katsiamis, E. M. Drakakis, and R. F. Lyon, "Practical gammatone-like filters for auditory processing," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2007, pp. 1–15, 2007.

[29] H. Park and C. D. Yoo, "Cnn-based learnable gammatone filterbank and equal-loudness normalization for environmental sound classification," *IEEE Signal Processing Letters*, vol. 27, pp. 411–415, 2020.

[30] I. Papadimitriou, A. Vafeiadis, A. Lalas, K. Votis, and D. Tzovaras, "Audio-based event detection at different snr settings using two-dimensional spectrogram magnitude representations," *Electronics*, vol. 9, no. 10, p. 1593, 2020.

[31] N. Takahashi, M. Gygli, B. Pfister, and L. Van Gool, "Deep Convolutional Neural Networks and Data Augmentation for Acoustic Event Detection," *arXiv:1604.07160 [cs]*, Dec. 2016, arXiv: 1604.07160.

[32] V. Boddapati, A. Petef, J. Rasmusson, and L. Lundberg, "Classifying environmental sounds using image recognition networks," *Procedia Computer Science*, vol. 112, pp. 2048–2056, 2017.

[33] Z. Zhang, S. Xu, S. Cao, and S. Zhang, "Deep Convolutional Neural Network with Mixup for Environmental Sound Classification," *arXiv:1808.08405 [cs, eess]*, Aug. 2018, arXiv: 1808.08405.

[34] H. Wang, Y. Zou, D. Chong, and W. Wang, "Environmental Sound Classification with Parallel Temporal-spectral Attention," *arXiv:1912.06808 [cs, eess]*, May 2020, arXiv: 1912.06808.

[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *arXiv:1512.03385 [cs]*, Dec. 2015, arXiv: 1512.03385.

[36] "Sound Effects Library," Mar. 2013.

[37] D. Dean, S. Sridharan, R. Vogt, and M. Mason, "Gthe qut-noise-timit corpus for evaluation of voice activity detection algorithms," in *2010 11th Annual Conference of the International Speech Communication Association*, 2010, pp. 3110–3113.

Improving Intrusion Detection for Imbalanced Network Traffic using Generative Deep Learning

Amani A. Alqarni
Department of Computer Science and Engineering
College of Computer Science and Engineering
University of Hafr Al Batin
Hafr Al Batin 39524, Saudi Arabia

El-Sayed M. El-Alfy
Information and Computer Science Department
Interdisciplinary Research Center for Intelligent Secure Systems
College of Computing and Mathematics
King Fahd University of Petroleum and Minerals
Dhahran 31261, Saudi Arabia

***Abstract*—Network security has become a serious issue since networks are vulnerable and subject to increasing intrusive activities. Therefore, network intrusion detection systems (IDSs) are an essential component to defend against these activities. One of the biggest issues encountered by IDSs is the class imbalance problem which leads to a biased performance by most machine learning models to normal activities (majority class). Several techniques were proposed to overcome the class-imbalance problem such as resampling, cost-sensitive, and ensemble learning techniques. Other issues related to intrusion detection data include mixed data types, and non-Gaussian and multimodal distributions. In this study, we employed a conditional tabular generative adversarial network (CTGAN) model with common machine learning algorithms to construct more effective detection systems while addressing the imbalance issue. CTGAN can generate samples of the minority class during training to make the dataset more balanced. To assess the effectiveness of the proposed IDS, we combined CTGAN with three machine learning algorithms: support vector machine (SVM), K-nearest neighbor (KNN), and decision tree (DT). The imbalanced NSL-KDD dataset was used and several experiments were conducted. The results showed that CTGAN can improve the performance of imbalance learning for intrusion detection with SVM and DT. On the other hand, KNN showed no improvement in the performance since it is less sensitive to the class imbalance problem. Moreover, the results proved that CTGAN can capture the distribution of discrete features better than continuous features.**

***Keywords*—Intrusion detection; machine learning; imbalance learning; conditional tabular generative adversarial networks**

I. Introduction

In the current era of transformation to digital services and with the evolution of Internet technologies, cybersecurity has become a serious issue, especially with the growing volume and diversity of attacks. People utilize the Internet to conduct most of their work (e.g. online shopping, payment, banking, access to governmental services, file sharing, communication, and more). Moreover, it is a vital part of several cyber-physical systems in critical infrastructures such as Internet-of-Things (IoT) and smart grids. Therefore, intrusion detection systems are an essential component of cybersecurity and provide crucial services to protect information systems from cyberattacks that can lead to catastrophic consequences (e.g. sensitive data leakage, physical harm, and financial loss). It can monitor the system operations and network traffic to detect anomalous patterns [1].

Intrusion detection is a recurrent research topic due to the emergence of more sophisticated adversarial incidents. One of the challenges facing intrusion detection is the scarcity of intrusive samples compared to the abundance of normal operation samples; resulting in insufficient data samples to train the model on a representative collection of attack scenarios. With such an imbalanced dataset, the number of samples in a majority (normal or negative) class is significantly higher than the number of samples in a minority (abnormal or positive) class. Training a machine learning algorithm on an imbalanced dataset can adversely impact the performance and lead to unsatisfactory results since it can be more biased towards the majority class, i.e. normal patterns.

Imbalanced learning is very common in several problems with rare events and different techniques have been proposed to address the imbalance issue. These techniques can be divided into five main categories: data-level, algorithm-level, cost-sensitive, ensemble-based, and hybrid techniques. The data level are preprocessing techniques known as re-sampling because they either increase the frequency of the minority class (oversampling) or reduce the frequency of the majority class (undersampling/downsampling). Random oversampling (ROS), random undersampling (RUS), synthetic minority oversampling technique (SMOTE), and generative adversarial networks (GAN) are examples of data-level techniques. A systematic literature review of the challenges and solutions for imbalanced data in machine learning is provided in [2].

The focus of our study is on oversampling techniques, specifically, an enhanced version of GAN, i.e. the conditional tabular GAN (CTGAN) [3]. CTGAN is a recent deep learning model and can be thought of as an oversampling technique. It can augment the tabular dataset and increase the frequency of the minority class samples while handling other issues such as mixed data types, multimodality, and non-Gaussian distributions. It proves its efficiency in addressing the imbalance problem and improving the classification accuracy in different domains.

This paper aims to investigate the role of CTGAN in improving the classification performance of support vector machines (SVM), K-nearest neighbors (KNN), and decision trees when applied to imbalanced data to detect various types of network intrusions. Different metrics have been computed to evaluate the quality of the generated data for various attacks in the multi-class NSL-KDD dataset. Moreover, the performance measures of the trained intrusion detection models have been computed and discussed.

The main contributions of this paper are:

- Handling the imbalance problem in intrusion detection datasets by employing CTGAN which is not investigated very well in the literature.
- Evaluating the quality of the generated data by CTGAN in terms of different metrics.

The rest of this paper is organized as follows: Section II provides a brief background about intrusion detection, class imbalance problem, and techniques to deal with this problem. Section III reviews related studies on intrusion detection. Section IV describes the methodology we followed in our study. Finally, Section V describes the experiments we conducted to evaluate the proposed prototype.

II. Background

A. Intrusion Detection

An intrusion detection system is an essential component responsible for analyzing and monitoring networks to detect intrusions and alert administrators on ongoing attack activities [4]. Intrusion detection is still a significant research field for two reasons. First, there are continuous updates and changes of network intrusions resulting in continually changing patterns [5]. Second, the number of available intrusion detection datasets is increasing over time, making it possible to investigate and compare new approaches [5]. Examples of intrusion categories include Denial of Service (DoS) such as smurf, User to Root (U2R) such as buffer overflow, Probing (Prob) such as portsweep, and Root to Local (R2L) such as password guessing [6]. An effective intrusion detection system should have not only low false negative but also low false positive. These measures can be greatly affected by the quality of having a representative training dataset. However, real scenarios may have several challenges. For example, besides the class imbalance problem, the intrusion detection traces may have several other characteristics that need a special treatment, e.g. mixed data types (continuous, discrete, ordinal, and categorical) as well as non-Gaussian and multimodal distributions.

B. Class Imbalance Problem

Most of the intrusion detection datasets are imbalanced datasets, which causes a degradation in the classification performance for certain types of intrusions [7]. Usually, the number of normal traces in intrusion detection datasets is much higher than the number of intrusion traces. Thus, as a minority class, the intrusion class might not be well-represented and hence not classified correctly.

The misclassification of the minority class costs more than the misclassification of the majority class, as it could cause a serious problem. Misclassifying a normal behavior class can lead to the need for more tests to explore the intrusion. On the other hand, misclassifying intrusions may lead to disaster impacts (e.g., privacy loss, unauthorized access to network assets, or damage of the whole system). Even if the detection rate of the minority class is low, classification accuracy could be high because the classification accuracy does not consider the distribution of classes. Consequently, machine-learning based intrusion detection systems can be accuracy biased, since they can give more attention to the majority class (a.k.a. normal behavior).

C. Imbalance Techniques

Many different techniques can be used to address the imbalance issue. These can be categorized into five categories, i.e. data-level, algorithm-level, cost-sensitive, ensemble-based, and hybrid techniques. Data-level techniques modify the class distributions before the training process. This modification is done either by removing some instances from the majority class or by adding more instances to the minority class [8]. The former method is known as undersampling while the latter is known as oversampling. Random oversampling (ROS), random undersampling (RUS) and synthetic minority oversampling technique (SMOTE) are examples of data-level preprocessing techniques.

Unlike data-level techniques, algorithm-level techniques do not modify the distribution of classes; but rather, they modify the algorithm [9]. In contrast, cost-sensitive techniques assign different costs to give the minority class higher importance than the majority class [10]. Ensemble methods combine more than one algorithm to achieve superior performance than would normally be attained separately such as bagging, boosting, stacking, and cascading classifiers [10]. Moreover, hybrid techniques combine two or more of the aforementioned techniques to produce an efficient technique for handling imbalance [9]. For instance, SMOTEENN combines oversampling by SMOTE with undersampling by the edited nearest neighbor (ENN) method [11].

D. Conditional Tabular GAN

Generative Adversarial Network (GAN) is one of the top innovative deep learning models [12]. It has been widely used in various applications to process different types of data (e.g., images, voice, and text). Hence, it became one of the most critical research fields in deep learning. It combines two networks: generator and discriminator [13]. The generator is responsible for generating synthetic data that resemble to the original data whereas the discriminator is responsible for classifying the real or synthetic data with their corresponding classes [13]. GAN in tabular data has various challenges. One of the challenges is that the structured data can follow non-Gaussian and multimodal distributions. Tabular GAN (TGAN) resolved this issue by using mode-specific normalization [3].

In GAN, the generator does not consider the imbalance issue. Data that belong to the minority class will not be presented sufficiently as the data belong to the majority class do. Conditional generator in conditional GAN can be used to enforce the synthetic sample to match a specific class (category) [14]. Hence, it can generate more intrusive samples to overcome the imbalance issue in intrusion detection datasets. To use the conditional generator, in CGAN, instead of the original generator in GAN, three issues must be mitigated:

- A way must be found to represent the condition and the input to the generator.
- The generated samples must match the chosen category (condition).

- The conditional generator should learn the conditional distribution.

Conditional Tabular GAN (CTGAN) combines the advantages of CGAN and TGAN. Therefore, it can be used to solve the class imbalance issue by controlling the class labels of the generated samples. It also overcomes the non-Gaussian and multimodal distributions of structured data. Furthermore, CTGAN utilizes fully connected networks to enhance the quality of the model [3]. The conditional generator can be interpreted as

$$\hat{\mathbf{r}} \sim P_g(row|D_{i*} = k^*)$$

where k^* is the chosen category from the discrete column D_{i*} that must be generated by conditional generator and $\hat{\mathbf{r}}$ is the sample generated by the generator. Fig. 1 shows a typical architecture of the CTGAN.

III. Related Work

With the growth of the amount of data related to intrusion detection, and the evolution in machine learning and deep learning techniques, many studies have been conducted in this field. Most of the previous studies ignore the class imbalance problem and use datasets of balanced distributions. In [1], the authors proposed a framework called scale-hybrid-IDS-AlertNet which can trace the network traffic and detect abnormal activities. The building of this framework came after a comprehensive analysis of different machine learning and deep learning models on different intrusion detection datasets. They found that the deep neural network (DNN) model outperforms other machine learning models such as Logistic Regression (LR), Naïve Bayes (NB), K-nearest neighbor (KNN), and support vector machines (SVM).

There are some researchers who have used imbalance techniques to deal with this problem in network intrusion detection systems. For example, Razan Abdulhammed et al. [15] compared the performance of different data-level techniques on the CIDDS-001 dataset. For data preprocessing, they considered ROS, RUS, class balanced and spread subsample. For classification, they utilized deep neural networks (DNN), random forest (RF), voting technique, stacking technique and variational autoencoder (VA). The superior performance was achieved by RF on the original distribution, class balancer and RUS (99.9%). Moreover, the accuracy of voting in the original distribution and using RUS was high (99.99%).

SMOTE is a known and effective oversampling technique, and many studies have been conducted to prove its efficiency. In [16], the imbalance issue was mitigated in a CICIDS2017 dataset using SMOTE oversampling. As SMOTE works only with binary classification, the researchers examined two classes at a time—a normal class with one of the minority classes (i.e., botnet, web attack, or brute force attack). Two experiments were conducted: one on the imbalanced dataset and one on the balanced dataset. Three algorithms were utilized to conduct the experiments (i.e., RF, NB, and KNN). On the imbalanced dataset, the accuracy was high with all classifiers, but the precision, recall, and F1-score were low. To mitigate this degradation, the researchers applied the SMOTE technique and the result showed better performance based on F1-score and recall.

Generative adversarial networks (GAN) is considered as a data-level technique because it modifies the distribution of data by generating new samples. It was applied in [17] by Yilmaz et al. to improve the performance of intrusion detection. The result after applying GAN was found to be more accurate than without using GAN. Different versions of GAN were proposed to enhance its performance. For example, Shuokang Huang and Kai Lei. [18] proposed Imbalanced GAN (IGAN) that includes a data imbalance filter, a generator, and a discriminator. It can force the generator to generate samples of the minority class only. IGAN with fuzzy neural network (FNN) obtained a superior performance over Convolutional Neural Network (CNN), RUS with SVM, FNN, and SMOTE with multilayer perception (MLP).

Punam Bedi et al. [19] proposed an intrusion detection system called Siam IDS which is based on Siamese Neural Network (Siamese-NN). It can handle the imbalance issue in intrusion detection systems. It achieved high recall values of the minority classes (U2R and R2L intrusions). Moreover, it outperformed CNN-based IDS and DNN-based IDS. In [20], m-RIGFS and RWIGFS were used with weighted-SVM to improve the imbalance learning for the intrusion detection systems. m-RIGFS and RWIGFS are feature selection techniques for imbalanced classes. This approach obtained a good performance in terms of the overall accuracy, sensitivity, and specificity. However, it should be noted that the sensitivity of the U2R, a rare class was low.

In [7], a CNN model was utilized to classify UNSW-NB15 and CICIDS2017, which are relatively recent intrusion detection datasets. They applied the CNN model, after addressing the class imbalance issue using their proposed approach, i.e. SGM which combines an oversampling technique (SMOTE) with an undersampling technique (Gaussian Mixture (GMM)). The proposed approach achieved a higher result (more than 96%) compared to other sampling techniques and classification models. Furthermore, in [21], SMOTE was combined with a genetic fuzzy system that includes a fitness function designed to deal with the imbalance problem. The proposed approach outperformed other approaches, which are the KDDCup-99 winner [22], GFS(Pittsburgh) [23], MOG-FIDS [24], EFRID [25] and RIPPER [26].

Data-level and cost-sensitive techniques can be combined to yield better performance. Alabdallah et al. in [27] combined a stratified sampling with a cost function. This approach assigns the minority class samples higher weights than the majority class samples to improve the classification performance and decrease the accuracy paradox.

To our knowledge and based on exploring earlier work in the literature, limited studies have used GAN-based methods to overcome the imbalance problem in intrusion detection systems. Moreover, none of the reviewed studies have evaluated the generated samples. Therefore, in our study, we explore CTGAN machine-learning based models to deal with the imbalance learning for intrusion detection. Moreover, we evaluated the generated samples in terms of different metrics (e.g. Chi-Squared test and Continuous Kullback–Leibler Divergence).

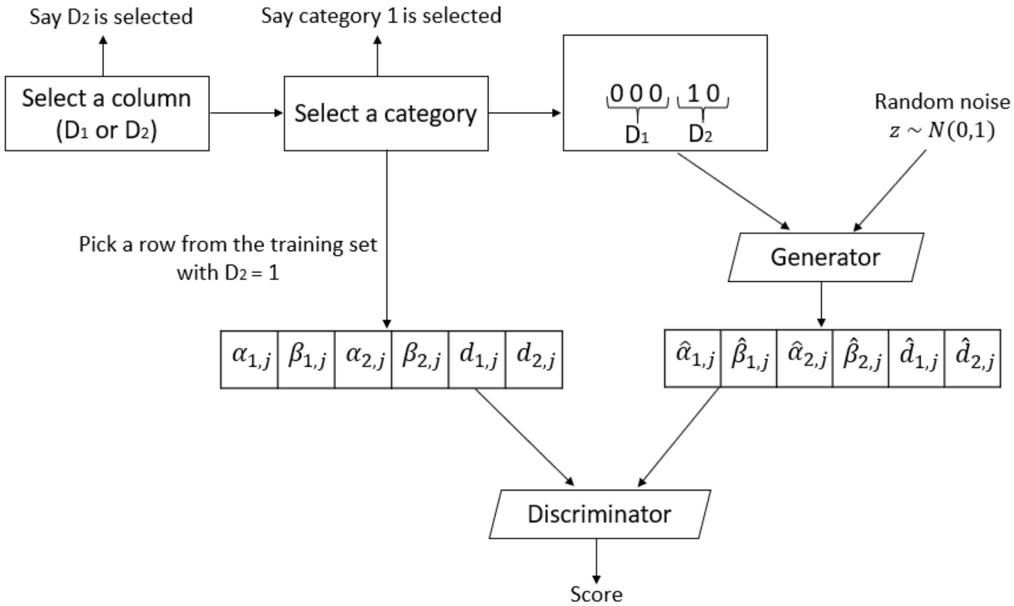


Fig. 1. CTGAN Architecture.

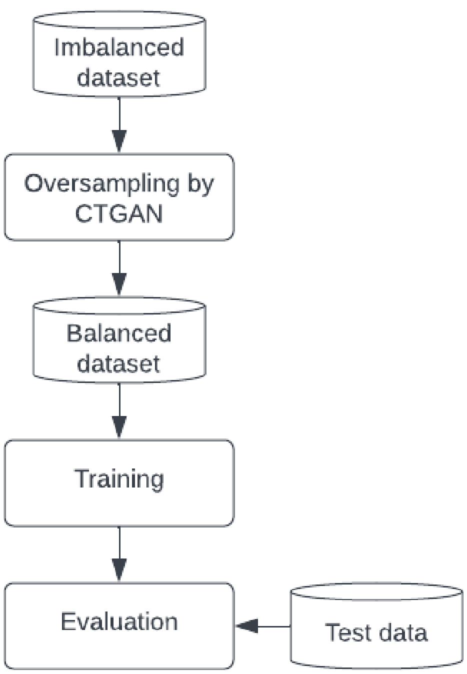


Fig. 2. General Layout of the Workflow.

IV. Methodology

Fig. 2 depicts the general layout of the workflow. The aim of this work is to investigate the performance of CTGAN in improving the detection rate of intrusive samples in intrusion detection systems. To achieve this aim, we applied the following steps:

- Oversampling by CTGAN: Since the intrusion data suffers from several issues including class imbalance, mixed data types, multimodality, and non-Gaussian distributions, we first applied CTGAN to increase the intrusive samples in the training dataset.
- Data preprocessing: The aim of this step is converting raw data into a suitable format for machine learning models. In this study, two preprocessing techniques were applied to the dataset before classification:
  - One-hot encoding: Some machine learning algorithms work only with numeric data, hence one-hot encoding is important when dealing with a dataset that contains categorical features. One-hot encoding is a method of sorting each categorical value into a distinct column and setting a value to it (either 0 or 1). Therefore, one-hot encoding was applied to the NSL-KDD dataset to convert its categorical features (i.e., protocol type, service, and flag).
  - Standard scalar: This is a significant step for some machine learning algorithms because features that are not scaled to have zero mean and unit variance could negatively affect the performance of the algorithm. This method scales the features to have a mean of zero and a standard deviation of one. Hence, all values will be in the same range. The following equation is used to perform standard scalar:

    $$z = \frac{x - \mu}{\sigma}$$

    where x is a sample, μ is the mean of the samples and σ is the standard deviation of the samples.
- Machine learning training and testing: To assess the efficiency of using CTGAN to overcome the imbal-

ance problem and other data issues, three machine learning algorithms (i.e. SVM, DT, and KNN) were used to train and test the NSL-KDD dataset before and after applying CTGAN.

- Evaluation: To accurately evaluate the performance of the proposed framework, various metrics (e.g. F1-score, geometric mean (G-mean), and Matthews correlation coefficient (MCC)) were used. These metrics are appropriate for evaluating the performance of imbalance learning.

V. Experiments and Evaluation

A. Dataset Description

NSL-KDD is an imbalanced dataset with huge amount of captured traffic under various normal and attack scenarios. It is an improved and revised version of the KDD99 dataset. It contains 41 features extracted from traffic traces of normal and abnormal activities as shown in Table I. Intrusion traffic is categorized into four main categories: Denial-of-Service (DoS), Probing (Probe), Remote-to-Local (R2L), and User-to-Root (U2R). Each category of these attacks contains several sub-categories as shown in Table II. In this study, only 20% of the dataset was used to conduct the experiments. Each class includes a different number of samples as shown in Table III. Therefore, this dataset is imbalanced because of the obvious difference in the number of samples in each class. The imbalance ratio of normal and attack classes in the dataset ranges from 1.44 to 305.77. Fig. 3 shows the distribution of intrusive and normal samples.

TABLE I. Features of NSL-KDD Dataset

No.	Feature	No.	Feature	No.	Feature
1	duration	16	num_root	31	srv_diff_host_rate
2	protocol_type	17	num_file_creations	32	dst_host_count
3	service	18	num_shells	33	dst_host_srv_count
4	flag	19	num_access_files	34	dst_host_same_srv_rate
5	src_bytes	20	num_outbound_cmds	35	dst_host_diff_srv_rate
6	dst_bytes	21	is_host_login	36	dst_host_same_src_port_rate
7	land	22	is_guest_login	37	dst_host_srv_diff_host_rate
8	wrong_fragment	23	Count	38	dst_host_serror_rate
9	urgent	24	srv_count	39	dst_host_srv_serror_rate
10	hot	25	serror_rate	40	dst_host_rerror_rate
11	num_failed_logins	26	srv_serror_rate	41	dst_host_srv_rerror_rate
12	logged_in	27	rerror_rate		
13	num_compromised	28	srv_rerror_rate		
14	root_shell	29	same_srv_rate		
15	su_attempted	30	diff_srv_rate		

TABLE II. Attack Categories in NSL-KDD

Attack category	Attack Types
DoS	Back, Apache2, Smurf, Neptune, Udpstorm, Land,Worm, Pod, Processtable, Teardrop
Prob	Satan, Portsweep, Nmap, Ipsweep, Mscan, Saint
R2L	Warezmaster, Phf, Ftp_write, Named, Snmpguess,Httptunnel, Xlock, Spy, Xsnoop, Sendmail, Imap,Guess_Password, Warezclient, Multi-hop, Snmpgetattack
U2R	Loadmodule, Buffer_overflow, Rootkit, Xterm,Sqlattack, Perl, PS

B. Performance Measures

1) Evaluation Metrics of Machine Learning Models: The following metrics were used to evaluate the performance of the applied classifiers:

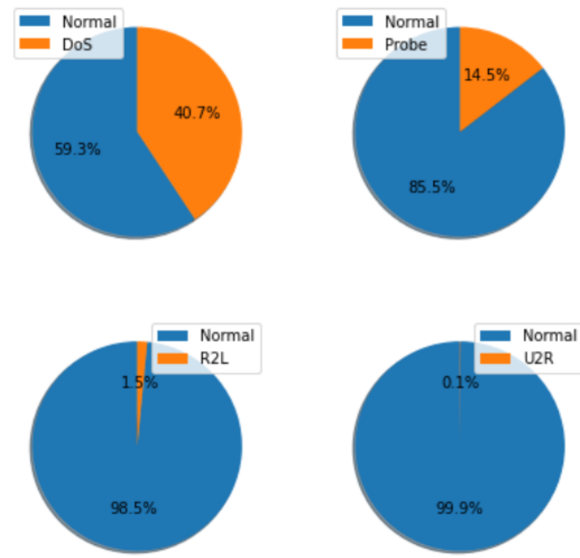


Fig. 3. Distribution of the Samples.

TABLE III. Samples in NSL-KDD

Class	Train set	Test set
Normal	13449	9711
DoS	9234	7458
Prob	2289	2421
R2L	209	2754
U2R	11	200
Total	25192	22544

a) Accuracy (ACC): This is the most common metric to evaluate the performance of a model. It is the number of samples that are correctly predicted over the number of all samples. It can be calculated using the following equation:

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

b) Recall: It refers to the ability of the model to predict positive samples. It can be calculated by dividing the number of the samples that are correctly classified as true positive over all positive samples. It can be calculated using the following equation:

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

c) Precision: It is the number of samples that are correctly classified as true positive over the number of samples that are predicted as positive. The following equation can be used to calculate the precision:

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

d) F1-score: It is a way to combine recall and precision into a single metric. It is called the harmonic mean of recall and precision. It can be calculated using Equation 4:

$$F1\_score = \frac{2\times Precision\times Recall}{Precision+Recall} \tag{4}$$

e) Geometric mean (G-mean): It is based on the true positive rate (sensitivity or recall) and true negative rate (specificity). G-mean is a combination of sensitivity (recall), and specificity. Specificity and G-mean can be calculated using the following equations:

$$Specificity = \frac{TN}{FP+TN} \tag{5}$$

$$G\_mean = \sqrt{Sensitivity \times Specificity} \tag{6}$$

f) Matthews correlation coefficient (MCC): It is a good performance metric for binary classification and combines all parts of the confusion matrix (i.e., true positives, false positives, true negatives, and false negatives). It can be calculated using the following equation:

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP+FP) \times (TP+FN) \times (TN+FP) \times (TN+FN)}} \tag{7}$$

C. Synthetic Data Evaluation Metrics

The overall evaluation score of the synthetic data is an aggregation of the following metrics:

a) Chi-Squared (CSTest): It is a statistical metric that compares the distributions of two discrete columns using the Chi-squared test.

b) Inverted Kolmogorov-Smirnov D statistic (KSTest): It is a statistical metric that compares the distributions of the continuous columns using the Kolmogorov–Smirnov test.

c) KSTestExtended: It is an extension of the KSTest metric that transforms all columns into numerical columns before applying the KSTest.

d) Continuous Kullback–Leibler Divergence (Continuous KLDivergence): This metric calculates the Kullback-Leibler (KL) divergence on all pairs of the numerical columns.

e) Discrete Kullback–Leibler Divergence (Discrete KLDivergence): This metric calculates the Kullback-Leibler divergence on all pairs of the Boolean and categorical columns.

D. Experiments

In this study, one-vs-one classification was performed. We divided the dataset into four subsets. Each subset contains two class labels, i.e. normal and one of the intrusions (i.e., DoS, Prob, R2L, or U2R).

The experiments were conducted in Python and CTGAN was used to oversample the subsets using different number of epochs. Tuning epoch size in deep learning models is important, as it has a direct effect on the performance. It is the number of iterations the model performs over the training dataset. The default epoch size value in CTGAN is 300. In this study, different epoch size values (e.g., 300, 400, and 1600) were tested to obtain the best results.

Then, we evaluated the quality of the synthetic samples using multiple evaluation metrics that are combined to produce an overall score. This score provides a general indication of how good the synthetic samples are. Lastly, SVM, KNN, and decision tree models were run on the NSL-KDD dataset before and after implementing CTGAN.

E. Result and Discussion

Table IV presents the evaluation scores of the synthetic data used to balance the subsets. The overall evaluation score is the average of multiple scores that evaluate the data from different aspects (e.g., statistical, detection, and likelihood). The overall score gives an estimation of how similar the synthetic data and the real data are (i.e., the quality of the generated data). This score ranges from 0 to 1, where 0 is the worst possible score and 1 is the best possible score. As shown in Table IV all synthetic data of all attack categories achieved overall scores around 0.5.

We observed that CSTest obtained high scores ranging from 0.85 to 0.99, whereas KSTest achieved fairly good scores ranging from 0.77 to 0.82. Moreover, KL Divergence scores for discrete columns are reasonable unless for DoS; it is slightly high. On the other hand, KL Divergence scores for Continuous columns are high which indicates worse performance. Therefore, we can conclude that CTGAN can fairly capture distributions of both continuous and discrete columns, but its performance is better in the discrete columns.

TABLE IV. Evaluation Scores of the Synthetic Samples

Metric	DoS	Probe	R2L	U2R
Overall evaluation score	0.54	0.50	0.49	0.51
CSTest	0.99	0.98	0.85	0.92
KSTest	0.80	0.77	0.82	0.80
KSTestExtended	0.79	0.76	0.81	0.80
ContinuousKLDivergence	0.83	0.73	0.80	0.84
DiscreteKLDivergence	0.40	0.29	0.28	0.21

TABLE V. Results of one-vs-one SVM Classification

Imbalanced dataset	Normal-vs-DoS	Normal-vs-Probe	Normal-vs-R2L	Normal-vs-U2R
Accuracy	0.90	0.92	0.79	0.98
Precision	0.99	0.87	0.99	0
Recall	0.77	0.69	0.07	0
F1-score	0.87	0.77	0.13	0
G-mean	0.87	0.81	0.26	0
MCC	0.80	0.72	0.231	-0.002
Balanced dataset	Normal-vs-DoS	Normal-vs-Probe	Normal-vs-R2L	Normal-vs-U2R
Accuracy	0.93	0.96	0.79	0.98
Precision	0.99	0.87	0.60	0.38
Recall	0.85	0.94	0.17	0.24
F1-score	0.91	0.90	0.26	0.30
G-mean	0.91	0.95	0.40	0.49
MCC	0.86	0.87	0.234	0.29

Although accuracy is the most common evaluation metric of machine learning models, it is an inappropriate metric for imbalanced classification because it does not distinguish between the number of correctly classified samples of the majority and minority classes. Therefore, it is obvious from Tables V, VI, and VII that there is a degradation in accuracy values of some attack categories after implementing CTGAN.

Moreover, the precision values of all classifiers decreased for some attacks categories, but improved for others. At the same time, recall values increased for all attack categories, except for DoS in the decision tree and KNN, where there was no improvement.

Precision and recall separately are not enough to evaluate the performance of the imbalanced classification. F1-score is

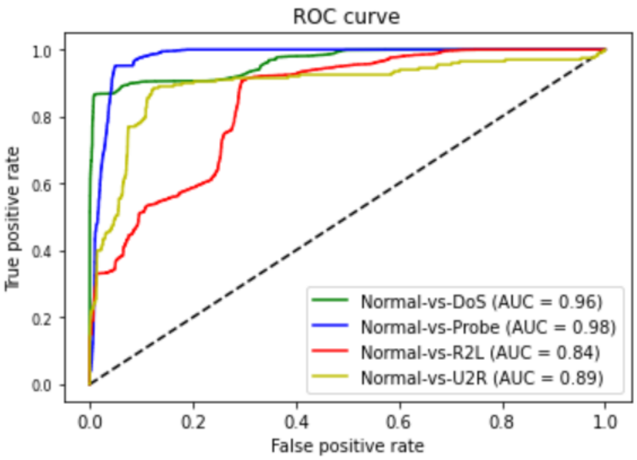


Fig. 4. ROC Curves of SVM on the Imbalanced Dataset.

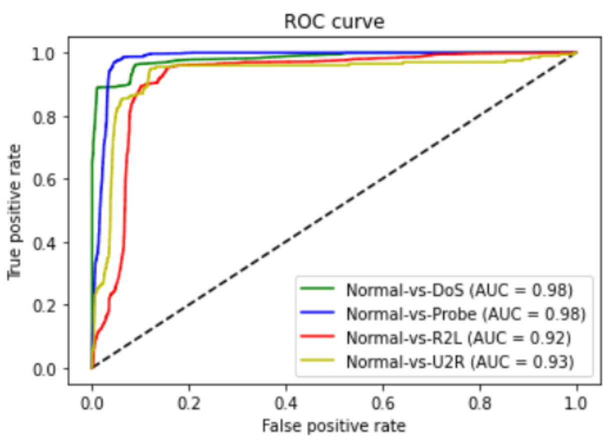


Fig. 5. ROC Curves of SVM on the Balanced Dataset.

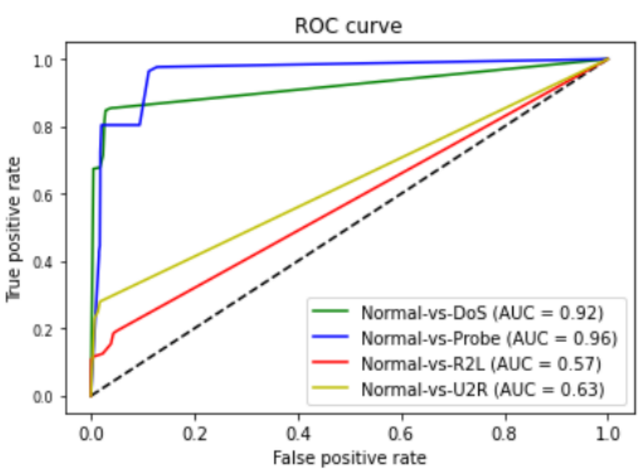


Fig. 6. ROC Curves of DT on the Imbalanced Dataset.

a balance of precision and recall; thus, it is a good metric for the imbalanced classification. Tables V, VI, and VII show an improvement in the F1-score values in all attack categories after using CTGAN, except for DoS with KNN, where there was no improvement.

TABLE VI. RESULTS OF ONE-VS-ONE DT CLASSIFICATION

Imbalanced dataset	Normal-vs-DoS	Normal-vs-Probe	Normal-vs-R2L	Normal-vs-U2R
Accuracy	0.91	0.94	0.80	0.98
Precision	0.96	0.91	0.98	0
Recall	0.82	0.80	0.11	0
F1-score	0.88	0.85	0.20	0
G-mean	0.89	0.88	0.33	0
MCC	0.81	0.821	0.291	0
Balanced dataset	Normal-vs-DoS	Normal-vs-Probe	Normal-vs-R2L	Normal-vs-U2R
Accuracy	0.91	0.96	0.80	0.94
Precision	0.98	0.88	0.68	0.22
Recall	0.82	0.91	0.21	0.73
F1-score	0.89	0.86	0.32	0.33
G-mean	0.90	0.92	0.45	0.83
MCC	0.83	0.828	0.299	0.37

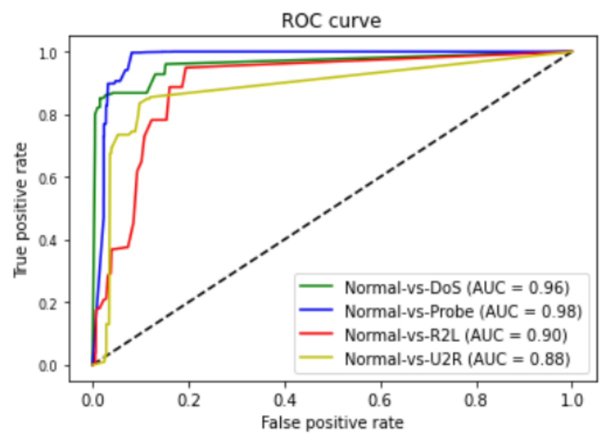


Fig. 7. ROC Curves of DT on the Balanced Dataset.

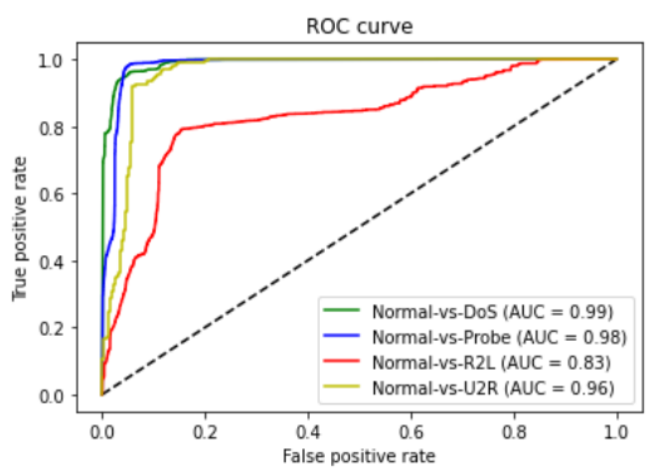


Fig. 8. ROC Curves of KNN on the Imbalanced Dataset.

G-mean is the balance between the accuracy of the algorithm on the majority class and the accuracy of the algorithm on the minority class. Hence, it is an appropriate evaluation metric for imbalanced classification. We notice from Tables V, VI, and VII that there is an improvement in G-mean values of all classifiers on all attack categories after using CTGAN, except for DoS with KNN there was no improvement.

TABLE VII. RESULTS OF ONE-VS-ONE KNN CLASSIFICATION

Imbalanced dataset	Normal-vs-DoS	Normal-vs-Probe	Normal-vs-R2L	Normal-vs-U2R
Accuracy	0.90	0.93	0.79	0.98
Precision	0.99	0.88	0.99	0.75
Recall	0.78	0.77	0.04	0.03
F1-score	0.87	0.82	0.07	0.06
G-mean	0.87	0.86	0.19	0.17
MCC	0.80	0.78	0.17	0.14
Balanced dataset	Normal-vs-DoS	Normal-vs-Probe	Normal-vs-R2L	Normal-vs-U2R
Accuracy	0.90	0.96	0.78	0.97
Precision	0.99	0.86	0.48	0.21
Recall	0.78	0.93	0.09	0.11
F1-score	0.87	0.89	0.15	0.15
G-mean	0.87	0.94	0.29	0.33
MCC	0.80	0.86	0.12	0.14

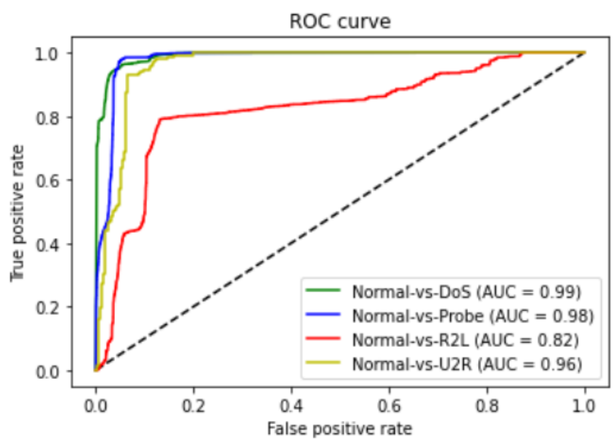


Fig. 9. ROC Curves of KNN on the Balanced Dataset.

While F1-score and G-mean are good evaluation metrics for imbalanced classification, MCC is more informative and reliable because it is determined based on the values of all of the cells of the confusion matrix. It is high only if all values of the confusion matrix are good. Moreover, MCC is the metric least affected by the imbalance issue. Thus, we notice from Tables V and VI that MCC values of SVM and decision tree increased on the balanced datasets for all attacks categories. On the other hand, there was no improvement in MCC values of KNN on DoS and U2R balanced datasets as shown in Table VII. Also, the MCC value of KNN was reduced on the balanced R2L dataset. One possible reason for this issue is that the KNN is less sensitive to the imbalance problem, which is consistent with [28]. Thus, oversampling the dataset did not improve the performance.

ROC curves and AUC scores are commonly used evaluation metrics for imbalanced classification. Fig. 4, Fig. 5 show the ROC curves and AUC scores of SVM on the imbalanced and balanced datasets of all attacks categories. We notice that AUC scores of all attack categories in the balanced datasets are clearly better than the AUC scores in the imbalanced datasets, which indicates an improved performance. Fig. 6 and Fig. 7 demonstrate the ROC curves and AUC scores of decision tree on the imbalanced and balanced datasets of all attacks categories. The AUC scores of all attacks categories are significantly higher after implementing CTGAN, especially for the rare classes (i.e., R2L and U2R). Fig. 8 and Fig. 9 depict the ROC curves and AUC scores of KNN on the imbalanced and balanced datasets of all attacks categories. We observed that the AUC scores of KNN did not show an improvement after the balancing process except with R2L, where there was a slight reduction.

Based on the F1-score, G-mean, MCC and AUC values obtained, using CTGAN to generate synthetic samples has improved the performance of the SVM and decision tree models. At the same time, we notice that the positive impact of CTGAN was not obvious on KNN, instead it causes a slight reduction in some evaluation metrics. This is due to the insensitivity of the KNN to the imbalance problem.

VI. CONCLUSION

Most of the intrusion detection datasets are imbalanced due to the natural difference between the number of intrusive and normal samples. There are many techniques to deal with the imbalance problem. One of these techniques is oversampling (e.g., ROS, SMOTE, and GAN-based methods). In this study, the focus is on CTGAN which can generate more samples of a specific class. CTGAN was applied on NSL-KDD to increase the number of intrusive samples and make the dataset balanced. The effectiveness of CTGAN was evaluated by running SVM, decision tree, and KNN models on the dataset before and after using CTGAN. Experiments using various types of attacks and one-vs-one classification were conducted in this study. CTGAN proved its effectiveness in generating synthetic data resembling real intrusions to improve the performance of SVMs and decision trees on the imbalanced datasets in terms of F1-score, G-mean, MCC, and AUC values. On the other hand, CTGAN did not show an improvement in KNN performance because it is less sensitive to class imbalance problem. For future work, more machine learning and deep learning models could be used. Furthermore, other imbalance techniques could be evaluated and compared with CTGAN.

ACKNOWLEDGMENT

The authors would like to thank King Fahd University and the University of Hafr Al Batin for support during this work.

REFERENCES

[1] R. Vinayakumar, M. Alazab, K. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, “Deep learning approach for intelligent intrusion detection system,” *IEEE Access*, vol. 7, pp. 41 525–41 550, 2019.

[2] H. Kaur, H. S. Pannu, and A. K. Malhi, “A systematic review on imbalanced data challenges in machine learning: Applications and solutions,” *ACM Computing Surveys (CSUR)*, vol. 52, no. 4, pp. 1–36, 2019.

[3] L. Xu, M. Skoularidou, A. Cuesta-Infante, and K. Veeramachaneni, “Modeling tabular data using conditional gan,” in *Advances in Neural Information Processing Systems*, 2019, pp. 7335–7345.

[4] A. Shafee, M. Baza, D. A. Talbert, M. M. Fouda, M. Nabil, and M. Mahmoud, “Mimic learning to generate a shareable network intrusion detection model,” in *2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 2020, pp. 1–6.

[5] J. Lee and K. Park, “Gan-based imbalanced data intrusion detection system,” *Personal and Ubiquitous Computing*, pp. 1–8, 2019.

[6] C. Yin, Y. Zhu, J. Fei, and X. He, “A deep learning approach for intrusion detection using recurrent neural networks,” *IEEE Access*, vol. 5, pp. 21 954–21 961, 2017.

[7] H. Zhang, L. Huang, C. Q. Wu, and Z. Li, "An effective convolutional neural network based on smote and gaussian mixture model for intrusion detection in imbalanced dataset," *Computer Networks*, p. 107315, 2020.

[8] S. E. Gómez, L. Hernández-Callejo, B. C. Martínez, and A. J. Sánchez-Esguevillas, "Exploratory study on class imbalance and solutions for network traffic classification," *Neurocomputing*, vol. 343, pp. 100–119, 2019.

[9] Z. Hosenie, R. Lyon, B. Stappers, A. Mootoovaloo, and V. McBride, "Imbalance learning for variable star classification," *Monthly Notices of the Royal Astronomical Society*, vol. 493, no. 4, pp. 6050–6059, 2020.

[10] Y. Sun, M. Li, L. Li, H. Shao, and Y. Sun, "Cost-sensitive classification for evolving data streams with concept drift and class imbalance," *Computational Intelligence and Neuroscience*, vol. 2021, 2021.

[11] X. Zhang, J. Ran, and J. Mi, "An intrusion detection system based on convolutional neural network for imbalanced network traffic," in *2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT)*. IEEE, 2019, pp. 456–460.

[12] A. Aggarwal, M. Mittal, and G. Battineni, "Generative adversarial network: An overview of theory and applications," *International Journal of Information Management Data Insights*, vol. 1, no. 1, p. 100004, 2021.

[13] S. Park and H. Park, "Performance comparison of multi-class svm with oversampling methods for imbalanced data classification," in *International Conference on Broadband and Wireless Computing, Communication and Applications*. Springer, 2020, pp. 108–119.

[14] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[15] R. Abdulhammed, M. Faezipour, A. Abuzneid, and A. AbuMallouh, "Deep and machine learning approaches for anomaly-based intrusion detection of imbalanced network traffic," *IEEE sensors letters*, vol. 3, no. 1, pp. 1–4, 2018.

[16] A. A. ALFRHAN, R. H. ALHUSAIN, and R. U. Khan, "Smote: Class imbalance problem in intrusion detection system," in *2020 International Conference on Computing and Information Technology (ICCIT-1441)*. IEEE, 2020, pp. 1–5.

[17] I. Yilmaz, R. Masum, and A. Siraj, "Addressing imbalanced data problem with generative adversarial network for intrusion detection," in *2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI)*. IEEE, 2020, pp. 25–30.

[18] S. Huang and K. Lei, "Igan-ids: An imbalanced generative adversarial network towards intrusion detection system in ad-hoc networks," *Ad Hoc Networks*, vol. 105, p. 102177, 2020.

[19] P. Bedi, N. Gupta, and V. Jindal, "Siam-ids: Handling class imbalance problem in intrusion detection systems using siamese neural network," *Procedia Computer Science*, vol. 171, pp. 780–789, 2020.

[20] B. Setiawan, S. Djanali, and T. Ahmad, "Analyzing the performance of intrusion detection model using weighted one-against-one support vector machine and feature selection for imbalanced classes," *Int. J. Intell. Eng. Syst*, vol. 13, pp. 151–160, 2020.

[21] S. M. Gaffer, M. E. Yahia, and K. Ragab, "Genetic fuzzy system for intrusion detection: Analysis of improving of multiclass classification accuracy using kddcup-99 imbalance dataset," in *2012 12th International Conference on Hybrid Intelligent Systems (HIS)*. IEEE, 2012, pp. 318–323.

[22] C. Elkan, "Results of the kdd'99 classifier learning," *ACM SIGKDD Explorations Newsletter*, vol. 1, no. 2, pp. 63–64, 2000.

[23] M.-Y. Su, C.-Y. Lin, S.-W. Chien, and H.-C. Hsu, "Genetic-fuzzy association rules for network intrusion detection systems," in *Proc. IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011)*, 2011, pp. 2046–2052.

[24] C.-H. Tsang, S. Kwong, and H. Wang, "Genetic-fuzzy rule mining approach and evaluation of feature selection techniques for anomaly intrusion detection," *Pattern Recognition*, vol. 40, no. 9, pp. 2373–2391, 2007.

[25] J. Gomez and D. Dasgupta, "Evolving fuzzy classifiers for intrusion detection," in *Proc. IEEE Workshop on Information Assurance*, vol. 6, no. 3, 2002, pp. 321–323.

[26] R. Agarwal and M. V. Joshi, "Pnrule: A new framework for learning classifier models in data mining (a case-study in network intrusion detection)," in *Proc SIAM International Conference on Data Mining*. SIAM, 2001, pp. 1–17.

[27] A. Alabdallah and M. Awad, "Using weighted support vector machine to address the imbalanced classes problem of intrusion detection system," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 12, no. 10, pp. 5143–5158, 2018.

[28] P. Thanh Noi and M. Kappas, "Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery," *Sensors*, vol. 18, no. 1, p. 18, 2017.

A Machine Learning Model for the Diagnosis of Coffee Diseases

Fredy Martínez
Universidad Distrital
Francisco José de Caldas
Bogotá D.C., Colombia

Holman Montiel
Universidad Distrital
Francisco José de Caldas
Bogotá D.C., Colombia

Fernando Martínez
Universidad Distrital
Francisco José de Caldas
Bogotá D.C., Colombia

***Abstract*—The growing and marketing of coffee is an important source of economic resources for many countries, especially those with economies dependent on agricultural production, as is the case of Colombia. Although the country has done a lot of research to develop the sector, the truth is that most of its cultivation is carried out by small coffee families without a high degree of technology, and without major resources to access it. The quality of the coffee bean is highly sensitive to diverse diseases related to environmental conditions, fungi, bacteria, and insects, which directly and strongly affect the economic income of the entire production chain. In many cases the diseases are transmitted rapidly, causing great economic losses. A quick and reliable diagnosis would have an immediate effect on reducing losses. In this sense, this research advances the development of an embedded system based on machine learning capable of performing on-site diagnoses by untrained personnel but taking advantage of the know-how of expert coffee growers. Such a system seeks to instrument the visual characteristics of the most common plant diseases on low-cost, robust, and highly reliable hardware. We identified a deep network architecture with high performance in disease categorization and adjusted the hyperparameters of the model to maximize its characterization capacity without incurring overfitting problems. The prototype was evaluated in the laboratory on real plants for recognized disease cases, tests that matched the performance of the model validation dataset.**

Keywords*—*Cercospora Coffeicola; convolutional neural network; coffee leaf miner; coffee leaf rust; deep learning; image processing; phoma leaf spot

I. Introduction

For developing countries, agriculture is one of the most important economic sectors, both for foreign exchange earnings and for ensuring the food sustainability of their citizens. Colombian coffee enjoys great importance in the international markets because has characteristics that make it stand out, such as its excellent quality and its soft flavor [1]. The importance of coffee is so great that it has been the main source of foreign exchange for the country with 5.3 of the Gross Domestic Product (GDP) and with a production of one million fifty thousand sacks by January 2020. However, its cultivation is mainly carried out by low-income coffee families, with very little access to technologies that help reduce the effect of the plagues that affect the plant [2], [3]. While real-time image processing can be computationally expensive [4], a low-cost artificial system reduces costs for damage and care of the plant because farming families can access these tools at low cost and use them to reduce the spread of disease and artificial intelligence strategies can increase crop performance if they are made accessible to people with modest education and purchasing power [5], [6].

The production and conservation of quality coffee are very difficult for small producers [7]. In Colombia, only Arabic coffees are cultivated, which differ from the Canephora coffees (Robusta coffees) because they are soft, and of greater acceptance in the world market. The harvest is mostly done by small coffee-growing families of medium and low profile. Some plagues attack and make the plant sick, reducing the production and affecting the quality and flavor [8]. These problems have increased considerably in the last decades worldwide, which has affected quality and quantity indicators [9]. Among the most important pests that affect the coffee plant are Coffee Leaf Rust (CLR) [10], the Coffee Borer Beetle (Hypothenemus Hampei) [11], the Coffee Leaf Miner (Leucoptera Coffeella) [11], the Citrus Mealybug (Planococcus Citri), the Coffee Stem, and Root Borer (Plagiohammus Colombiensis), and the red spider. Also of importance are the Iron Spot (Cercospora Coffeicola) [12], the Lint Disease (Corticium Koleroga), the Cock's Eye (Mycena Citricolor) [13], and the Anthracnose (Colletotrichum Coffeanum) [14]. The varieties of Arabic found in Colombia are Tipica (susceptible to CLR), Borbón, Maragogipe, Tabi (resistant to CLR), Caturra (susceptible to CLR), and Colombia variety (resistant to CLR).

Another important factor that negatively affects the cultivation of coffee, and that favors the propagation of plagues and their diseases, is related to the climatic variations of the planting areas [15], [16]. These climatic variations in addition to affecting the growth of the plants tend to increase the aggressiveness of the pests [17]. It has been observed that height affects the intensity of CLR aggression, which is greater in the lower areas with higher temperatures [18], [19].

Prevention and timely diagnosis are essential to stop the advance of pests [20]. Identifying pests at an early stage of infection greatly increases the chances of successful treatment. There are methods for determining the diseases of any plant, such as taking samples of vegetative tissue to a specialized laboratory or bringing an expert agronomist to the crop site. In any of these cases, the disadvantages for the farmer are centered on the time needed to obtain the results and the costs involved. This is why the design of autonomous systems using artificial vision and pattern recognition techniques, as well as some classification algorithms, has been considered for the development of preliminary diagnostic tasks [21], [22]. In this way, the coffee grower can identify the possible disease, its propagation, and with experts and specialists coordinate more quickly and with less cost the correct treatment [23], [24].

Several of the diseases and plagues that are threatening the cultivation of coffee also produce visually detectable effects [25]. The visible effects have been studied as possible indicators of their presence, thanks to the fact that they present specific characteristics [26]. Among these specific characteristics are abnormal coloring of the leaves, deformation of the leaves, and signs of dehydration. These particular characteristics can be used for the process of diagnosis of the disease, or in the opposite case, to diagnose the plant as healthy. RLC is considered by many to be the most severe disease of the coffee crop since it causes the premature fall of the leaves, leading to the death of the plant. The disease has caused great production losses in countries in Asia, Africa, and the Americas. Once the disease appears and establishes itself in a place, it has not been possible to eradicate it, despite multiple strategies implemented by the producing families [27]. It is characterized by pale spots on the underside of the leaves that over time become large yellow or orange spots with the presence of a yellow powder (the spores of the fungus) [28].

In the case of the Cock's Eye disease (Mycena Citricolor), small circular or oval spots are observed, slightly sunken, with a diameter of 6-10 mm on the leaves [29], [13]. The lesions start as dark brown spots with an undefined border, which when reaching their final size present a well-marked border, with little or no chlorosis around them, and can be light brown, grayish, or reddish-brown, with a papery and dry appearance.

Iron Spot (Cercospora Coffeicola) is another important disease that attacks coffee cultivation. It is caused by a fungus that affects the plant in various stages, beginning in the nursery [12]. It is visually characterized by brown spots with a yellowish halo that contrasts with the normal leaf tissue. As the disease progresses, the size of the spot increases, causing the tissue to die. The most serious damage occurs to the fruit, but also affects the leaves. It is transmitted by the fungus Cercospora Coffeicola, and its spot is particularly prevalent in the nursery and on unshaded coffee plantations. In the fruits the infection starts through wounds or exposure to the sun forming lesions similar to those on the leaves, but which eventually stop being circular to become elongated and dark.

Each disease produces characteristic damage to the plant. These damages visually generate geometrical and colorimetric parameters that can be identified through digital image processing [30], [31]. One of the most powerful strategies for image categorization is the convolutional neural networks, which have demonstrated to have a very high capacity to identify information in unknown images after training with categorized cases [32], [33]. Therefore, it is possible to use a neural model to design an embedded, autonomous, and low-cost system capable of identifying in real-time diseases of the coffee plant leaves [34].

The rest of this article is organized as follows. Section II describes the functional characteristics of the embedded system and the working environment of the equipment, which define the design profile of the system. Section III describes the model developed for the detection of anomalies in the coffee leaf, as well as the characteristics of the hardware used, and its configuration. The results that demonstrate the behavior of the classification model are given in Section IV, and in Section V the conclusions of the research and development are presented.

II. Problem Statement

The sustainability of agriculture depends on many factors, including the ability to reduce food losses due to infections caused by bacteria, viruses, and fungi. In this sense, early detection of crop diseases drastically reduces the spread of illnesses, and therefore economic losses. Solution strategies should be developed focusing not only on the nature of the crop in question but also on the social conditions under which production takes place. Our research focuses on the identification of diseases common to the coffee plant, which is why we sought to develop a system that could examine in real-time the leaf of the plant, the place where diseases can be identified. This system aims to detect possible changes in the leaf of the plant that could signal an infection.

The objective of this research is to develop an embedded system for the autonomous and on-site diagnosis of coffee diseases. Other important features of the system include low cost and ease of operation. Among the design features, the need for autonomous operation stands out given the impossibility of connection for the deployment of complex models. In addition to these features, portability and high performance also limit the hardware characteristics to be used.

Among the machine learning schemes evaluated as automatic categorization schemes, those based on deep networks presented the highest values in the evaluation metrics. Consequently, a deep model that can be run in real-time on limited processing hardware should be chosen for implementation. Such a system should have a digital camera for image capture, and the appropriate framework for digital processing. The categorization model must extract the image parameters with the diseases to be identified, so a specific dataset for the problem is required. It should also facilitate the interpretation of results by the user, so the images captured by the user should be labeled according to the diagnosis (Fig. 1).

To design the model, the most frequent diseases that cause the most damage to the plant and coffee production were selected. For these images, we used public databases categorized by experts in the plant. We used 1250 images with a size of each of 2048×1024 pixels, corresponding to Arabica coffee leaves separated into five categories, each category with 250 images. The number of images in each category was kept the same (250) to avoid bias in the model. The first category (category 1) corresponds to healthy leaves, the other four categories correspond to leaves affected by four common plant diseases (each leaf has only one of the diseases): Coffee Leaf Miner (CLM, category 2), Coffee Leaf Rust (CLR, category 3), Phoma Leaf Spot (Phoma Tarda, category 4), and Iron Spot (Cercospora Coffeicola, category 5). Fig. 2 show the detail of the images in each of the categories.

Before training the model, the images will be pre-processed using segmentation and labeling filters to remove the background of the image and keep only the leaf. Color adjustment filters will also be used to enhance the images. In this way, we seek to ensure that each image has the visual information that a human expert would identify. The same processing is applied to the images used in the training as well as those used for model validation (Fig. 3). The system must have a visual

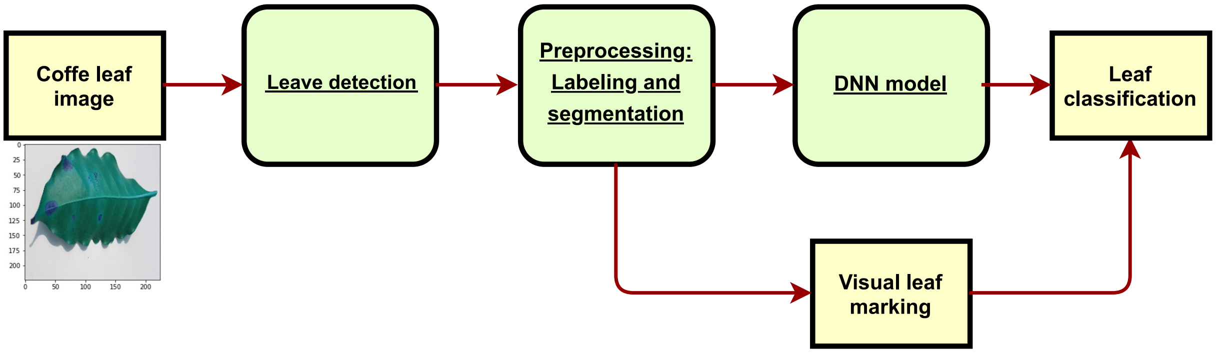


Fig. 1. Pipeline of the Proposed Embedded System.

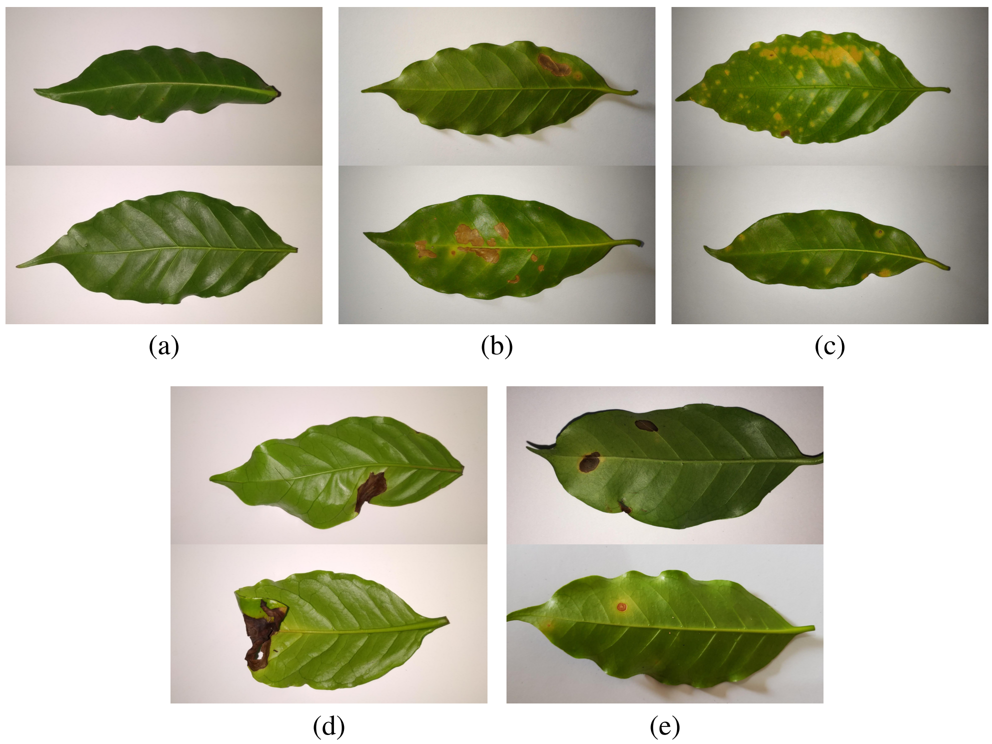


Fig. 2. Sample Images from the Dataset. (a) Healthy Leaves, (b) Coffee Leaf Miner (CLM), (c) Coffee Leaf Rust (CLR), (d) Phoma Leaf Spot (*Phoma Tarda*) and (e) Iron Spot (*Cercospora Coffeicola*)

output in which the user can observe the damage identified on the leaf in real-time. In principle, a screen should be available in which this image is constructed by superimposing on the frame captured by the camera the information related to this labeling and the information related to the categorization.

Among the possible deep models, the best performance was obtained with the ResNet (Residual Neural Network). Convolutional neural networks have convolution layers (convolution filters) that have the effect of filtering the image with a previously trained kernel, capable of detecting primitive features such as lines or curves. Over several layers, the neural network learns to identify these features along with the training data set. The ResNet50 architecture is selected as the topology given its smaller comparative size (fewer parameters), and high initial test results. This feature is achieved thanks to its design, the network topology contemplates short forward connections from the previous layers, which has been observed to increase its accuracy.

III. Methods

The system is composed of three processing modules: leaf detection unit, preprocessing unit, and DNN (Deep Neural Network) based model (Fig. 1). These modules are sequential, the output of one functions as input to the next. The first one corresponds to a set of filters applied to the input image that seeks to identify the morphological characteristics of the leaf in the video frames. These filters look for leaf shape regardless of orientation or background, but prioritize shapes of relative size to the frame, thus requiring the user to focus on individual plant leaves. These initial filters reduce processing

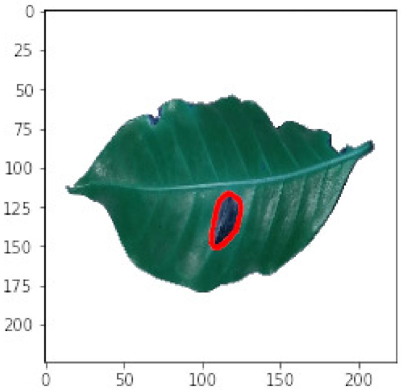


Fig. 3. Image after Segmentation, Labeling, Filtering and Scaling.

requirements by identifying an area in which the second module's preprocessing is applied. The preprocessing module receives as input an area in the frame on which segmentation and labeling are applied to identify areas of the region with characteristics different from those expected in a healthy leaf. This information is transferred to the output screen for user documentation but is also used to precisely delimit the region containing the leaf, which feeds the next module. Finally, this information enters the DNN module, which propagates the network in the trained model, and defines the most likely disease. This information is also displayed on the screen for the user.

The ResNet network was trained with public images corresponding to different databases. The selection of the images considered criteria related to the effect of the disease in the region of interest, the severity of leaf damage, and image capture conditions (real environment and/or laboratory). The images in the dataset were filtered to remove the background, center the leaf on the image, and improve its color level [35]. Also, they were randomly mixed within the stack to improve the performance of the network. To facilitate training and reduce resource consumption, the images were scaled to 256 $\times$ 256 pixels in RGB format. Although the aspect ratio of the images was altered, this does not alter the visual information related to the images, but it does facilitate the design of the neural network.

For neural network training, the color matrices of the images, which make up the input parameters, were normalized to color depths in the range of zero to one. Besides, the 1250 images were randomly separated into two groups, the first group with 80% of the images (1000 images) for neural network training, and a second group with the remaining 20% (250 images) for model validation purposes. For the design of the network structure, the size of the input images is taken into account, $256\times256\times3$ = 196,608, which defines the total number of input nodes. The number of output nodes is defined by the number of network categories, which in our case are five categories, so five output nodes. In the output, a one-hot coding structure was defined to define these five output categories.

The ResNet50 model is a variant of ResNet with a total of 48 convolution layers, along with 1 MaxPool layer and 1 Average Pool layer. The network has a total of 23,597,957 parameters, of which 23,544,837 were adjusted during training. Of these parameters, 10245 corresponded to the dense output network. As optimization function in the model, we use the stochastic gradient descent function. In the optimization we use as error measure the categorical hinge function. During the training, we calculated in each epoch the values of accuracy (or hit rate) and MSE (mean quadratic errors) metrics to observe the performance of the network throughout the training. The final model was trained over 300 epochs with a batch size of 32. Throughout the training, the accuracy increased from 23.3% to 96.5% for the training data.

We selected Arrow Electronics' DragonBoard 410c development board as the platform to evaluate the performance of our neural model as an embedded system. We chose this board for both cost and performance. This board has a Qualcomm APQ8016e 64-bit quad-core processor, Wi-Fi, Bluetooth, and GPS connectivity, and support for Windows 10 IoT Core, Android 5.1, and Debian 8.0. To evaluate the performance of our model, we use Keras 2.4.3 and Tensorflow 2.3.0 installed above Linux Debian OS. Additionally, we used numpy 1.18.5, scipy 1.4.1, scikit-learn 0.22.2, Pillow 7.0.0, glob2 0.7, matplotlib 3.2.2, cv2 4.1.2.30, seaborn 0.11.0, and pandas 1.1.2.

IV. Result and Discussion

The performance of the model was evaluated based on the behavior of the categorization system with the validation images, in this way it was possible to quantify the performance under ideal conditions. The final tests of the prototype were performed in the laboratory with leaves collected directly in the field by the research group. These tests allowed validation of the detection and preprocessing modules.

For the case of the final DNN model tuned for implementation, training was performed over 300 epochs, and accuracy (Fig. 4) and loss values (Fig. 5) were recorded for both training and validation data. The error produced by the training data is continuously reduced throughout the whole process, reaching a final value of 0.07. An equivalent behavior is observed for the accuracy of the training data, which increases continuously throughout the training process from 23.3% to 96.5%. The behavior of the validation data is not as uniform, but an overall reduction of the error at the end of the training process is observed, which although it is lower than that achieved by the training data, keeps decreasing in parallel, guaranteeing the non-existence of overfitting (final loss value of 0.56). The accuracy of the validation data also has a uniformly increasing behavior, parallel to the training data, increasing continuously throughout the training process (final value of 71.5%). These data, while not guaranteeing a perfect classification, do provide for the application of an adequate classification of the analyzed images.

The confusion matrix provides a quick picture of the classification capability of the model as it explicitly shows when one category is confused with another. This allows working separately with different types of error, as well as calculating different model performance metrics. We calculate the confusion matrix for our model using the images from the validation group (unknown images for the model) and assign a heatmap with light colors for the highest number of true

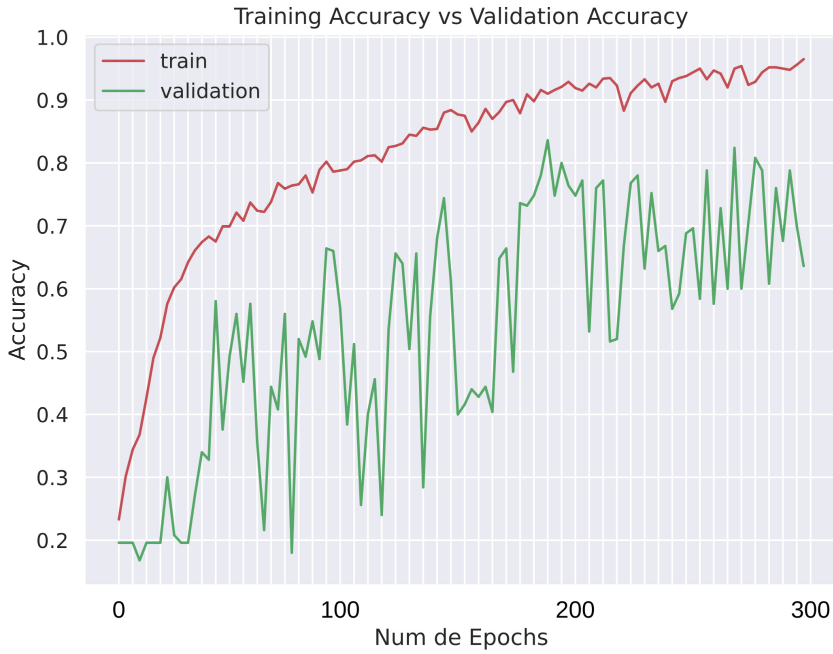


Fig. 4. Model Behavior: Training Accuracy vs Validation Accuracy.

positives, and dark colors for the opposite cases (Fig. 6). The diagonal of the curve clearly shows that the model correctly classifies most of the unknown images. For example, for the healthy leaves' category, 22 of the images were correctly classified in the first category, and for the CLM category, the best performing category, 45 of the images were correctly classified.

To evaluate the performance of the model in a specific way, we calculate the accuracy, recall, f1-score, and support metrics for each of the categories with the validation images (the 250 unknown images for the model). The average precision of the model (percentage of correct positive predictions among all positive predictions) was 73%, with an exceptional classification of diseased leaves with Phoma Leaf Spot (84% precision) and healthy leaves (92% precision). However, the classification of diseased leaves with Coffee Leaf Rust was considerably low (42% precision). The values of recall and f1-score show similar results to those shown by the precision, in the case of recall (percentage of correct positive predictions among all positive predictions that could have been made) some measure of the wrong positive predictions is presented, in this case, the average value drops a little to 64%, which is very similar to the precision value, but the recall for the leaves that are healthy drops to 41%, and the value for the leaves that are sick with Coffee Leaf Rust goes up to 88%. The f1-score corresponds to the harmonic mean of precision and recall, so the above peaks are averaged out at 64%. For the classification model of our project, these values are good enough to support the development of the prototype.

We also calculated the ROC curve (Receiver Operating Characteristic) of the neural model (Fig. 7). This curve graphically shows the sensitivity of the model (ratio of true positives to the ratio of false positives) to variations in the discrimination threshold between categories. In this sense, high average values (0.87) and high values per category (0.85 to 0.93) of true positives versus false positives are observed.

Laboratory tests of the prototype showed not only the correct operation of the classification model within the metric margins but also how the leaf detection and preprocessing modules facilitate the work of the deep model. The need to evaluate the impact of these modules on the overall performance of the system is raised in future work. The capability of the DragonBoard 410c development board to run the software in real-time will also be verified.

V. Conclusion

Early and on-site detection of diseases in coffee crops is of great importance to avoid harvest losses, and to schedule the correct spraying processes. In this sense, in this work, we propose an embedded system based on machine learning for the detection of diseases in the coffee plant. This system is intended to be used directly in crops by farmers without technical knowledge, so its design, in addition to the characteristics of the plant and its diseases, considers aspects of use, cost, and performance. These characteristics of the system constitute the major contribution of the authors in the research.

For the design of the classification model, we selected four high impact diseases for this crop: Coffee Leaf Miner (CLM), Coffee Leaf Rust (CLR), Phoma Leaf Spot (Phoma Tarda), and Iron Spot (Cercospora Coffeicola). Healthy leaves were also assigned a category. These diseases produce visible damage in the coffee leaf that can be identified and classified by image processing. In this sense, we selected a deep neural network type ResNet (Residual Neural Network) to identify and learn the characteristics of the leaves and their diseases. This neural network was selected due to its high performance and lower number of parameters compared to other topologies, including other larger ResNet models. The architecture of the

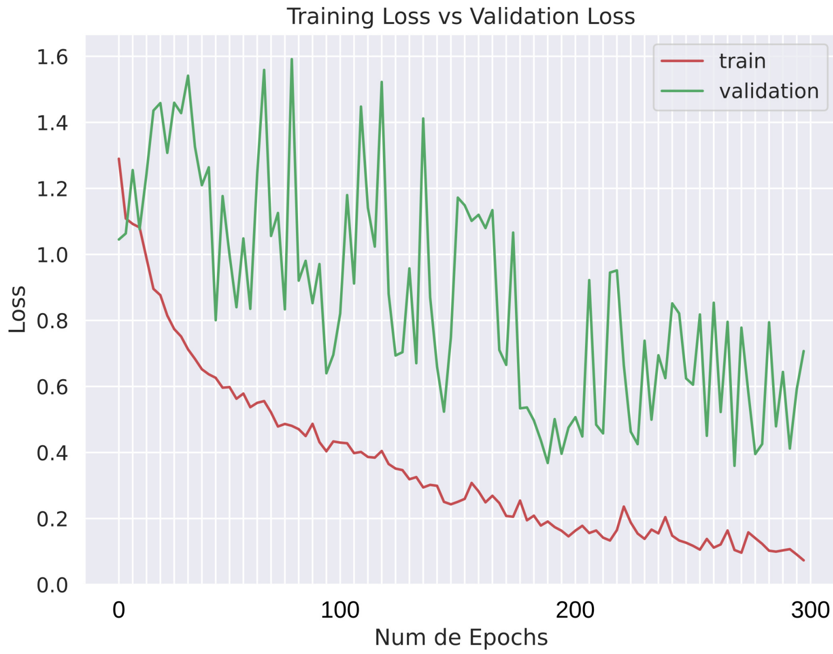


Fig. 5. Model Behavior: Training Loss vs Validation Loss.

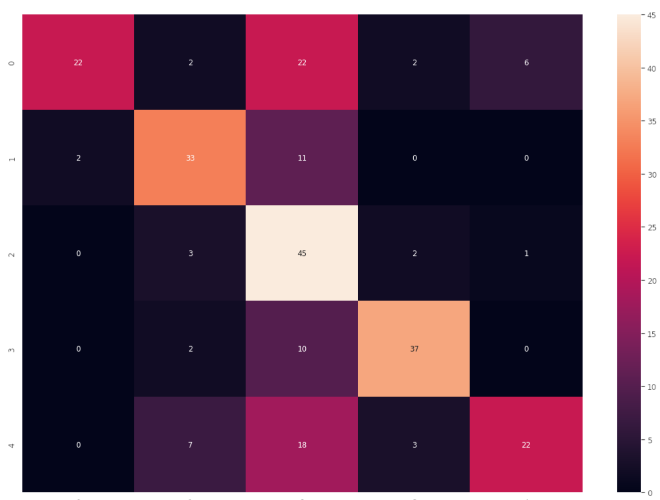


Fig. 6. Confusion Matrix.

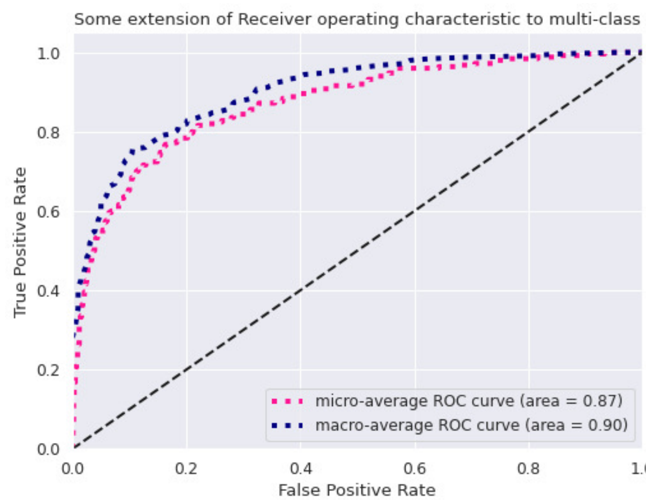


Fig. 7. ROC Curve.

ResNet network was adjusted for input images of 256×256 pixels in RGB format, 50 layers of depth (ResNet50), and five output categories. The database was made up of 250 images in each category, and 80% of them were used for training (1000 images) and 20% for model validation. The training was carried out over 300 epochs taking care not to overfitting the network. To fine-tune the parameters, the error was evaluated using the categorical hinge function, and optimized using the stochastic gradient descent function. The final accuracy achieved by the model was 96.5% for the training data and 63.6% for the validation data (images unknown to the model). This model was implemented on a DragonBoard 410c from Arrow Electronics, running a Debian OS. Preliminary results show low resource consumption and acceptable performance for real-world implementation. Detection of diseased leaves exceeds 91% of cases, and correct disease identification is 64% in the worst case. Research continues to strengthen the training database, apply further fine-tuning to the hyperparameters, and evaluate the impact on the performance of the digital image processing modules.

Acknowledgment

This work was supported by the Universidad Distrital Francisco José de Caldas, specifically by the Technological Faculty. The views expressed in this paper are not necessarily endorsed by Universidad Distrital. The authors thank all the students and researchers of the research group ARMOS for their support in the development of this work.

REFERENCES

[1] D. Bebber, À. Castillo, and S. Gurr, "Modelling coffee leaf rust risk in Colombia with climate reanalysis data," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 371, no. 1709, 2016.

[2] J. Avelino, M. Cristancho, S. Georgiou, P. Imbach, L. Aguilar, G. Bornemann, P. Läderach, F. Anzueto, A. Hruska, and C. Morales, "The coffee rust crises in Colombia and Central America (2008–2013): Impacts, plausible causes and proposed solutions," *Food Security*, vol. 7, no. 2, pp. 303–321, 2015.

[3] D. Velásquez, A. Sánchez, S. Sarmiento, M. Toro, M. Maiza, and B. Sierra, "A method for detecting coffee leaf rust through wireless sensor networks, remote sensing, and deep learning: Case study of the Caturra variety in Colombia," *Applied Sciences (Switzerland)*, vol. 10, no. 2, 2020.

[4] A. Mengistu, S. Mengistu, and D. Melesew, "An automatic coffee plant diseases identification using hybrid approaches of image processing and decision tree," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 9, no. 3, pp. 806–811, 2018.

[5] A. Cadena, E. Peña, and F. Martínez, "Desarrollo de prototipo de quadrotor con sistema de comunicación inalámbrica bidireccional," *Tekhnê*, vol. 13, no. 1, pp. 13–30, 2016.

[6] B. Ortega, R. Biswal, and E. DelaCruz, "Detección de enfermedades en el sector agrícola utilizando Inteligencia Artificial," *Research in Computing Science*, vol. 148, no. 7, pp. 419–427, 2019.

[7] R. Cerda, J. Avelino, C. Gary, P. Tixier, E. Lechevallier, and C. Allinne, "Primary and secondary yield losses caused by pests and diseases: Assessment and modeling in coffee," *PLoS ONE*, vol. 12, no. 1, 2017.

[8] L. Belan, W. de Jesus, A. de Souza, L. Zambolim, J. Filho, D. Barbosa, and W. Moraes, "Management of coffee leaf rust in Coffea canephora based on disease monitoring reduces fungicide use and management cost," *European Journal of Plant Pathology*, vol. 156, no. 3, pp. 683–694, 2020.

[9] L. de Sousa, M. da Silva, and J. Costa, "Leaf-associated bacterial microbiota of coffee and its correlation with manganese and calcium levels on leaves," *Genetics and Molecular Biology*, vol. 41, no. 2, pp. 455–465, 2018.

[10] L. Ehrenbergerová, A. Kučera, E. Cienciala, J. Trochta, and D. Volařík, "Identifying key factors affecting coffee leaf rust incidence in agroforestry plantations in Peru," *Agroforestry Systems*, vol. 92, no. 6, pp. 1551–1565, 2018.

[11] L. de Oliveira and G. de Souza, "Models for simulating the frequency of pests and diseases of Coffea arabica L." *International Journal of Biometeorology*, vol. 64, no. 7, pp. 1063–1084, 2020.

[12] S. Arteaga and D. Espinoza, "Eficacia de dos productos a base de trichoderma, en tres dosis de aplicación para el control de la mancha de hierro (cercospora coffeicola), en café (coffea canephora) variedad robusta a nivel de vivero en el Cantón Francisco de Orellana provincia de Orellana," *Caribeña de Ciencias Sociales*, pp. 1–24, 2019.

[13] J. Silva, M. Resende, A. Monteiro, M. Pádua, L. Guerra-Guimarães, F. Medeiros, S. Martins, and D. Botelho, "Resistance inducers applied alone or in association with fungicide for the management of leaf rust and brown eye spot of coffee under field conditions," *Journal of Phytopathology*, vol. 167, no. 7-8, pp. 430–439, 2019.

[14] B. Zewdie, A. Tack, G. Adugna, S. Nemomissa, and K. Hylander, "Patterns and drivers of fungal disease communities on Arabica coffee along a management gradient," *Basic and Applied Ecology*, vol. 47, pp. 95–106, 2020.

[15] I. Merle, P. Tixier, E. Virginio, C. Cilas, and J. Avelino, "Forecast models of coffee leaf rust symptoms and signs based on identified microclimatic combinations in coffee-based agroforestry systems in Costa Rica," *Crop Protection*, vol. 130, 2020.

[16] K. Belachew, G. Senbeta, W. Garedew, R. Barreto, and E. Del Ponte, "Altitude is the main driver of coffee leaf rust epidemics: A large-scale survey in Ethiopia," *Tropical Plant Pathology*, vol. 45, no. 5, pp. 511–521, 2020.

[17] F. Hinnah, P. Sentelhas, C. Meira, and R. Paiva, "Weather-based coffee leaf rust apparent infection rate modeling," *International Journal of Biometeorology*, vol. 62, no. 10, pp. 1847–1860, 2018.

[18] G. Daba, K. Helsen, G. Berecha, B. Lievens, A. Debela, and O. Honnay, "Seasonal and altitudinal differences in coffee leaf rust epidemics on coffee berry disease-resistant varieties in Southwest Ethiopia," *Tropical Plant Pathology*, vol. 44, no. 3, pp. 244–250, 2019.

[19] T. Liebig, F. Ribeyre, P. Läderach, H. Poehling, P. van Asten, and J. Avelino, "Interactive effects of altitude, microclimate and shading system on coffee leaf rust," *Journal of Plant Interactions*, vol. 14, no. 1, pp. 407–415, 2019.

[20] S. McCook and J. Vandermeer, "The Big Rust and the Red Queen: Long-term perspectives on coffee rust research," *Phytopathology*, vol. 105, no. 9, pp. 1164–1173, 2015.

[21] K. Lewis, M. Quioc, and J. Espineli, "Image processing techniques and data mining algorithms for coffee plant's leaves classification," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 2, pp. 1101–1106, 2020.

[22] F. Montalbo and A. Hernandez, "Classifying barako coffee leaf diseases using deep convolutional models," *International Journal of Advances in Intelligent Informatics*, vol. 6, no. 2, pp. 197–209, 2020.

[23] L. Rodrigues, G. Sera, O. Filho, L. Beriam, and I. de Almeida, "First report of mixed infection by Pseudomonas syringae pathovars garcae and tabaci on coffee plantations," *Bragantia*, vol. 76, no. 4, pp. 543–549, 2017.

[24] A. Chemura, O. Mutanga, M. Sibanda, and P. Chidoko, "Machine learning prediction of coffee rust severity on leaves using spectroradiometer data," *Tropical Plant Pathology*, vol. 43, no. 2, pp. 117–127, 2018.

[25] A. Sánchez and S. Arguijo, "Análisis de imágenes multiespectrales para la detección de cultivos y detección de plagas y enfermedades en la producción de café," *Research in Computing Science*, vol. 147, no. 7, pp. 309–317, 2018.

[26] N. Thiep, K. Soytong, N. Oanh, and P. Hung, "Study on nematodes (Pratylenchus spp.) on Arabica coffee in the Northwestern Vietnam," *International Journal of Agricultural Technology*, vol. 15, no. 4, pp. 675–684, 2019.

[27] L. Belan, L. Belan, A. da Matta, C. Gonçalves, F. Alves, W. Cintra, and W. Moraes, "Standard area diagram with color photographs to estimate the severity of coffee leaf rust in Coffea canephora," *Crop Protection*, vol. 130, 2020.

[28] A. Chemura, O. Mutanga, and T. Dube, "Separability of coffee leaf rust infection levels with machine learning methods at Sentinel-2 MSI spectral resolutions," *Precision Agriculture*, vol. 18, no. 5, pp. 859–881, 2017.

[29] K. Zamora, L. Castro, A. Wang, L. Arauz, and L. Uribe, "Uso potencial de lixiviados y tés de vermicompost en el control del ojo de gallo del cafeto Mycena citricolor," *Agronomía Costarricense*, vol. 41, no. 1, pp. 33–51, 2017.

[30] J. Cepeda and M. Pereira, "Identificación de formas geométricas y colores mediante procesamiento y reconocimiento de imágenes," *Tekhnê*, vol. 2, no. 1, pp. 9–17, 2004.

[31] C. Cortes, J. Leguizamon, and E. Jacinto, "Sistema de procesamiento de imágenes vasculares infrarrojas para aplicación en dispositivos biométricos de control de acceso," *Tekhnê*, vol. 12, no. 1, pp. 13–22, 2015.

[32] F. Martínez and F. Martinez, "Evaluation of deep neural network architectures in the identification of bone fissures," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 18, p. 807, 2020.

[33] F. Martínez, F. Martinez, and E. Jacinto, "Performance Evaluation of the NASNet Convolutional Network in the Automatic Identification of COVID-19," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 10, p. 662, 2020.

[34] F. Martínez and A. Rendón, "Identifier of human emotions based on convolutional neural network for assistant robot," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 18, p. 1499, 2020.

[35] P. Vasavi, A. Punitha, and T. Rao, "Crop leaf disease detection and classification using machine learning and deep learning algorithms by visual symptoms: a review," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 2, pp. 2079–2086, 2022.

Deep Learning-based Hybrid Model for Efficient Anomaly Detection

Frances Osamor, Briana Wellman
Computer Science and Information Technology
University of the District of Columbia, Washington, USA

***Abstract*—It is common among security organizations to run processes system call trace data to predict its anomalous behavior, and it is still a dynamic study region. Learning-based algorithms can be employed to solve such problems since it is typical pattern recognition problem. With the advanced progress in operating systems, some datasets became outdated and irrelevant. System calls datasets such as Australian Defense Force Academy Linux Dataset (ADFA-LD) are amongst the current cohort containing labeled data of system call traces for normal and malicious processes on various applications. In this paper, we propose a hybrid deep learning-based anomaly detection system. To advance the detection accurateness and competence of anomaly detection systems, Convolution Neural Network (CNN) with Long Short Term Memory (LSTM) is employed. The raw sequence of system call trace is fed to the CNN network first, reducing the traces' dimension. This reduced trace vector is further fed to the LSTM network to learn the sequences of the system calls and produce the concluding detection outcome. Tensorflow-GPU was used to implement and train the hybrid model and evaluated on the ADFA-LD dataset. Experimental results showed that the proposed method had reduced training time with an enhanced anomaly detection rate. Therefore, this method lowers the false alarm rates.**

Keywords—Anomaly detection; system call sequence; convolution neural network; long short term memory

I. Introduction

Intrusion detection is the procedure of recognizing malicious behaviors on the network [1]. There are two kinds of intrusion-based detection systems. They are network-based [2] and host-based [3] —network-based examines the traffic on the network and the set of protocols to govern the conceivable intrusions. Pattern matching procedures were used earlier by analysts [4]. Packet header of the data, string and port data are among the few matching features used. Such features improve the consistency and suitability. Particle swarm optimization, gray wolf algorithm, genetic algorithm are some of the current feature selection algorithms [5, 6]. The genetic algorithm runs to the issue of moderately outsized randomness; the gray wolf algorithm traps easily in the local optimum. Host-based intrusion detection system investigates the system maneuver data such as the log files and audits for anomalous pattern behavior. Learning-based methods are popular compare to the rule-based system with the increasing complexity of the environment [7]. Support Vector Machine, Decision Tree, and Random Forest are some of the traditional machine learning algorithms [8, 9, 10, and 11] applied for host-based intrusion detection systems. Hinton projected the idea of deep learning first in 2006 [12]. Deep Learning learns the complex patterns of the data in both supervised and unsupervised ways. Hence, deep learning based algorithms are widely used in natural language processing areas, image processing areas, etc. In recent years, scholars started smearing deep Learning based algorithms to solve intrusion detection problems. To reduce the dimensionality of features and extract meaningful patterns, autoencoders are widely used. It encodes the high-dimensional input data into lower dimension subspace. Convolution Neural Network extracts the significant features from the image data (Gray Scale or Color). Chawla et al. [13] applied stacked CNN and achieved an accuracy of 0.81. Diep et al. [14] focused on custom CNN with word embedding followed by Bi-LSTM and achieved a score of 0.96. In this work, we study the application of learning-based Convolution Neural Network (CNN) and Long Short Term Memory (LSTM) algorithm with optimization of hyperparameters to detect the anomalies [15] with reduced and efficient neural architecture and in optimized training time. The second focus is on the training time of the hybrid model. The training time of the proposed hybrid model is significantly reduced compared to the baseline models. The rest of the divisions are as follows. In Section 2, we discuss the related work on anomaly detection systems. The dataset is explained in Section 3. Next, we provide explanation of proposed framework in Section 4. Next, in Section 5 we explain the experimental results and finally we conclude in section 6 with future work discussion.

II. Related Work

A system call is a call made by a program to the kernel for a service. Its system calls sequence can analyze the behavior of the process. Such traces are used to classify a process as normal or malicious in a host-based intrusion detection system. For such behavior classification of a process, numerous data representation methods can be found in the literature. Sequence n-gram model [16, 17] and pair-gram [18, 19] are among the few representations to extract the meaningful features from the trace of system call. Information retrieval and Natural language processing (NLP) techniques can be used by treating an individual system call as word and the trace of system call as the document. Vector space model and Boolean model-based document representation approach are used to extract the features from the traces. X. Wang et al. [20] extracted the feature using the Boolean model with n-gram approach and further applied the Support Vector Machine algorithm for classification purposes. Vector space model was used by K. Rieck et al. [21] with polynomial function for categorizing the traces of system call. As a distance metric, Y. Liao [22] used a vector space model with a k-nearest neighbor classifier with cosine similarity. Nonetheless, such a classification system considers the system calls frequencies and not its sequence. System call trace datasets like DARPA [23] and University of New Mexico [24] were widely used by researchers for training the learning-based algorithms for analyzing the behavior of the processes. Nevertheless, with the advanced modernization of complex operating systems, these datasets are becoming irrelevant. ADFA datasets by G. Creech et al. [25, 26] are currently used to benchmark the evaluation of intrusion detection systems based on a system call. They used short sequences and a projected model for anomaly detection. It has an extensive pool of system call traces. They trained one-class SVM, Hidden Markov Model, and Extreme Learning Machine algorithms. With one-class SVM, the author achieved 80% accuracy and 90% with ELM [27]. It is a time-intensive task for learning a vocabulary of all sequences. Miao Xie et al. [28] uses principal component analysis [29] to reduce the dimensionality of features on the ADFA-LD dataset and further train k-means clustering and k-nearest neighbor algorithms. There result showed an additional endeavor, they achieved an accuracy of 70% by training one-class SVM. The results and experiments in the proposed work are mostly on the supervised classification [30, 31]. Thus, in this work, we implemented a hybrid learning-based approach.

III. ADFA-LD Dataset

Australian Defense Force Academy generated the ADFA-LD host-based intrusion detection dataset. Linux system calls are being recorded in this dataset. To communicate between the user and kernel mode, standard interfaces are provided by the Linux kernel. The programs in user-mode have very partial reference to hardware devices for accessing the resources of systems, read and write to the device, and new process creation. Every system call has a unique identifier number on the sequence trace. The host designed to characterize a recent Linux server which logs the traces of system calls during a specified period. Authentic programs functioned in normal behavior during such sampling period. The constituted Linux server comprises of file distribution, remote access, database, and web server type of functionality. Entirely patched Ubuntu 11.04 operating system with Linux kernel 2.6.38 was used. For access to services on the web, Apache 2.2.17 and PHP 5.3.5 were configured. Secure Shell Protocol, MySQL, and File Transfer Protocol with their default ports were permitted. As a collaborative web-based tool, TikiWiki 8.1 was installed. Subsequently, various cyberattacks listed in Table I are injected to generate the abnormal traces.

TABLE I. ADFA-LD Dataset

Trace Type	Count	Label
Hydra-FTP	162	Abnormal
Hydra-SSH	148	Abnormal
Webshell	118	Abnormal
Java Meterpreter	125	Abnormal
Meterpreter	75	Abnormal
Add user	91	Abnormal
Training	833	Normal
Validation	4373	Normal

IV. Proposed Framework

Fig. 1 depicts the overall framework. There are three phases. Dataset Collection, Data Engineering and Training of the Detection Algorithm with the final testing.

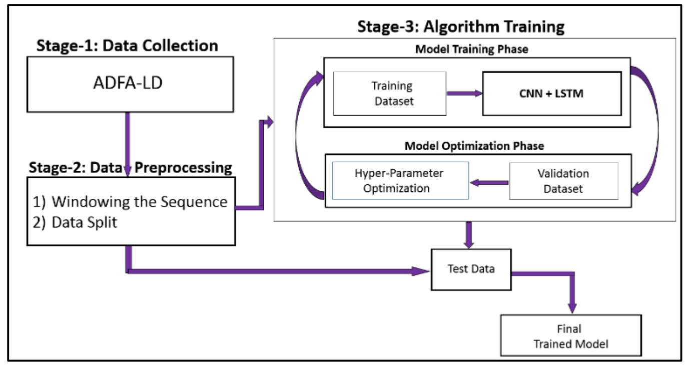


Fig. 1. Proposed Framework.

Phase-1: Data Collection

We employed the ADFA-LD dataset which is explained in Section 3.

Phase-2: Data Engineering (Preprocessing)

The following tasks are performed:

- Windowing the sequence of the system calls as input and output and performing the encoding of the target category [32].
- Applying stratified sampling and divide the dataset into train, validate, and test.

Phase-3: Algorithm Training.

In this phase, the hybrid model is trained as discussed below:

A. *Training of the Model*

A hybrid deep learning-based Convolution Neural Network with long short term memory algorithm is trained in this phase.

1) Convolution neural network: It is a type of neural network, as depicted in Fig. 2, which is widely used for images dataset. It is used for image segmentation, classification, etc.

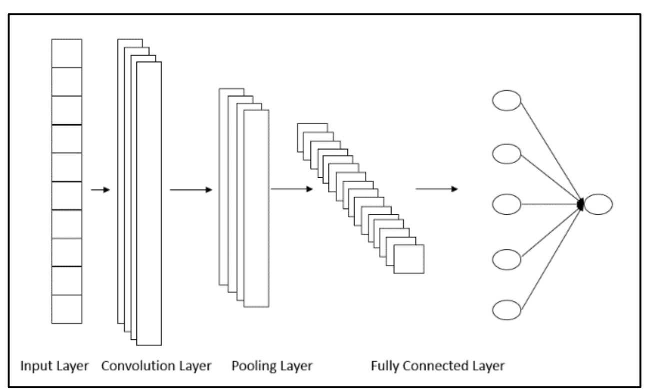


Fig. 2. Convolution Neural Network.

There are two layers in CNN architecture. They are:

a) Convolution layer: This layer applies convolution operation on the input image to extract the meaningful feature. Each pixel value of the image is multiplied by the corresponding filter value, and finally, all the values are added.

b) Pooling layer: This layer is used to reduce the dimensionality of the images. There are different types: Max-Pooling, Min-Pooling, Mean-Pooling, etc.

Rectifier Linear Unit (ReLu) activation function is applied to deal with the non-linearity of the data. Finally, a fully connected dense layer is employed to predict the outcome.

2) Long short term memory: LSTM has memory cells. Numerous gates are attached to the LSTM system to mitigate the vanishing gradient problem of the RNN [33]. These gates behave like a memory. The memory update occurs with the read of an input by the cell. It has the following four gates:

I gate: It enhances the updation of the new incoming memory.

O gate: It updates the new hidden state by selecting the new memory cell info.

F gate: It regulates the amount of old information to be discarded.

M gate: It creates new memory.

For a example dataset of $U=(U_1, U_2, U_3, U_{N-1}, U_N)$ as input to network, it updates the above four gates values to learn the output variable. It is updated as follows:

$$(z_t, d_{t-1}, s_{t-1}) \rightarrow (d_t, s_t) \tag{1}$$

$$i_t = \sigma\,(k_{zi}z_t + k_{di}d_{t-1} + k_{si}a_{t-1} + j_i) \tag{2}$$

$$f_t = \sigma\,(k_{zf}z_t + k_{df}d_{t-1} + k_{af}a_{t-1} + j_f) \tag{3}$$

$$a_t = f_t * a_{t-1} + i_t * tanh(k_{za}z_t + k_{da}d_{t-1} + j_a) \tag{4}$$

$$y_t = \sigma\,(k_{zy}z_t + k_{dy}d_{t-1} + k_{ay}a_t + j_y) \tag{5}$$

$$d_t = y_t * tanh(a_t) \tag{6}$$

where j_i, j_f ,j_y ,j_a represents the bias units for I, F, O, and the memory cell, respectively. Next, k denotes the weight vector, and *a* denotes memory state with output of intermediate layer's as d as shown in Fig. 3.

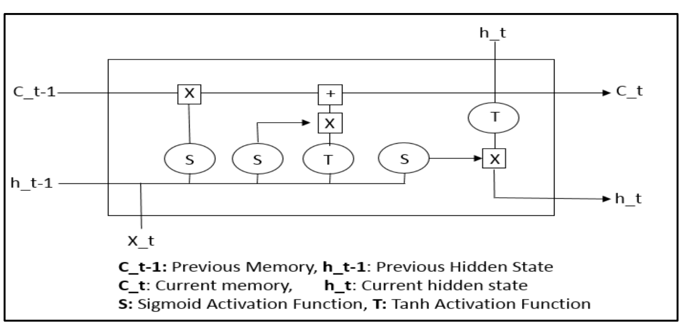


Fig. 3. Long Short Term Memory.

We developed and trained the hybrid CNN [34] with LSTM model architecture shown in Fig. 4. To optimize the hyper-parameters value, we use the K-fold technique with K being 10 [35]. The proposed architecture has a custom ConvLSTM2D layer. This layer takes the raw system call sequences of normal operation as input. The convolution layer applies the convolution operation followed by the pooling technique to extract the meaningful features from the sequences. We use a kernel size of (1,2) with 64 filters. Next, the extracted sequence is fed to the LSTM layer, which learns the patterns of the normal behavior of the sequences. TimeDistributed layer is applied to pass the hidden output at every time step. Finally, the output is flattened and passed to a fully connected dense layer, predicting the final result.

The following metric is used to evaluate the model.

Mean Squared Error (MSE): It finds the square deviations between predicted and actual value.

$$\text{MSE} = \frac{1}{N}\sum\left(Y - \overline{Y}\right)^2 \tag{7}$$

N is the total data points, Y is the actual ground-truth label, and $\overline{Y}$ is the predicted label.

B. Optimization of the Model

We use the validation part of the dataset for optimizing the hyper-parameter [36, 37] for the hybrid model. Parameters that are tuned are as follows:

- Number of Epochs: Total amount of time, data is passed to the model.
- Batch-size: Total count of data sequence given to the model to calculate the loss and update the weight. After the fine-tune optimization, test data is used to evaluate the model.

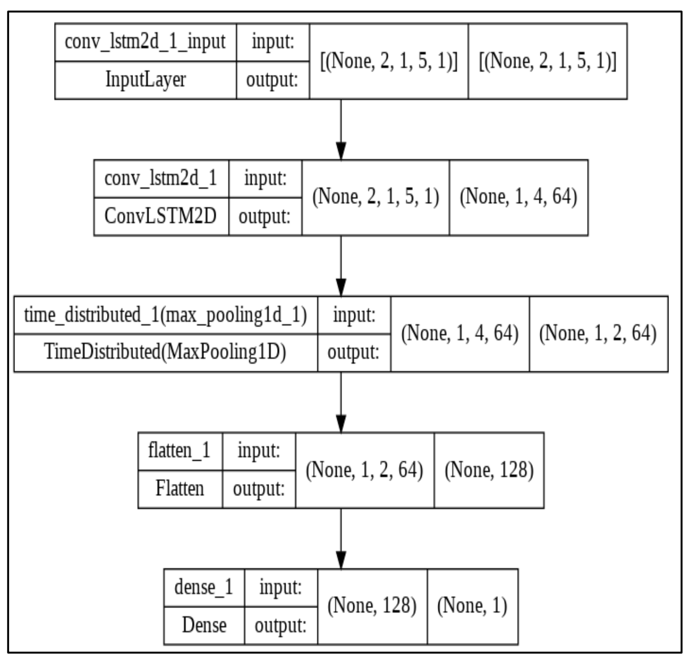


Fig. 4. Model Architecture.

V. Experimental Results

The hybrid model was trained with activation function namely Rectifier Linear Unit. The optimal accuracy rate is achieved with a batch-size of 128. While training the deep neural network, the size of the batch is one of the critical hyper-parameter to be tuned. Stochastic Batch, Batch, and Mini-Batch are the three variants of batch sizes. The model waits till the end of the processing to update the weights in batch variant. Next, model updates the weights after every input sequence in stochastic batch approach. Model updates the weights after every batch in minibatch approach. We train the hybrid model with mini-batch of {16, 32, 64, 128, and 256}. At each batch value, the accuracy value is shown in Fig. 5. The highest accuracy of 0.961 is achieved with a batch value of 128. Fig. 6 and 7 depicts the accuracy and loss at each epoch. After 100^{th} Epoch, the accuracy and loss didn't change much. With the proposed hybrid approach, we achieved 96% accuracy rate and 0.04% loss. The comparison with baseline is given in Table II.

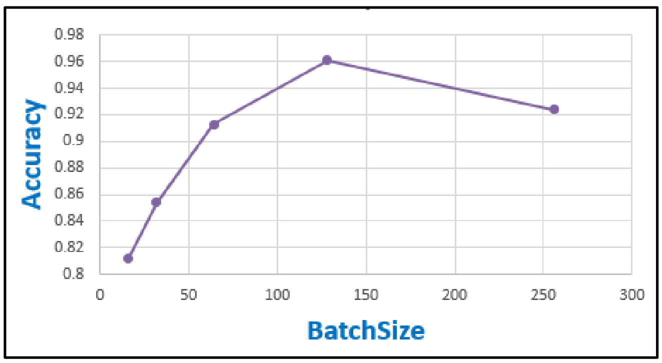


Fig. 5. Accuracy for Individual Batch.

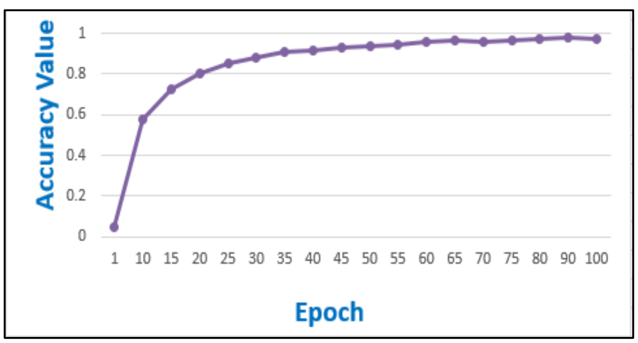


Fig. 6. Accuracy for Individual Epoch.

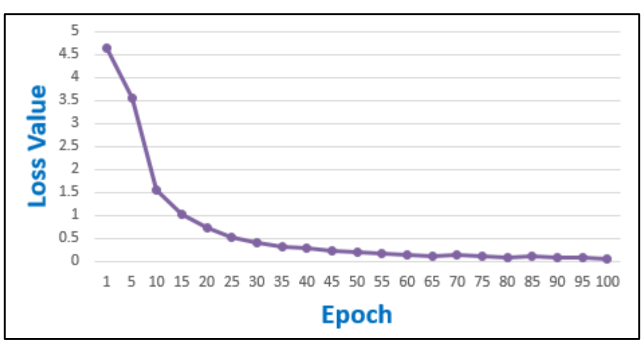


Fig. 7. Loss for Individual Batch.

TABLE II. Comparison with Baseline

Model	Accuracy
CNN	0.81
LSTM	0.88
CNN+LSTM	**0.96**

VI. Conclusion

Intrusion-based detection algorithms exert on the postulate that normal events differ from abnormal events. Anomaly detection algorithms learn a program's behavior during its normal operation. Process behavior is defined by the occurrence of the system call in a particular sequence. We proposed a hybrid deep learning-based CNN with an LSTM model to detect the anomaly in the sequence of system calls. CNN was used to extract the meaningful features, and LSTM was used to learn the patterns of the sequence from the reduced features. The model is trained with the normal process behavior and tested against the normal and malware-infected process. We use the ADFA-LD dataset to test our proposed hybrid model. We achieved an accuracy rate of 96% with a reduced time.

This work can be extended by applying the AutoEncoder neural network to reduce the dimensionality of the data further. Furthermore, numerous natural language processing-based algorithms can be trained, such as Bi-Directional LSTM and Transformers for comparative analysis.

References

[1] Tidjon, L. N., Frappier, M., & Mammar, A. (2019). Intrusion detection systems: A cross-domain overview. *IEEE Communications Surveys & Tutorials*, *21*(4), 3639-3681.

[2] Hamed, T., Dara, R., & Kremer, S. C. (2018). Network intrusion detection system based on recursive feature addition and bigram technique. *computers & security*, *73*, 137-155.

[3] Marteau, P. F. (2018). Sequence covering for efficient host-based intrusion detection. *IEEE Transactions on Information Forensics and Security*, *14*(4), 994-1006.

[4] Kim, H., Hong, H., Kim, H. S., & Kang, S. (2009). A memory-efficient parallel string matching for intrusion detection systems. *IEEE communications letters*, *13*(12), 1004-1006.

[5] Wei, B., Zhang, W., Xia, X., Zhang, Y., Yu, F., & Zhu, Z. (2019). Efficient feature selection algorithm based on particle swarm optimization with learning memory. *IEEE Access*, *7*, 166066-166078.

[6] Ghamisi, P., & Benediktsson, J. A. (2014). Feature selection based on hybridization of genetic algorithm and particle swarm optimization. *IEEE Geoscience and remote sensing letters*, *12*(2), 309-313.

[7] Porras, P. A., & Kemmerer, R. A. (1992, January). Penetration state transition analysis: A rule-based intrusion detection approach. In *Proceedings Eighth Annual Computer Security Application Conference* (pp. 220-221). IEEE Computer Society.

[8] Soni, J., Prabakar, N., & Upadhyay, H. (2019). Feature extraction through deepwalk on weighted graph. In *Proceedings of the 15th international conference on data science (ICDATA'19), Las Vegas, NV*.

[9] Soni, J., & Prabakar, N. (2018). Effective machine learning approach to detect groups of fake reviewers. In *Proceedings of the 14th international conference on data science (ICDATA'18), Las Vegas, NV* (pp. 3-9).

[10] Soni, J., Prabakar, N., & Kim, J. H. (2017, May). Prediction of component failures of telepresence robot with temporal data. In *30th Florida conference on recent advances in robotics*.

[11] Thejas, G. S., Soni, J., Chandna, K., Iyengar, S. S., Sunitha, N. R., & Prabakar, N. (2019, April). Learning-based model to fight against fake like clicks on instagram posts. In *2019 SoutheastCon* (pp. 1-8). IEEE.

[12] Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, *18*(7), 1527-1554.

[13] Chawla, A., Lee, B., Fallon, S., & Jacob, P. (2018, September). Host based intrusion detection system with combined CNN/RNN model. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 149-158). Springer, Cham.

[14] Diep, N. N., Thuy, N. T. T., & Duy, P. H. (2018). Combination Of Multi-Channel Cnn And Bilstm For Host-Based Intrusion Detection. Southeast Asian Journal Of Sciences, 6(2), 147-159.

[15] Soni, J., Prabakar, N., & Upadhyay, H. (2019, December). Behavioral Analysis of System Call Sequences Using LSTM Seq-Seq, Cosine Similarity and Jaccard Similarity for Real-Time Anomaly Detection. In *2019 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 214-219). IEEE.

[16] Forrest, S., Hofmeyr, S. A., Somayaji, A., & Longstaff, T. A. (1996, May). A sense of self for unix processes. In *Proceedings 1996 IEEE Symposium on Security and Privacy* (pp. 120-128). IEEE.

[17] Hofmeyr, S. A., Forrest, S., & Somayaji, A. (1998). Intrusion detection using sequences of system calls. *Journal of computer security*, *6*(3), 151-180.

[18] Hubballi, N., Biswas, S., & Nandi, S. (2011, January). Sequencegram: n-gram modeling of system calls for program based anomaly detection. In *2011 Third International Conference on Communication Systems and Networks (COMSNETS 2011)* (pp. 1-10). IEEE.

[19] Hubballi, N. (2012, January). Pairgram: Modeling frequency information of lookahead pairs for system call based anomaly detection. In *2012 Fourth International Conference on Communication Systems and Networks (COMSNETS 2012)* (pp. 1-10). IEEE.

[20] Wang, X., Yu, W., Champion, A., Fu, X., & Xuan, D. (2007, September). Detecting worms via mining dynamic program execution. In *2007 Third International Conference on Security and Privacy in Communications Networks and the Workshops-SecureComm 2007* (pp. 412-421). IEEE.

[21] Rieck, K., Holz, T., Willems, C., Düssel, P., & Laskov, P. (2008, July). Learning and classification of malware behavior. In *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment* (pp. 108-125). Springer, Berlin, Heidelberg.

[22] Liao, Y., & Vemuri, V. R. (2002). Using text categorization techniques for intrusion detection. In *11th USENIX Security Symposium (USENIX Security 02)*.

[23] DARPA Intrusion Detection Dataset. http://www.ll.mit.edu/ideval/data/

[24] Forrest, S. (2010). University of New Mexico (UNM) Intrusion Detection Dataset. http://www.cs.unm.edu/~immsec/systemcalls.htm

[25] Creech, G. and Hu, J. (2013) Generation of a New IDS Test Dataset: Time to Retire the KDD Collection. Wireless Communications and Networking Conference (WCNC 2013), Shanghai, 7-10 April 2013, 4487-4492.

[26] Creech, G. (2014) Developing a High-Accuracy Cross Platform Host-Based Intrusion Detection System Capable of Reliably Detecting Zero-Day Attacks. Ph.D. Dissertation, University of New South Wales, Sydney.

[27] Creech, G. and Hu, J. (2014) A Semantic Approach to Host-Based Intrusion Detection Systems Using Contiguous and Discontiguous System Call Patterns. IEEE Transactions on Computers, 63, 807-819.

[28] Xie, M., Hu, J. and Slay, J. (2014) Evaluating Host-Based Anomaly Detection Systems: Application of the One-Class SVM Algorithm to ADFA-LD. Proceedings of the 11th IEEE International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2014), Xiamen, 19-21 August 2014, 978-982.

[29] Soni, J., Peddoju, S. K., Prabakar, N., & Upadhyay, H. (2021). Comparative Analysis of LSTM, One-Class SVM, and PCA to Monitor Real-Time Malware Threats Using System Call Sequences and Virtual Machine Introspection. In *International Conference on Communication, Computing and Electronics Systems* (pp. 113-127). Springer, Singapore.

[30] Xie, M. and Hu J. (2013) Evaluating Host-Based Anomaly Detection Systems: A Preliminary Analysis of ADFA-LD. Proceedings of the 6th IEEE International Congress on Image and Signal Processing (CISP 2013), Hangzhou, 16-18 December 2013, 1711-1716.

[31] Xie, M., Hu, J., Yu, X. and Chang, E. (2014) Evaluating Host-Based Anomaly Detection Systems: Application of the Frequency-Based Algorithms to ADFA-LD. Proceedings of 8th International Conference on Network and System Security (NSS 2014), Lecture Notes in Computer Science, 8792, 542-549.

[32] Soni, J., Prabakar, N., & Upadhyay, H. (2019). Comparative Analysis of LSTM Sequence-Sequence and Auto Encoder for real-time anomaly detection using system call sequences.

[33] Medsker, L. R., & Jain, L. C. (2001). Recurrent neural networks. *Design and Applications*, *5*, 64-67.

[34] Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017, August). Understanding of a convolutional neural network. In *2017 international conference on engineering and technology (ICET)* (pp. 1-6). IEEE.

[35] Rathore, P., Soni, J., Prabakar, N., Palaniswami, M., & Santi, P. (2021). Identifying groups of fake reviewers using a semisupervised approach. *IEEE Transactions on Computational Social Systems*, *8*(6), 1369-1378.

[36] Thejas, G. S., Soni, J., Boroojeni, K. G., Iyengar, S. S., Srivastava, K., Badrinath, P., & Upadhyay, H. (2019, December). A multi-time-scale time series analysis for click fraud forecasting using binary labeled imbalanced dataset. In *2019 4th International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS)* (Vol. 4, pp. 1-8). IEEE.

[37] Soni, J., Prabakar, N., & Upadhyay, H. (2020). Visualizing High-Dimensional Data Using t-Distributed Stochastic Neighbor Embedding Algorithm. In *Principles of Data Science* (pp. 189-206). Springer, Cham.

An Ensemble Deep Learning Approach for Emotion Detection in Arabic Tweets

Alaa Mansy, Sherine Rady, Tarek Gharib
Department of Information Systems, Faculty of Computers and Information, Ain Shams University, Cairo, Egypt

***Abstract*—Now-a-days people use social media websites for different activities such as business, entertainment, following the news, expressing their thoughts, feelings, and much more. This initiated a great interest in analyzing and mining such user-generated content. In this paper, the problem of emotion detection (ED) in Arabic text is investigated by proposing an ensemble deep learning approach to analyze user-generated text from Twitter, in terms of the emotional insights that reflect different feelings. The proposed model is based on three state-of-the-art deep learning models. Two models are special types of Recurrent Neural Networks RNNs (Bi-LSTM and Bi-GRU), and the third model is a pre-trained language model (PLM) based on BERT and it is called MARBERT transformer. The experiments were evaluated using the SemEval-2018-Task1-Ar-Ec dataset that was published in a multilabel classification task: Emotion Classification (EC) inside the SemEval-2018 competition. MARBERT PLM is compared to one of the most famous PLM for dealing with the Arabic language (AraBERT). Experiments proved that MARBERT achieved better results with an improvement of 4%, 2.7%, 4.2%, and 3.5% regarding Jaccard accuracy, recall, F1 macro, and F1 micro scores respectively. Moreover, the proposed ensemble model showed outperformance over the individual models (Bi-LSTM, Bi-GRU, and MARBERT). It also outperforms the most recent related work with an improvement ranging from 0.2% to 4.2% in accuracy, and from 5.3% to 23.3% in macro F1 score.**

Keywords*—*Deep learning; emotion detection; transformers; RNNs; Bi-LSTM; Bi-GRU

I. INTRODUCTION

Twitter is a famous 24/7 active social media platform with many signed users sharing their activities, thoughts, and feelings at any time. People post tweets, stream live videos, chat with each other, companies create and manage a lot of marketing campaigns to promote their products, and even much more services are provided.

These days no one can give up using online social networks because it makes them feel connected all the time. Also, they can express their feelings and emotions whether they are happy, sad, surprised, anticipated, or any other feelings during their online activities.

A lot of expressions and words in our daily written text over the web may reflect our feelings. Not only that but also it may affect other people significantly because we believe that every simple word reflects an impact. For example, posting a tweet like that:

"I got COVID-19 twice even though I have been vaccinated the vaccine is useless",

Such simple words can kill a lot of people affected by that virus. Elderly people who have chronic diseases will realize that death is their next step because it makes them feel frustrated. By analyzing that content, everything that may affect a lot of people can be controlled. For Example, social networks can utilize a model for emotion detection in their platforms as an option to prevent such disappointing statuses from being appeared in their customers' timelines. In this way, they can control and restrict anxiety, frustration, and much more.

Emotion Detection or ED is one of the hottest research topics in the field of Natural Language Processing (NLP). ED is considered different from Sentiment Analysis (SA), where SA task is to recognize polarities from text such as positive, negative, or neutral. On the other hand, ED aims to get emotional insights from what has been typed [1].

When reading a tweet, it may reflect one of the following feeling polarities (positive, negative, or neutral). This study is not focusing on the detection of these polarities, but it goes deeper to detect different emotions like (joy, anger, surprise, etc.).

Scientists have summarized ED activities in a set of approaches that determine how exactly emotions are represented. The most famous emotional model is the Discrete Emotion Model (DEM) like Ekman's model which contains six basic emotions which are anger, fear, disgust, happiness, sadness, and surprise. The other models are Dimensional Emotion Model (DiEM) like Plutchik's Emotion Model and Russell's Circumplex Model [2][3].

Suppose a text presented in a user-generated tweet like this: "غضب دفين يفقد الاشياء الوانها لتصبح رمادية وتنعدم لذة الحياة", humans can simply understand the context of this sentence by understanding each word based on the understanding of previous and next words. Also, they can understand the implied emotion of the user who posted the tweet (tweeter) which is sad or angry.

Words in a sentence are linked with each other's in a certain sequence to form a meaning, understanding that meaning is called "Contextual Understanding". Traditional machine learning techniques cannot understand the context very well. Deep learning (DL) sequence models can be utilized to make machines simulate human understanding.

Sequence models such as Recurrent Neural Networks (RNNs) can understand the context by memorizing words and getting the relationships between them. But they have shown some shortages known by the problems of vanishing and exploding gradients. Accordingly, new generations of RNNs have been developed to overcome that shortage. For example, LSTM and GRU models can deal with long-term dependencies and tackle the problems mentioned above.

Although there is a lack of Arabic resources and research studies on Arabic contextual understanding, different Arabic language models were developed to support this point. The most famous one is called AraBERT [4] which is an Arabic pre-trained language model (PLM) based on Bidirectional Encoder Representations from Transformers (BERT) [5]. AraBERT was pretrained using more than 20GB of Arabic text from different sources like Arabic Wikipedia, Arabic news websites, and others. It is based on Modern Standard Arabic (MSA), and it gets better results when fine-tuned using MSA datasets. PLMs are considered a part of Transfer Learning (TL) that support the research in this area and tackle the problem of limited resources. Other models were pretrained based on both MSA and Arabic Dialects (AD) like MARBERT PLM that is utilized in this study.

State-of-the-art neural networks (Bi-LSTM, Bi-GRU, and MARBERT) have been ensembled to deal with a multilabel classification task for emotion detection in user-generated Arabic tweets that were collected and shared during SemEval-2018 task-1: Affect in Tweets.

In Section II, related work is discussed. Section III discusses the proposed ensemble model. In Section IV, empirical results and discussion are investigated. Finally in Section V, the conclusion and future work.

II. Related Work

A lot of research studies have been conducted to get emotional insights and understand the context of English text. Unfortunately, there exist few studies related to the Arabic language because of different challenges related to the complexity of this language, the lack of existing Arabic resources, and different available Arabic dialects. Attention to the analysis of the Arabic language has increased in the last decade due to the need for digital transformation in Arab communities. The Arabic language has a lot of different dialects that are spoken by around 422 million speakers all over the world which is considered a big challenge in the analysis. In the following subsections, ED generic and closely related studies are discussed.

A. Survey Studies

Alswaidan et al. [6] surveyed the state-of-the-art approaches related to emotion detection ED in the textual content for English and some other languages. They mentioned the available resources (corpora and lexicons) for working with ED tasks and addressed some challenges like (I) The challenge of detecting implicit emotions which are hidden in the text. (II) The problems related to size and quality in the available datasets. (III) Limited resources in some languages like Arabic.

Another survey study by Acheampong et al. [7] investigated the ED problem in text content by mentioning all available emotion-related datasets like (ISEAR, SemEval, EMOBANK, EmoInt, Cecilia Ovesdotter Alm's Affect data, Daily Dialog, AMAN'S Emotion, Grounded Emotion data, Emotion-Stimulus data, Crowdsourcing, MELD, Emotion and Smile dataset). Also, they have mentioned the different approaches used to analyze and detect the emotional insights from that data (the rule construction approach, ML approach, and the hybrid approach). And they have made a comparison between different related works in terms of (used approaches, datasets, and limitations).

Similarly, a Systematic Literature Review (SLR) was introduced by William et al. [8] and listed the closely related studies used for text-based depression detection. Also, they aimed to identify and analyze different text-based approaches for the early detection of depression in social media posts. Their results showed that using BiLSTM along with the attention model performs well on depression-related textual data. They also made an experiment by using a BERT-based model and achieved better results compared to the studies mentioned in the SLR. Their experiments used a BERT-based model for the classification task. They also suggested a new method to deal with long sequences by summarizing the text before feeding it into the model. The model depends on a dataset crawled from Reddit.

B. Utilizing Traditional ML and DL Models

Mohammad et al. [9] shared a task called “Affect in Tweets” in the SemEval-2018 competition, which includes a list of subtasks for detecting the emotional states of the tweeters from their text-based tweets. They streamed and annotated some Arabic tweets to form twitter-based labeled datasets represented in three different languages English, Arabic, and Spanish. About 200 team members participated in this competition. Different ML and DL algorithms like (Bi-LSTM, CNN, Gradient Boosting, Linear Regression, Logistic Regression, LSTM, Random Forest, RNN, and SVM) were used. Badaro et al. [10] improved the performance of the emotion classification task by utilizing a pre-trained word embedding model (Aravec) and achieved the best evaluation metrics for (Arabic EC subtask) by using SVC L1 classifier that achieved 48.9%, 61.8%, 46.1% for accuracy, micro f1, and macro F1 scores respectively.

Baali et al.[11] presented a study for classifying emotions in tweets written in the Arabic language. They have used Convolutional Neural Networks (CNN) trained on top of trained word vectors. They compared the results of their approach with three ML algorithms (SVM, NB, and MLP). Their proposed approach was evaluated on the Arabic dataset provided by Sem- Eval for the emotion intensity ordinal classification task (EI-oc). Their results were 99.90% as training accuracy, and 99.82% as validation accuracy.

Khalil et al. [12] proposed a Bi-LSTM deep learning model for the task of emotion classification (EC) in Arabic tweets that were shared SemEval-2018 competition. They have merged the dataset files into only one file to use in the cross-validation process. Aravec with CBOW for the word embedding phase has been used. Their results have shown [Jaccard Accuracy 0.498, Micro Precision 0.695, Micro Recall 0.551, and Micro F1 score 0.615].

C. Utilizing Pre-Trained Language Models (PLMs)

One of the challenges of ED in Arabic text is the limited resources of the Arabic language. As a result, Transfer Learning have been emerged to help pre-train of an NLP model on one large dataset and then quickly fine-tune the model to adapt to other NLP tasks. Also, the nature of that dataset may affect the fine-tuning process i.e., if the model was pre-trained on a dataset containing emotional-related content it will give the best results in ED tasks compared to the model that was pre-trained using other natures of data. Also, some of the existing PLMs were pre-trained using Arabic MSA like AraBERT introduced by Antoun et al. [4] which gets lower results when compared to other PLMs like Abdul-Mageed et al. [13] who introduced MARBERT PLM that was pretrained using both Arabic MSA and different Arabic dialects. Another research study by Abdelali et al. [14] trained five different Arabic BERT models of QARiB using the original implementation of the BERT model implemented by google for both Arabic MSA and Arabic Dialects. Also, they have compared their results with three existing PLMs (mBERT, ARABERTv0.1&v1, ArabicBERT). And the evaluation was conducted using 5 different datasets represented in the following tasks (1) Named Entity Recognition (2) Emotion detection [SemEval2018-Ar-Ec] (3) QADI Arabic Dialects Identification (4) Offensive language detection (5) Sentiment Analysis. Macro-averaged F1 score was used as an evaluation metric, and the results related to the (EC task) using the dataset SemEval2018-Ar-Ec showed that the QARiB25 mix achieved the best macro-averaged F1 score equal to 46.8 %.

Researchers continued to investigate the development of Arabic language models thought conducting a lot of experiments like Al-Twairesh [15] who conducted ten experiments using different models starting from traditional TF-IDF to the recent state-of-the-art BERT models (TF-IDF, AraVecCBOW100, AraVecSG100, AraVecCBOW300, AraVecSG300, AraBertv01, AraBertv1, ArabicBertBase, ArabicBertLarge, Multi-Dialect Bert) on SemEval-2018 dataset. And the results showed that the Arabic BERT-Large model achieved the best results compared to other models.

Others utilized the contextualized embeddings of the PLMs to support other DL models like Elfaik et al. [2] who investigated the problem of Arabic Emotion detection (multilabel emotion classification) in tweets by combining the generated contextualized embeddings using AraBERT and an attention-based LSTM-BiLSTM deep model. The attention mechanism is applied to the output of LSTM-BiLSTM to guarantee different words. Their proposed approach was evaluated using the dataset of SemEval-2018-Task1-Ar-Ec (Affect in Tweets). Their results show that the proposed approach achieves accuracy (53.82%).

Samy et al. [16] researchers utilized some social intelligence and proposed a context-aware gated recurrent unit (C-GRU) to solve the problem of multi-label classification in Arabic-related tweets represented in the SemEval-2018-Task1-Affect in tweets (EC subtask). They have related each tweet with a specific topic, and they depend on what is called social influence where people in the same network can share topics and in the same topic, they can find similar emotions. They have used SemEval-2017 for the topic classification task and SemEval-2018-Ec-Ar for the emotion detection task. They have used Jaccard-similarity for accuracy, F1 macro average, and F1 micro average which achieved results of 0.532, 0.648, and 0.495 respectively.

D. Utilizing Ensemble Techniques

AlZoubi et al. [1] implemented an ensemble approach that contains [bidirectional GRU_CNN (BiGRU_CNN), conventional neural networks (CNN), and XGBoost regressor (XGB)] to be used in solving the emotion intensity (EI-reg) subtask of the SemEval-2018 Task1 (Affect in Tweets). Their proposed ensemble approach was evaluated using the dataset of the SemEval-2018 Task1 El-reg. Results show that their model achieved a Pearson of (69.2%).

Alswaidan et al. [17] proposed three different models, a human-engineered feature-based (HEF) model, a deep feature-based (DF) model, and a hybrid of both models (HEF+DF) for the emotion detection task in Arabic text. And they measured the performance of the proposed models using three different datasets (SemEval2018-Ar-Ec, IAEDS, and AETD). Regarding the SemEval2018-Ar-Ec dataset, the hybrid model achieved the best results of 0.512, 0.631, and 0.502 for Jaccard accuracy, F^{micro}, and F^{macro} scores respectively.

Talafha et al. [18] investigated the Arabic dialect identification problem and trained Arabic-BERT [19] using 10M unlabeled tweets shared in Nuanced Arabic Dialect Identification Task 1 (NADI) and the result was a new pre-trained language model called Multi-dialect-Arabic-BERT. Also, they utilized an ensemble technique (element-wise average) to get the highest value of the predicted probabilities per class for each of the four models. Their results are 44.07 for accuracy and 29.03 for the F1 score.

Closely related studies have been analyzed and concluded in the chart area shown in Fig. 1 that presents the progress till now regarding the EC task using the SemEval-2018-Ar-Ec dataset. As shown in the figure, utilizing PLMs has shown some progress in accuracy compared to other models. Similar studies that use PLMs for the EC task of SemEval-2018 didn't use the most suitable PLMs because most of them have used models that were pre-trained using non-emotional related content. To the best of our knowledge, no one has fine-tuned the MARBERT model using the SemEval-2018-Task1-Ec-Ar dataset. Also, we have used the ensemble model to combine different contextual understanding experiences that can help in getting better results.

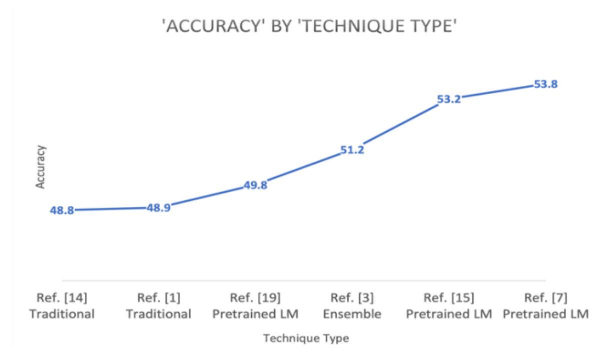


Fig. 1. Progress of Emotion Classification Task against SemEval-Ec Dataset.

III. Proposed Model

The proposed ensemble model for multi-label emotion classification EC in the Arabic language is shown in Fig. 2 which consists of six layers (a) Preprocessing layer (b) Word embedding layer (c) Processing Layer (d) Testing Layer (e) Ensemble layer (f) Classification layer. The details of these layers are explained in the following subsections.

A. Preprocessing Phase

Data preprocessing is considered one of the most important phases in machine learning applications to avoid misleading results and get better insights. In this section, the preprocessing steps will be discussed in detail with an example from our dataset.

As shown in Table I, a user-generated tweet from the SemEval2018-Ar-Ec dataset has been preprocessed using the most common preprocessing techniques like removing English characters, numbers, stop words, repeating chars, punctuation marks, and Arabic diacritics. Also, text normalization and emojis replacement steps have been added.

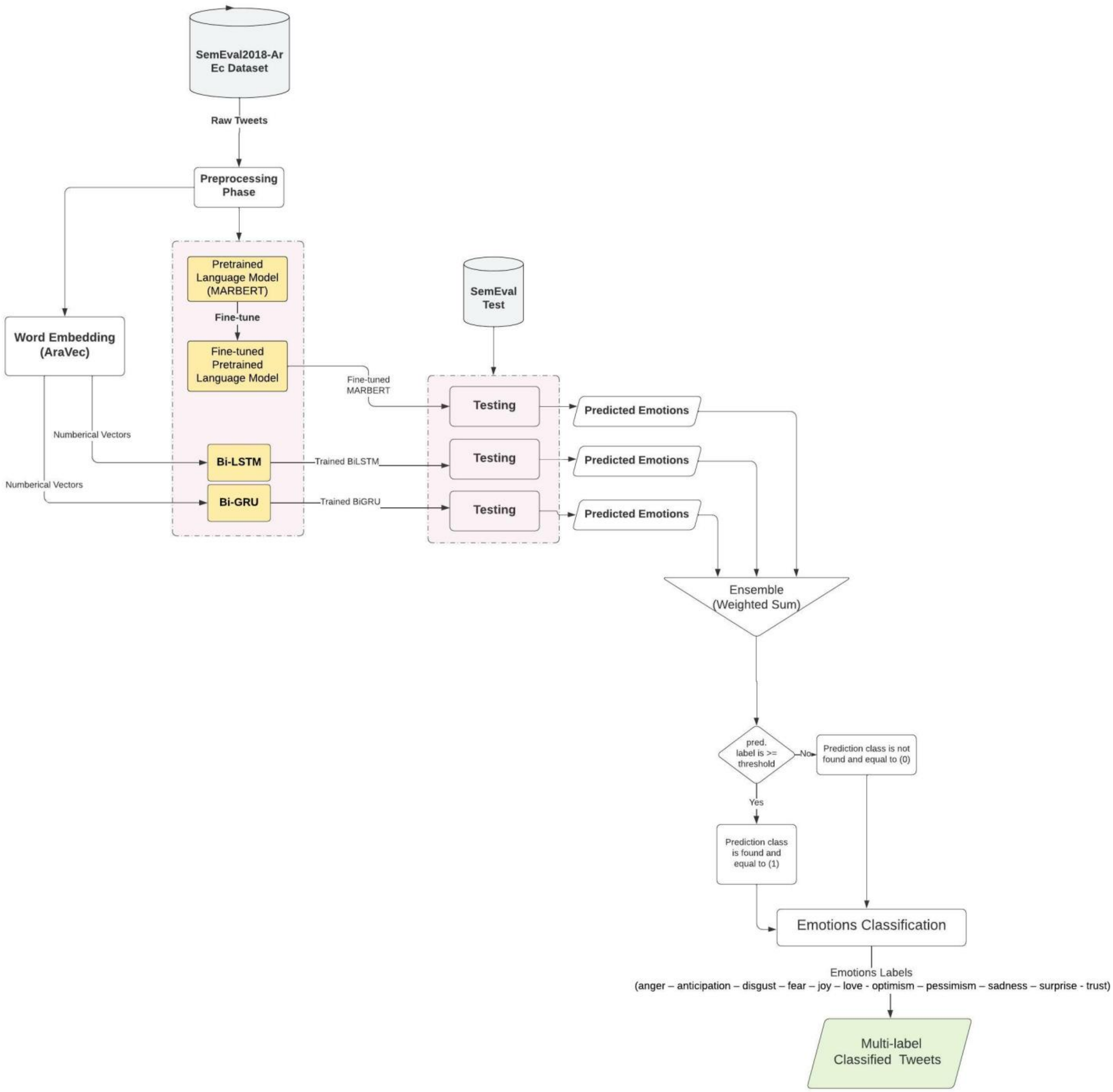


Fig. 2. The Proposed Ensemble Model.

TABLE I. PREPROCESSING EXAMPLE

Preprocessing Method	Example
Raw Tweet	الواحد حواليه كميه negative energy تكفيه قروون لقداام والله ..الدنيا كلها بقيت بؤس وكآبه ..متعرفش ده غضب من ربنا ولا ايه 💔😔
Remove English Characters	الواحد حواليه كميه تكفيه قروون لقداام والله .. الدنيا كلها بقيت بؤس وكآبه .. متعرفش ده غضب من ربنا ولا ايه💔😔
Remove Numbers	الواحد حواليه كميه تكفيه قروون لقداام والله .. الدنيا كلها بقيت بؤس وكآبه .. متعرفش ده غضب من ربنا ولا ايه💔😔
Remove Stop Words	الواحد حواليه كميه تكفيه قروون لقداام والله .. الدنيا كلها بقيت بؤس وكآبه .. متعرفش ده غضب ربنا ايه💔😔
Normalize Arabic	الواحد حواليه كميه تكفيه قروون لقداام والله .. الدنيا كلها بقيت بءس وكابه.. متعرفش ده غضب ربنا ايه💔😔
Remove Arabic Diacritics	الواحد حواليه كميه تكفيه قروون لقداام والله.. الدنيا كلها بقيت بءس وكابه .. متعرفش ده غضب ربنا ايه💔😔
Replace Emojis	الواحد حواليه كميه تكفيه قروون لقداام والله .. الدنيا كلها بقيت بءس وكابه .. متعرفش ده غضب ربنا ايه وجه متأمل قلب مجروح
Remove Repeating Char	الواحد حواليه كميه تكفيه قرون لقدام والله .الدنيا كلها بقيت بءس وكابه . متعرفش ده غضب ربنا ايه وجه متأمل قلب مجروح
Remove Punctuations	الواحد حواليه كميه تكفيه قرون لقدام والله الدنيا كلها بقيت بءس وكابه متعرفش ده غضب ربنا ايه وجه متأمل قلب مجروح

The original user-generated tweet was:

الواحد حواليه كميه negative energy تكفيه قروون لقداام والله ..الدنيا كلها
بقيت بؤس وكآبه .. متعرفش ده غضب من ربنا ولا ايه💔😔

and the result after preprocessing is:

الواحد حواليه كميه تكفيه قرون لقدام والله الدنيا كلها بقيت بءس وكابه
متعرفش ده غضب ربنا ايه وجه متأمل قلب مجروح

- English Characters Removal: In this step, all English characters in both lower and upper cases (A-Z & a-z) are removed.
- Numeric Values Removal: Removing any numbers in the tweet's text.
- Stop words Removal: Removing all stop words that will not affect the meaning of the tweet like (هو – هي – الذى – ولا – هناك - ...).
- Arabic Normalization: Returning chars to its original form like (["[إأٱآا]", "ا"], ["ى", "ي"], ["ؤ", "ء"], ["ئ", "ء"], ["ة", "ه"]).
- Arabic Diacritics Removal: Removing all diacritics like (fatHa فتحة َ◌, kasrah كسرة ِ, DHammah ضمة ُ◌, tanween ً◌ and others).
- Emojis Replacement: Emojis reflect a tweeter's feelings and their existence is very important for understanding the context, so we have replaced them with plain text because it may affect the classification of tweets. A dataset from Kaggle[1] contains 4581 emojis and their meanings in the English language have been used. All English text has been translated into Arabic using a python script and a google translation API. As an example, (😔) is replaced with "وجه متألم" which reflects sadness.
- Repeated Chars Removal: In this step, any repeating characters will be removed like "قروون", it will be "قرون". Also, special words have been managed like (الله – اللهم - ...) to keep it with its original text.
- Punctuations Removal: Removing all punctuation marks like [`÷×؛<>_()*&^%#][−"…"!|+¦~{}',.؟":/،_–]

By the end of this section, a cleared output is now ready for input into the next layers.

B. Word Embedding Layer

In this layer, the output of preprocessing layer is entered into the word embedding layer AraVec[2], a pre-trained word embedding model first introduced by [19]. Text is converted into numerical vectors to be ready for the training phase. This layer is used only when preparing the input for both (Bi-GRU and Bi-LSTM models).

C. Processing Layer

This layer consists of three different models (MARBERT transformer, Bi-LSTM, and Bi-GRU).

1) Pretrained language model (MARBERT Transformer): As NLP is a challengeable task, researchers are seeking continuously to improve the performance of models for a better contextual understanding of user-generated content. Pretrained language models have emerged during the last decade as a part of transfer learning in the ML field. The most famous one in dealing with Arabic text is ARABERT which was introduced by [4] and is based on BERT [5] which is a deep learning transformer model. DL transformer is first introduced by [20] and it relies on the self-attention mechanism. Self-attention is used to relate different positions of a single sequence to compute a representation of the same sequence and help more in contextual understanding.

As mentioned by [13] ARABERT has some limitations related to dealing with different Arabic dialects and the number of tasks that have been used. As a result, a pretrained language model based on BERT has been introduced which is called MARBERT. It was pretrained based on both MSA and Arabic dialects, unlike ARABERT which was pre-trained only on Arabic MSA. ARABERT had been tested and evaluated using only three tasks: named entity recognition NER, question answering QA, and sentiment analysis SA while MARBERT had been used with a set of tasks represented in five basic

[1] https://www.kaggle.com/eliasdabbas/emoji-data-descriptions-codepoints?select=emoji_df.csv

[2] https://github.com/bakrianoo/aravec

categories (1) Sentiment Analysis SA (2) Named Entity Recognition NER (3) Dialect Identification DI (4) Topic Classification TC (5) Social Meaning SM like (emotion, irony, sarcasm, …). Because MARBERT transformer was used before in the emotion detection task, it has an emotion-related contextual understanding experience. In this paper, a fine-tuned MARBERT on the SemEval-2018-Ec-Ar task has been proposed.

2) Bi-LSTM model: Bidirectional Long Short Term Memory Model or Bi-LSTM is an extension of the normal LSTM introduced by Hochreiter & Schmidhuber in 1997 [21]. LSTM was developed to avoid the short-term dependency problem as it can remember information for long periods, unlike traditional RNNs.

The core component in any LSTM cell is called “cell state”, which maintains information from previous time steps. Addition or deletion to the cell state is controlled by three main gates (forget gate, input gate, and output gate). The input of an LSTM cell is a combination of the input from the current time step and the previous hidden state. This combination outputs a numerical vector whose values are squished between 0 and 1 after applying a sigmoid function as shown in (1). Values closer to 0 will be forgotten while values closer to 1 will be kept and this is called the forget gate.

$$F_t = s\ (\ W_f\ .\ [h_{t-1}\ ,\ x_t] + b_f\ ) \tag{1}$$

The same combination of input will be copied to be an input for two different activation functions (sigmoid and tanh) which are the main components of the input layer. The output of the sigmoid function is a vector whose values are squished between 0 and 1 by using (2) while the output of the tanh function is a vector whose values are squished between -1 and 1 by using (3). A pointwise multiplication is conducted between the output of these two activation functions which outputs a candidate cell state represented in a vector C_t after filtering non-important information using the sigmoid function.

$$i_t = s\ (\ W_i \cdot [h_{t-1}\ ,\ x_t] + b_i\ ) \tag{2}$$

$$C_t = \tanh\ (\ W_c \cdot [h_{t-1}\ ,\ x_t] + b_c\ ) \tag{3}$$

To calculate the updated cell state, (4) is used which represents a pointwise addition between two parts, the first part is the result of a pointwise multiplication between (the previous cell state and the output of forget gate) while the second part is the result of input gate.

$$C_t = f_t * C_{t-1} + i_t * C_t \tag{4}$$

The final step is to determine the new hidden state, and this is the output gate. To calculate the new hidden state h_t, (6) is used which includes two main parts, the first one is the output of a sigmoid activation function O_t, (5) that accepts a combined input from both the previous hidden state and the current input while the second part is the output of a tanh activation function whose input is the newly updated cell state C_t. As a result, the final output h_t of the LSTM cell will be filtered values from the cell state C_t.

$$O_t = s\ (\ Wo\ .\ [h_{t-1}\ ,\ x_t] + b_0\ ) \tag{5}$$

$$h_t = O_t * \tanh\ (C_t) \tag{6}$$

Bi-LSTM is also a type of deep learning model that deals with sequential data. It is an extension of LSTM, and it accepts input data from both directions one from a forward direction and the other from a backward direction. Working in both directions can increase the contextual understanding of the user-generated text. In the following example, the word “احمد” in the first sentence is a noun (a person whose name is Ahmed) but in the second one, it is a verb (means thank). The model can understand the context by working in both directions to decide the meaning of each word in the context based on the next and previous words. In this way, Bi-LSTM can help more in a deep understanding of the user-generated text to get the emotional insights of the tweeter.

اعلان مجدي يعقوب للقلب ❤️ دنيا سمير غانم واحمد حلمي 😻 بجد ابدعته الاعلان حلو اووي

دائماً ارفع راسي واحمد ربنا أنى من شعب مصر العظيم.

3) Bi-GRU model: GRU or Gated Recurrent Unit is a newer version of the LSTM neural network, and it was introduced by [22]. Unlike LSTM, GRU has fewer steps because it has only two gates (update and reset gates). Also, it has no cell state and the role of maintaining information lies in the hidden state. Update gate acts like forget and update gates in LSTM cell i.e., it decides what information to maintain and what to drop. In most cases, the results of Bi-GRU are slightly faster and better than Bi-LSTM.

D. Testing Layer

After finishing the processing layer, a fine-tuned version of MARBERT is available besides a trained version of Bi-LSTM and Bi-GRU. The SemEval-2018-Ar-Ec test dataset is used to test and evaluate the models. The output of the testing phase is three prediction files ready to enter the ensemble layer.

E. Ensemble Layer

After each model is tested using the SemEval-2018-EC-Ar-test dataset, three prediction files are generated, combined, and processed using a weighted sum equation that balances contextual understanding according to the performance of each model.

F. Classification Layer

Fraction results are generated from the ensemble layer. To get correct values, a certain threshold had been used to determine which values are one “1” meaning class label is found or zero “0” meaning class label is not found.

IV. Empirical Results and Discussion

In this section, the results of the proposed ensemble approach are discussed.

A. Dataset

The Arabic dataset has eleven class labels (anger – anticipation – disgust – fear – joy – love - optimism – pessimism – sadness – surprise - trust). For every class label in a tweet there is one of the two binary classification numbers (zero or one) indicating the feature is found or not. “Zero” means that the emotion is not found while “One” means that it is found. As shown in Table II every tweet is classified into (zero or one) across one or more classes which represent the emotional state of the tweeter. Tweets available in the SemEval2018-Ar-Ec dataset [9] were collected using Twitter API[3] and they focused in their searching queries on the tweets related to some emotional words also they have used Best-Worst Scaling (BWS) to determine the annotation reliability. SemEval-2018 Dataset is available for free download from the official site of competition[4]. The dataset is divided into three main files (train, development, and test). Tweets’ count in each file is shown in Table III.

TABLE II. SEMEVAL-2018-AR-EC DATASET DESCRIPTION

Multilabel ***['anger', 'anticipation', 'disgust', 'fear', 'joy', 'love', 'optimism', 'pessimism', 'sadness', 'surprise', 'trust']***	**Tweet**
[0 – 0 – 0 – 1 – 0 – 0 – 0 – 0 – 0 – 0 – 0] [Fear]	مومعقول اللي قاعد يصير فيني هالايام يارب ماينتابني شعور الخوف والتوتر اللهم التركيز وأعلى الدرجات
[1 – 0 – 0 – 0 – 0 – 0 – 0 – 0 – 1 – 0 – 0] [Anger, Sadness]	احتاج افرغ غضبي على احد بس محد له ذنب فلذلك اتطرق للانعزال
[0 – 0 – 0 – 0 – 1 – 1 – 1 – 0 – 0 – 0 – 0] [Joy, Love, Optimism]	كل عام وانت بخير وعيد سعيد وحياة مليئة بالافراح والمسرات ان شاء الله
[0 – 0 – 0 – 0 – 1 – 0 – 1 – 0 – 0 – 1 – 0] [Joy, Optimism, Surprise]	وفـ لحظه واحده .. تتدخل ارادة ربنا و تحل كل حاجه .. الصبر ! 💚

TABLE III. NUMBER OF TWEETS IN DATASET FILES

File Name	**Number of Arabic tweets**
Train	2,278
Development	585
Test (gold labels)	1,518

B. Tools

This work has been implemented on a cloud-based environment “Google Colab” [5] owned by Google. Colab offers three different plans (Free, Colab Pro, and Colab Pro+) that have differences in RAM, GPUs, storage capacities, and other features. The free plan that provides [12.69 GB of RAM, Python3 Google Compute Engine Backend (GPU), 78.19 GB for Disk Storage] has been utilized. Libraries from “Huggingface” [6] for working with transformers were utilized. Also, the “simple transformers” library was used for implementing the transformer model.

C. Evaluation Metrics

For the evaluation [7] of the proposed ensemble model, different evaluation metrics were utilized:

$$\text{Jaccard Accuracy} = \frac{1}{|T|} \sum_{t \in T} \frac{|G_t \cap P_t|}{|G_t \cup P_t|} \tag{7}$$

$$\text{Micro-P} = \frac{\sum_{e \in E} \textit{number of tweets correctly assigned to emotion class } e}{\sum_{e \in E} \textit{number of tweets assigned to emotion class } e} \tag{8}$$

$$\text{Micro-R} = \frac{\sum_{e \in E} \textit{number of tweets correctly assigned to emotion class } e}{\sum_{e \in E} \textit{number of tweets in emotion class } e} \tag{9}$$

$$\text{Micro-avg F} = \frac{2 \times Micro\text{-}P \times Micro\text{-}R}{Micro\text{-}P + Micro\text{-}R} \tag{10}$$

$$\text{Precision } (P_e) = \frac{\textit{number of tweets correctly assigned to emotion class } e}{\textit{number of tweets assigned to emotion class } e} \tag{11}$$

$$\text{Recall } (R_e) = \frac{\textit{number of tweets correctly assigned to emotion class } e}{\textit{number of tweets in emotion class } e} \tag{12}$$

$$F_e = \frac{2 \times Pe \times Re}{Pe + Re} \tag{13}$$

$$\text{Macro-avg F} = \frac{1}{|E|} \sum_{e \in E} Fe \tag{14}$$

The values of True Positives (TP) and True Negatives (TN) are the correct predictions of the classifier while False (FP) and False Negatives (FN) are the mis-predicted values. And the target is to minimize FP and FN.

D. Choosing Best Word Embedding Model

Results of Bi-GRU and Bi-LSTM have been tracked when using two different word embedding models Fasttext [8] and Aravec[9]. It was found that when applying Aravec, better results are achieved than Fasttext. A comparison between Aravec and Fasttext results is shown in Table IV.

E. Models

1) Bi-LSTM deep learning model: The experiments applied using the BiLSTM model were made after defining the parameters shown in Table V.

TABLE IV. A COMPARISON BETWEEN THE RESULTS OF BIGRU AND BILSTM WHEN USING FASTTEXT AND ARAVEC

Algorithm	**Evaluation Metrics**				
	Jaccard Score	***Precision***	***Recall***	***F1 Score***	
				Macro	***Micro***
BiGRU *Fasttext/Aravec*	0.472 / **0.498**	0.434 / **0.599**	0.532 / **0.543**	0.477 / **0.503**	0.642 / **0.664**
BiLSTM *Fasttext/Aravec*	0.455 / **0.485**	0.417 / **0.522**	0.549 / **0.559**	0.469 / **0.509**	0.624 / **0.653**

[3] https://developer.twitter.com/en/docs/twitter-api
[4] https://competitions.codalab.org/competitions/17751
[5] https://colab.research.google.com/
[6] https://huggingface.co/

[7] https://competitions.codalab.org/competitions/17751#learn_the_details-evaluation
[8] https://fasttext.cc/
[9] https://github.com/bakrianoo/aravec

TABLE V. BI-LSTM MODEL PARAMETERS

Parameter	Value
Cell type	LSTM
Number of cells	2
bidirectional	True
Cell units	256
Max timesteps	64
Batch size	128
Embedding dimensions	300
Learning Rate (LR)	3e-4
No. of epochs	20
Number of classes	11

During the training and validation phases, the model monitors and saves the best checkpoints for different validation metrics like accuracy, precision, recall, and loss. The weights of each best checkpoint have been loaded and made our predictions using the test dataset.

As shown in Table VI, the best recall checkpoint achieved better results than other checkpoints with 0.485, 0.522, 0.559, 0.509, and 0.653 for Jaccard score, precision, recall, and F1 macro & micro scores respectively. Fig. 3, 4, and 5 show the relationship between training and validation for loss, accuracy, and recall at each epoch.

TABLE VI. BEST RESULTS OF BI-LSTM MODEL

Best Metrices	Evaluation Metrics				
	Jaccard Score	*Precision*	*Recall*	*F1 Score*	
				Macro	*Micro*
BiLSTM_best_precision	0.319	0.210	0.384	0.257	0.483
BiLSTM_best_loss	0.481	0.503	0.512	0.480	0.649
BiLSTM_best_accuracy	0.415	0.400	0.445	0.393	0.586
BiLSTM_best_recall	**0.485**	**0.522**	**0.559**	**0.509**	**0.653**

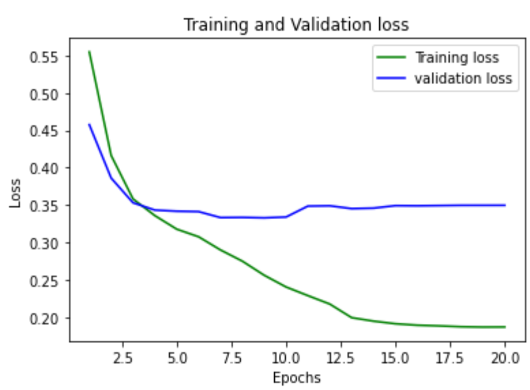


Fig. 3. BiLSTM Training and Validation Loss.

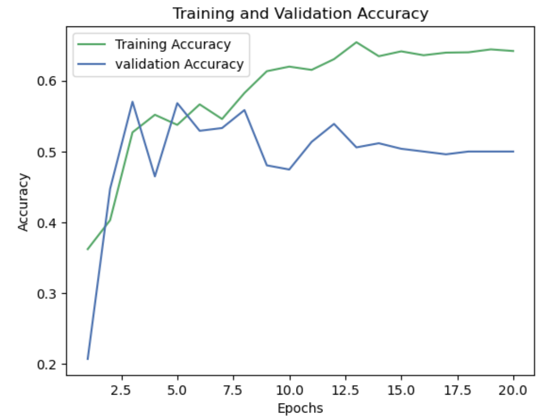


Fig. 4. BiLSTM Training and Validation Accuracy.

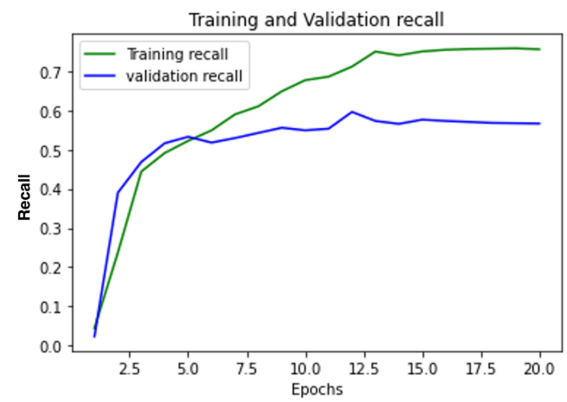


Fig. 5. BiLSTM Training and Validation Recall.

2) Bi-GRU deep learning model: The same work was done using Bi-GRU model with the parameters shown Table VII and the results of best checkpoints were compared in Table VIII.

As shown in the comparison, the best recall checkpoint achieved the best prediction results compared to other checkpoints with 0.498, 0.599, 0.503, 0.664 for Jaccard accuracy, precision, macro & micro F1 score, respectively. Fig. 6, 7, and 8 show the relationship between training and validation for loss, accuracy, and recall at each epoch.

TABLE VII. BI-GRU MODEL PARAMETERS

Parameter	Value
Cell type	GRU
Number of cells	2
bidirectional	True
Cell units	256
Max timesteps	64
Batch size	128
Embedding dimensions	300
Learning Rate (LR)	3e-4
No. of epochs	20
Number of classes	11

TABLE VIII. BEST RESULTS OF BI-GRU MODEL

Best Metrices	Evaluation Metrics				
	Jaccard Score	***Precision***	***Recall***	***F1 Score***	
				Macro	***Micro***
BiGRU_best_ precision	0.267	0.248	0.269	0.223	0.422
BiGRU_best_ loss	0.498	0.552	**0.545**	0.500	**0.664**
BiGRU_best_ accuracy	0.431	0.395	0.475	0.426	0.602
BiGRU_best_ recall	**0.498**	**0.599**	0.543	**0.503**	**0.664**

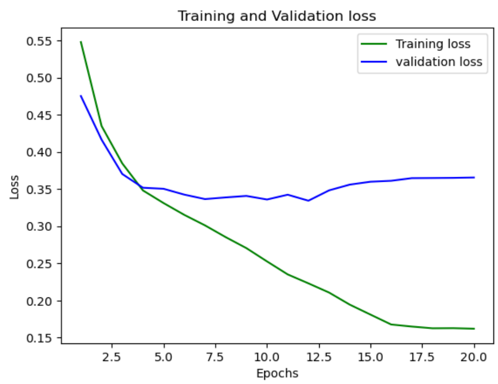


Fig. 6. Bi-GRU Training and Validation Loss.

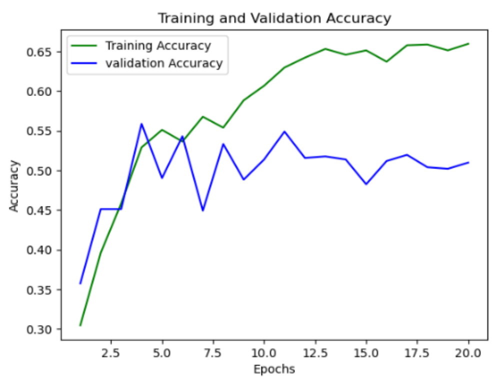


Fig. 7. Bi-GRU Training and Validation Accuracy.

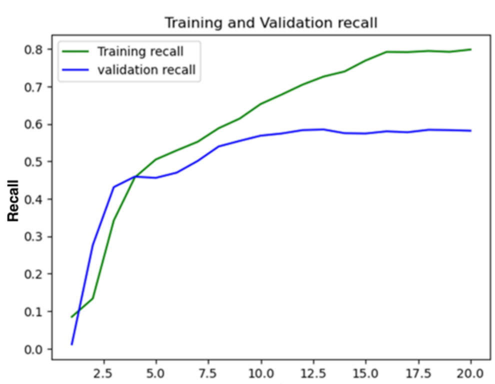


Fig. 8. Bi-GRU Training and Validation Recall.

3) MARBERT deep learning model: A pretrained language model "MARBERT" based on BERT was utilized. In [13] authors mentioned that the MARBERT transformer outperforms the recently used ARABERT transformer presented by [4] through their experiments using different datasets. They did not use SemEval-2018 E-c dataset for the emotion classification task. A comparison between the results of both ARABERT and MARBERT against the SemEval-2018-Ar-Ec dataset has been made and the results showed that MARBERT outperforms ARABERT by 4%, 2.7%, 4.2%, and 3.5% regarding Jaccard accuracy, recall, F1 macro, and F1 micro score respectively as shown in Table IX.

TABLE IX. COMPARISON BETWEEN MARBERT AND ARABERT AGAINST SEMEVAL-2018-EC-AR DATASET

Model	Evaluation Metrics				
	Jaccard Score	***Precision***	***Recall***	***F1 Score***	
				Macro	***Micro***
ARABERT	0.489	**0.545**	0.532	0.487	0.657
MARBERT	**0.529**	0.542	**0.559**	**0.529**	**0.692**

In fact, the outperformance is because MARBERT was pretrained on different tasks one of them is related to emotion recognition and it was pretrained on both MSA and Arabic dialects unlike ARABERT as mentioned in Section III.

4) Ensemble model: Because each model in the proposed ensemble model has its contextual understanding mechanism, there is a margin difference between the results of MARBERT transformer and the other two models. As a result, a weighted sum equation has been used to ensemble the results of all models i.e., predictions of each model are multiplied by weights according to the percentage of their understanding of the context as shown in (15). If the prediction of a single label is greater than or equal to a certain threshold, the predicted label will equal to "one" or the emotion is found otherwise the emotion is not found, and the predicted label will equal to "zero". To determine the best threshold and weights a manual grid search was made and gave the best threshold equals 0.34 and the best weights are w1 = 0.72, w2 = 0.1, and w3 = 0.18 for MARBERT, BiLSTM, and BiGRU respectively. The idea of manual grid search is that different weight values ranging from 0.01 to 1 are tested with different thresholds ranging from 0.01 to 1 to find the best result for the proposed ensemble model.

$$P^{total} = (p1 * w1 + p2 * w2 + p3 * w3) \quad (15)$$

where *p1*, *p2*, and *p3* are the predictions of MARBERT transformer, BiLSTM, and BiGRU respectively. P^{total} is the total result of the ensemble model after combining the results of the three models.

A comparison between the results of the closely related studies and the proposed ensemble model is shown in table X where it was found that the ensemble model has achieved the best results regarding Jaccard accuracy, and F1 macro score with an improvement ranging from 0.2% to 4.2% in accuracy, and from 5.3% to 23.3% in macro F1 score.

Table XI presents a comparison between the proposed ensemble model and each of the separate models that constitute the ensemble model. The table indicates that the best performance is for the MARBERT transformer model, while the ensemble approach combining all models still has a better effect on the overall performance compared to other models individually with an accuracy of 0.540 and a macro F1 score of 0.701.

TABLE X. CLOSELY RELATED WORKS AND THEIR RESULTS

Reference	***Publication***	***Year***	***Problem***	***Dataset***	***Methods***	***Performance Measures***
A context integrated model for multi-label emotion detection [16]	Elsevier	2018	Arabic Multilabel Emotions Classification	SemEval-2018 task 1-Ec-Ar	C-GRU (Context-aware GRU)	Mic F1: 0.495 Macro F1: 0.648 Jaccard Acc: 0.532
Hybrid Feature Model for Emotion Recognition in Arabic Text [17]	IEEE Access	2020	Arabic Multilabel Emotions Classification	SemEval-2018 task 1-Ec-Ar	HEF + DF Hybrid of human-engineered feature-based model + deep feature-based (DF) model	Micro F1: 0.631 Macro F1: 0.502 Jaccard Acc: 0.512
Pre-Training BERT on Arabic Tweets: Practical Considerations [14]	arXiv	2021	Arabic Multilabel Emotions Classification	SemEval-2018 task 1-Ec-Ar	QARiB Model	Macro F1: 0.468
Combining Context-aware Embeddings and an Attentional Deep Learning Model for Arabic Affect Analysis on Twitter [2]	IEEE Access	2021	Arabic Multilabel Emotions Classification	SemEval-2018 task 1-Ec-Ar	AraBERT word embeddings, attention-based LSTM and BiLSTM	Accuracy: 0.538
Deep Learning for emotion analysis in Arabic tweets [12]	Journal of big data	2021	Arabic Multilabel Emotions Classification	SemEval-2018 task 1-Ec-Ar	Bi-LSTM AraVec / CBOW	Micro F1: 0.61 Precision: 0.695 Recall: 0.551 Jaccard Acc: 0.498
Proposed Ensemble Model	-	2022	Arabic Multilabel Emotions Classification	SemEval-2018 task 1-Ec-Ar	MARBERT, Bi-LSTM, Bi-GRU	Accuracy: **0.540** Macro F1 Score: **0.701** Precision: 0.634 Recall: 0.550 Micro F1 Score: 0.527

TABLE XI. COMPARISON BETWEEN MARBERT, BI-LSTM, BI-GRU AND OUR PROPOSED ENSEMBLE MODEL

Model	**Evaluation Metrics**				
	Jaccard Score	***Precision***	***Recall***	***F1 Score***	
				Micro	***Macro***
MARBERT	0.529	0.542	**0.559**	**0.529**	0.692
Bi-LSTM	0.485	0.522	0.559	0.509	0.653
Bi-GRU	0.498	0.599	0.543	0.503	0.664
Proposed Ensemble Model	**0.540**	**0.634**	0.550	0.527	**0.701**

V. Conclusion and Future Work

ED can help communities in different domains, especially for social networks which have many signed users sharing their feelings and thoughts. Understanding the context of user-generated Arabic text still has a lot of challenges because of the language complexity, the limited Arabic resources, and different available Arabic dialects. Recent research studies have used pre-trained language models to overcome the issue of limited resources. In this paper, a pretrained language model (MARBERT) is fine-tuned using the SemEval-2018-task1-Ar-Ec dataset that was published in SemEval-2018-task1: Affect in Tweets to perform a multilabel classification task. Three state-of-the-art models (BiLSTM, BiGRU, and MARBERT) were ensembled and compared to recently published studies. The experimental results showed that the proposed ensemble model outperforms the best existing related work with an improvement ranging from 0.2% to 4.2% in accuracy, and from 5.3% to 23.3% in macro F1 score.

Also, it was noticed that the SemEval-2018 dataset we are using in this paper is not balanced. Three classes (anticipation, surprise, and trust) have low instances in the dataset making the dataset imbalanced. So, different data augmentation and oversampling techniques can be applied to solve this issue in future studies.

References

[1] O. AlZoubi, S. K. Tawalbeh, and M. AL-Smadi, “Affect detection from arabic tweets using ensemble and deep learning techniques,” *Journal of King Saud University - Computer and Information Sciences*, Oct. 2020, doi: 10.1016/j.jksuci.2020.09.013.

[2] H. Elfaik and E. H. Nfaoui, “Combining Context-Aware Embeddings and an Attentional Deep Learning Model for Arabic Affect Analysis on Twitter,” *IEEE Access*, vol. 9, pp. 111214–111230, 2021, doi: 10.1109/ACCESS.2021.3102087.

[3] Z. Liu, A. Xu, Y. Guo, J. U. Mahmud, H. Liu, and R. Akkiraju, “Seemo,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Apr. 2018, pp. 1–12. doi: 10.1145/3173574.3173938.

[4] W. Antoun, F. Baly, and H. Hajj, “AraBERT: Transformer-based Model for Arabic Language Understanding,” Feb. 2020, [Online]. Available: http://arxiv.org/abs/2003.00104.

[5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” Oct. 2018, [Online]. Available: http://arxiv.org/abs/1810.04805.

[6] N. Alswaidan and M. E. B. Menai, “A survey of state-of-the-art approaches for emotion recognition in text,” *Knowledge and Information Systems*, vol. 62, no. 8, pp. 2937–2987, Aug. 2020, doi: 10.1007/s10115-020-01449-0.

[7] F. A. Acheampong, C. Wenyu, and H. Nunoo-Mensah, “Text-based emotion detection: Advances, challenges, and opportunities,” *Engineering Reports*, vol. 2, no. 7, Jul. 2020, doi: 10.1002/eng2.12189.

[8] D. William and D. Suhartono, “Text-based Depression Detection on Social Media Posts: A Systematic Literature Review,” *Procedia Computer Science*, vol. 179, pp. 582–589, 2021, doi: 10.1016/j.procs.2021.01.043.

[9] S. Mohammad, F. Bravo-Marquez, M. Salameh, and S. Kiritchenko, “SemEval-2018 Task 1: Affect in Tweets,” in *Proceedings of The 12th International Workshop on Semantic Evaluation*, 2018, pp. 1–17. doi: 10.18653/v1/S18-1001.

[10] G. Badaro *et al.*, “EMA at SemEval-2018 Task 1: Emotion Mining for Arabic,” in *Proceedings of The 12th International Workshop on Semantic Evaluation*, 2018, pp. 236–244. doi: 10.18653/v1/S18-1036.

[11] M. Baali and N. Ghneim, “Emotion analysis of Arabic tweets using deep learning approach,” *Journal of Big Data*, vol. 6, no. 1, p. 89, Dec. 2019, doi: 10.1186/s40537-019-0252-x.

[12] E. A. H. Khalil, E. M. F. E. Houby, and H. K. Mohamed, “Deep learning for emotion analysis in Arabic tweets,” *Journal of Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00523-w.

[13] M. Abdul-Mageed, A. Elmadany, and E. M. B. Nagoudi, “ARBERT & MARBERT: Deep Bidirectional Transformers for Arabic,” Dec. 2020, [Online]. Available: http://arxiv.org/abs/2101.01785.

[14] A. Abdelali, S. Hassan, H. Mubarak, K. Darwish, and Y. Samih, “Pre-Training BERT on Arabic Tweets: Practical Considerations,” Feb. 2021, [Online]. Available: http://arxiv.org/abs/2102.10684.

[15] N. Al-Twairesh, “The Evolution of Language Models Applied to Emotion Analysis of Arabic Tweets,” *Information*, vol. 12, no. 2, p. 84, Feb. 2021, doi: 10.3390/info12020084.

[16] A. E. Samy, S. R. El-Beltagy, and E. Hassanien, “A Context Integrated Model for Multi-label Emotion Detection,” *Procedia Computer Science*, vol. 142, pp. 61–71, 2018, doi: 10.1016/j.procs.2018.10.461.

[17] N. Alswaidan and M. E. B. Menai, “Hybrid Feature Model for Emotion Recognition in Arabic Text,” *IEEE Access*, vol. 8, pp. 37843–37854, 2020, doi: 10.1109/ACCESS.2020.2975906.

[18] B. Talafha *et al.*, “Multi-Dialect Arabic BERT for Country-Level Dialect Identification,” Jul. 2020, [Online]. Available: http://arxiv.org/abs/2007.05612.

[19] A. Safaya, M. Abdullatif, and D. Yuret, “KUISAIL at SemEval-2020 Task 12: BERT-CNN for Offensive Speech Identification in Social Media,” in *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, 2020, pp. 2054–2059. doi: 10.18653/v1/2020.semeval-1.271.

[20] A. B. Soliman, K. Eissa, and S. R. El-Beltagy, “AraVec: A set of Arabic Word Embedding Models for use in Arabic NLP,” *Procedia Computer Science*, vol. 117, pp. 256–265, 2017, doi: 10.1016/j.procs.2017.10.117.

[21] A. Vaswani *et al.*, “Attention Is All You Need,” Jun. 2017, [Online]. Available: http://arxiv.org/abs/1706.03762.

[22] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[23] K. Cho *et al.*, “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation,” Jun. 2014, [Online]. Available: http://arxiv.org/abs/1406.1078.

Chatbots for the Detection of Covid-19: A Systematic Review of the Literature

Antony Albites-Tapia[1], Javier Gamboa-Cruzado[2], Junior Almeyda-Ortiz[3], Alberto Moreno Lázaro[4]
Facultad de Ingeniería Industrial y de Sistemas, Universidad Nacional Federico Villarreal, Lima, Perú[1, 2, 3]
Facultad de Medicina Humana, Universidad Privada Antenor Orrego, Trujillo, Perú[4]

***Abstract*—At present, the development of chatbots is one of the key activities for the diagnosis of Covid-19. The aim is to understand how these chatbots operate in the health area to make a respective diagnosis. The purpose of the research is to determine the state of the art on the use of chatbots and its impact on Covid-19 diagnosis during the last 2 years. The data sources that have been consulted are IEEE Xplore, Taylor & Francis Online, ProQuest, World Wide Science, Science Direct, Microsoft Academic, Google Scholar, ACM Digital Library, Wiley Online Library, and ETHzurich. The search strategy identified 5701 papers, of which 101 papers were selected through 8 selection criteria and 7 quality assessments. This review presents discussions regarding the methodologies used for chatbot development, i.e., what are the purposes and impact of using chatbots for Covid-19 diagnosis. In addition, this is presents the results of how important the development and implementation of chatbots are in the area of health in the face of this pandemic.**

Keywords*—*Covid-19 diagnosis; chatbot; NLP; digital assistants; health; systematic review

I. Introduction

The world is facing an unprecedented health crisis caused by the new coronavirus disease (Covid-19). Therefore, people are forced to go to health centers to rule out the disease, causing long queues. As a global pandemic, COVID-19 represents a problem in the care of patients in hospitals [107]. For this reason, many people initially turn to various sources on the Internet for self-diagnosis of their health before doing so with trained medical staff. The risk units of the Emergency Operations Center (EOC) attend online requests related to Covid-19 [108]. Direct telephone lines established by governments have collapsed due to the large volume of calls, generating long waiting times [109]. These health-related problematic situations are being solved with the help of Chatbots [110]. Thus, virtual agent applications are the latest inventions of the digital age. These applications are well known as conversational agents that run through programming or a type of artificial intelligence (AI) interaction between users and machines, with the intervention of natural language processing (NLP). Chatbots have been referred to as potentially the most promising and advanced form of human-machine interactions. It is therefore clear that chatbots play an important role in the process of Covid-19 diagnosis. These virtual agents (chatbots) are becoming involved in major global sectors such as health, banking, education, and agriculture. This paper focuses on the diagnosis of Covid-19 using chatbots. Likewise, the research is unique because a detailed review of the methodologies used for the development of chatbots has been carried out, what are the purposes and the impact of using these virtual agents, and 101 high-impact articles have also been rigorously reviewed. Also, because the research questions are original and have not been raised by any other researcher. The document is organized as follows. The present research is to close the gap concerning the lack of knowledge of the progress of research related to Covid-19 detection using chatbots. Section II describes the background this study and previous related research. Section III reviews the methodology applied. Section IV presents a systematic review of the results. Finally, Section V presents some concluding remarks and areas of future research.

II. Background and Related Studies

This section presents the background regarding the use of chatbots and discusses previous studies conducting systematic reviews of similar research.

A. Chatbots

A chatbot is a computer program through which it is possible to have a conversation, ask the user for information, or perform an action. This technology has helped to streamline processes in different areas, giving a faster response to users.

Several studies [1], [50] have emphasized the results provided by chatbots and have agreed on the rapid response they provide to users.

Based on previous studies, we identified the characteristics that chatbots provide, as well as their different areas of application, including their use in Covid-19 diagnosis and its impact. In this review, we thoroughly search and explore their application in the health field and more specifically, their role in the Covid-19 pandemic.

B. Previous Systematic Reviews

In recent months, reviews have been found on the use of AI for the diagnosis of Covid-19. A study was conducted on the use of AI for innovation in pandemic solutions [104]. Another study mapped the use of new technologies such as chatbots by health care organizations to improve mental health [103].

The authors investigated the use of various AI technologies for use in different solutions and the impacts they generate in times of pandemic. This shows the importance of AI in tackling the pandemic but does not mention much about the role of chatbots in diagnosing Covid-19. This review is then extended to the use of chatbots for the diagnosis of Covid-19 during this pandemic.

III. Methodology of Review

The review method used in this study was developed by considering the guidelines of B. Kitchenham [102] for conducting a systematic literature review (SLR). The review method elaborates the research questions, data sources, search procedure, exclusion criteria, quality assessment, data extraction, and data synthesis.

A. Research Problems

Table I shows the research objectives of this study through the correspondence between each research question and each motivation.

TABLE I. Research Objectives

Research Questions	**Motivation**
RQ1: What methodologies are used for the development of chatbots diagnosing Covid-19?	Identify the methodologies to be used for the development of a chatbot.
RQ2: In what areas are chatbots mostly being applied?	Identify the areas where chatbots are being used.
Q3: What are the purposes of using a chatbot for Covid-19 diagnosis?	Identify the purposes of using chatbots for diagnosing Covid-19.
Q4: What are the most important tools for developing a chatbot?	Identify the most important tools for developing a chatbot.
Q5: What are the most widely used programming languages in chatbot development?	Identify the most widely used programming languages in chatbot development.

B. Search Sources and Search Strategy

The source of the referenced data includes prominent digital libraries such as Taylor & Francis Online, ProQuest, World Wide Science, Science Direct, IEEE Xplore, Microsoft Academic, Google Scholar, ACM Digital Library, Wiley Online Library, and ETHzurich. The search strategy includes searches by keywords relevant to our study. The search has been focused on research questions and commonly used terms corresponding to machine learning and estimations of the software effort.

The search procedure has been carried out using search threads written as (chatbot) AND (covid or covid-19 or covid-19 diagnosis) where the set of Ai represents the keywords related to the automatic learning methods, i.e., Ai chatbot} and Bi represents the keywords related to the dependent variables, i.e., Bi {covid, covid-19, covid-19 diagnosis}. The results of the search procedure have been reduced to 5701 items that are discussed in more details below (see Table II).

C. Selection Criteria

Exclusion criteria have been defined to accurately assess the quality of the available literature. Papers have been reviewed and discussed by the authors for the exclusion decision with the industry advisor. The papers were checked against the given criteria:

EC1: The papers do not mention a methodology, model, or method.

EC2: The papers are not less than 2 years old.

EC3: The papers are not written in English.

EC4: They are scientific papers.

EC5: Titles and keywords are inappropriate.

EC6: The papers are not related to the topic.

EC7: It contains few citations.

EC8: The summary of papers is irrelevant.

The result of this stage shows 101 items.

D. Studio Selection

At the outset, 5701 results were obtained. When applying criteria 1 and 2, 2914 items were excluded. Likewise, when applying criteria 3 and 4, 1369 papers were excluded. Furthermore, when applying criteria 5 and 6, 565 papers were excluded. Finally, when applying criteria 7 and 8, 752 papers were excluded. This results in a total of 101 items (see Fig. 1).

TABLE II. Search Sources and Search Strategies

Source	**Search equations**
IEEE Xplore	(((“Full Text & Metadata”: chatbot) AND “Full Text & Metadata”: covid-19 or covid-19 diagnosis) AND “Full Text & Metadata”: method or methodology or model)
Taylor & Francis Online	[All: chatbot] AND [All: covid-19 OR covid-19 diagnosis] AND [All: method OR methodology OR model]
ProQuest	(chatbot) AND (covid-19 diagnosis OR covid-19 diagnosis) AND (model OR method OR methodology)
World Wide Science	chatbot and (covid-19 OR covid-19 diagnosis) and (model or Methodology or method)
Science Direct	chatbot and (covid-19 OR covid-19 diagnosis) AND (model OR method OR methodology)
Microsoft Academic	(chatbot) AND (covid-19 OR covid-19 diagnosis) AND (model OR method OR methodology)
Google Scholar	chatbot AND (covid-19 OR covid-19 diagnosis) AND (model OR method OR methodology)
ACM Digital Library	(chatbot) AND (covid-19 OR covid-19 diagnosis) AND (model OR method OR methodology)
Wiley Online Library	(chatbot) AND (covid-19 OR covid-19 diagnosis) AND (model OR method OR methodology)
ETHzurich	(chatbot) AND (covid-19 OR covid-19 diagnosis) AND (model OR method OR methodology)

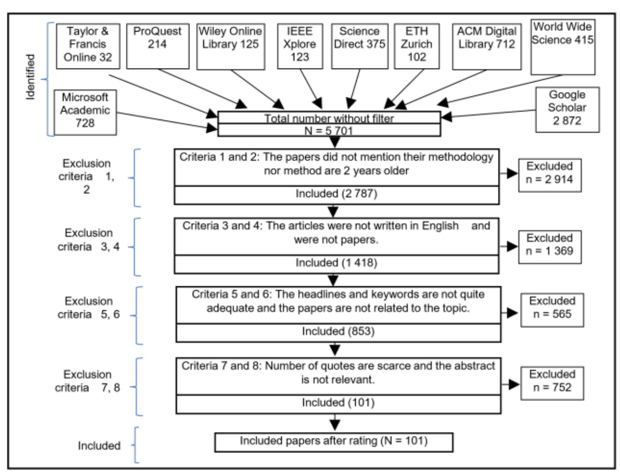


Fig. 1. Study Selection.

The figure shows how the exclusion criteria were applied and the result of this.

E. Quality Evaluation

Each of the 101 papers that remained after the selection criteria were evaluated independently using the following criteria.

QA1: Is the document well organized?

QA2: Does the topic belong to the fields of computer science, information systems, or management?

QA3: Are the objectives of the research clearly specified in the document?

QA4: Does the paper belong to a book, publication, or conference?

QA5: Are the results of the experiments performed clearly identified and reported?

QA6: Is the full text of the document available?

QA7: In general, is the document considered useful?

The result of this stage showed 101 papers, of which no papers were discarded because all met the quality criteria.

F. Data Extraction Strategy

During this stage, we extracted data from each of the 101 papers included in this systematic review according to 10 important data properties. The 10 properties considered in the extraction are Source, Year, Language, Research Areas, Type of Publication, Research Methodology, Authors, Number of Citations, Abstract, and Keywords.

Fig. 2 below shows the report of the studies collected in Mendeley to achieve a better control of the papers and facilitate the extraction of information.

G. Data Synthesis

The process of data synthesis includes gathering the data and finalizing the answers according to the research questions. The data synthesis has been achieved by analyzing the selected studies through multivariate regression techniques for a meta-analysis, as well as the following descriptive statistical measures: mean, mode, median, sum, count, and percentage. Tables, bar charts, frequency polygons, histograms, pie charts, sector diagrams, donut charts, scatter plots, georeferenced maps, and keyword clouds were also used.

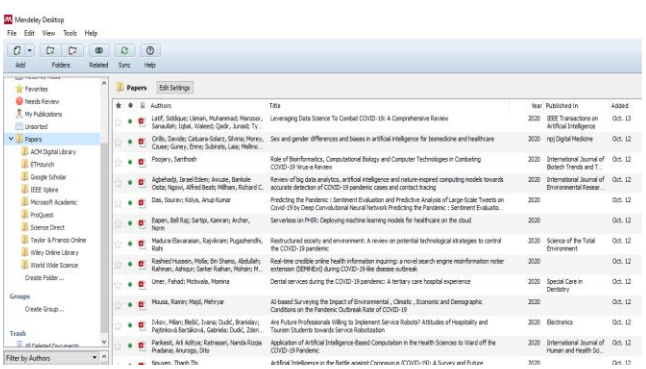


Fig. 2. Mendeley Paper Report.

IV. Result and Discussion

A. Overview of Studies

The study selection process conducted through the SLR resulted in 101 papers for data extraction and analysis from the most recognized sources, i.e., Google Scholar, ProQuest, Science Direct, Taylor & Francis Online, IEEE Xplore, and Microsoft Academic (Fig. 3). Fig. 4 shows the most recurrent words in the abstracts of the papers, through a word cloud and the number of repetitions of the most frequent terms.

It can be seen that the most frequently mentioned terms are Covid, health, pandemic, and digital. This means that, during the pandemic, the digital era has been of significant help in the area of health to combat Covid-19. With the use of machine learning, an analysis of publication results was executed, by country, institutions, keywords and citation counts [112].

In addition, Fig. 5 shows the number of papers per country of publication of this study.

It should be noted that the US is the country that has done the most research on Covid-19. The authors have reported that China has published 97 papers (42%) [111].

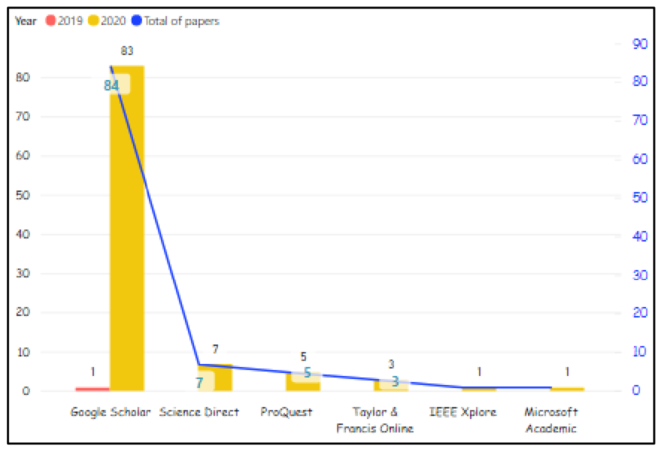


Fig. 3. Most Recognized Data Source.

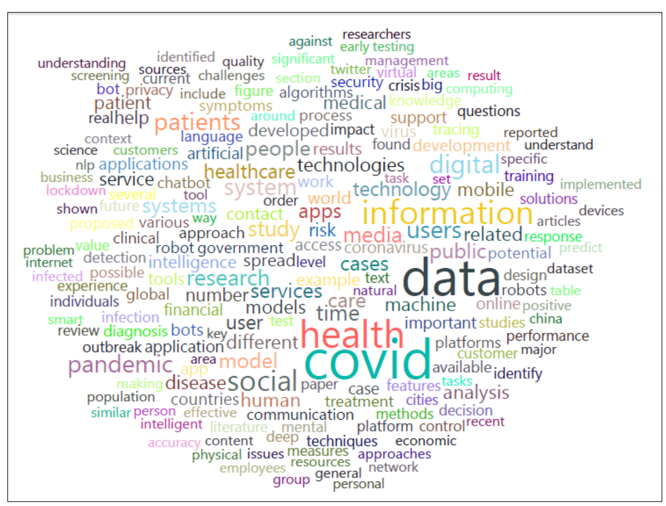


Fig. 4. Word Cloud in the Abstracts of Papers.

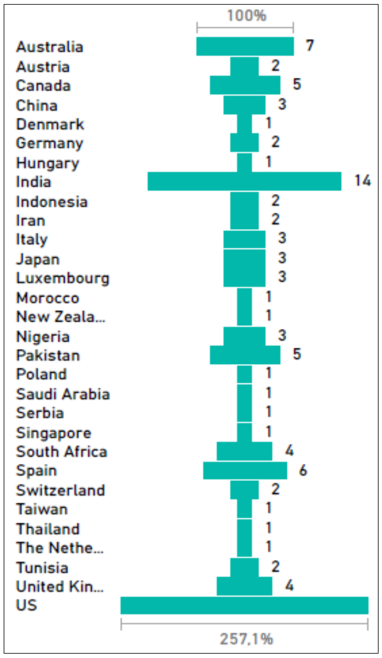


Fig. 5. Numbers of Papers by Country of Publication.

B. *Answers to Research Questions*

RQ1: What methodologies are being used for development?

The systematic review of the literature has resulted in 101 relevant papers. Each paper contains its own methodology for the development of chatbots used to diagnose Covid-19. We detected four different types of methodologies in the development of such chatbots. To provide a clearer view of our response, Table III was developed to answer the question more clearly.

The most widely used methodology for the development of chatbots is "sentence order prediction," which is used to predict the sentences a user may write and use the information to prepare a more effective response. The second methodology is "mask language modelling," which is used to predict the blank spaces in texts, which is necessary for the treatment of users. "Next sentence prediction" is being used because of the need to have a good user experience by speeding up the responses of the chatbots. The last methodology found is "learning based," which is a method based on continuous learning, where a chatbot learns based on constant conversations with the user.

TABLE III. METHODOLOGIES APPLIED FOR THE DEVELOPMENT OF CHATBOTS

Applied methodologies	**Papers**	**Qty. (%)**
Sentence Order Prediction	[2] [3] [7] [8] [3] [17] [21] [27] [39] [42] [48] [58] [65] [68] [71] [81] [86] [94] [99]	19 (34.5)
Mask Language Modelling	[1] [4] [8] [3] [6] [14] [20] [24] [45] [52] [62] [89] [90] [100]	14 (25.5)
Next Sentence Prediction	[13] [4] [8] [9] [14] [19] [24] [37] [45] [59] [68] [76] [91] [99]	14 (25.5)
Learning Based	[3] [7] [31] [61] [69] [76] [80] [89]	8 (14.5)

The use of different methodologies provides excellent results, and depending on the chatbots to be developed, several methodologies can even be implemented together [104].

RQ2: In what areas are chatbots being mostly applied?

A literature review demonstrated that there were five most important areas where chatbots have been applied. Table IV shows the areas where chatbots are being used.

TABLE IV. AREAS WHERE CHATBOTS ARE APPLIED

Application Areas	**Papers**	**Qty. (%)**
Health	[1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [17] [18[19] [20] [22] [23] [24] [25] [26] [27] [28] [29] [30] [31] [32] [33] [35] [36] [38] [39] [40] [41] [42] [43] [44] [45] [46] [47] [48] [49] [50] [51] [53] [54] [55] [56] [57] [58] [61] [62] [63] [64] [65] [66] [67] [68] [69] [70] [71] [72] [73] [74] [75] [76] [77] [78] [79] [81] [82] [83] [84] [85] [86] [87] [88] [90] [91] [92] [93] [94] [95] [96] [97] [98] [99] [100] [101]	93 (57.7)
Education	[1] [5] [7] [9] [12] [16] [19] [22] [26] [27] [30] [32] [34] [36] [37] [40] [42] [46] [51] [53] [54] [55] [57] [58] [77] [79] [81] [82] [84] [85] [89] [90] [98] [99]	34 (21.5)
Retail	[10] [12] [16] [33] [39] [43] [55] [60] [65] [59] [86] [89] [90] [97]	14 (8.8)
Banks	[12] [16] [27] [39] [55] [59] [72] [73] [80] [86] [90]	11 (8)
Tourism	[4] [22] [43] [59] [60] [65] [69] [77] [94]	9 (4)

One of the areas where chatbots are most applied is the health sector owing to its importance in human health and the impact it generates during the present pandemic. Another area where chatbots are applied is in the education sector where, through an agent that is able to respond to user needs 24 h a day, it is beneficial for different studies. In the banking and retail sectors, its application is increasing owing to the flexibility of chatbots to meet different business needs. Finally, in the tourism sector, the need to count on an agent that solves the frequent questions of the users and that is available at all times has led to the development of chatbots.

According to Alaa Ali Abd-Alrazaq [103], health is one of the areas where chatbots are most frequently applied, and the effectiveness and safety of chatbots in improving mental health has been measured.

RQ3 - What are the purposes of using chatbots for the diagnosis of covid-19?

As a result of a systematic review of the literature, we found three main purposes for using chatbots for the diagnosis of Covid-19. Table V shows the purposes of using chatbots for a Covid-19 diagnosis.

The most important purpose of using chatbots for Covid-19 is a high availability, in other words, to have a service that is always available to all users. User experience is one of the main reasons why we use chatbot for a Covid-19 diagnosis, which allows for a much faster, smoother, and better results. Agility is another extremely relevant reason because having a space where one can obtain much more agile and personalized answers generates a positive impact for users.

According to Adam S. Miner [104], in line with our results in which we determined that a high availability is the main purpose for using a chatbot, the possibility to offer recommendations and diagnosis at any time has been observed.

Q4 - What are the most important tools for developing a chatbot?

When reviewing the relevant literature, it was observed that there are five important tools for the development of chatbots, which are the most widely used in the scientific field of AI. Table VI shows the most important tools for chatbot development.

The first and most widely used tool for developing a chatbot is DialogFlow, which is capable of understanding natural language and provides the infrastructure for creating conversations and building dialogue to interact with the user in a fluid manner. The second tool is Watson, which allows you to create, train, and implement conversational interactions in any application, device, or channel. Another tool used is Language Understanding, which is used for automatic learning that allows users to understand natural language and extract its meaning and understand what aspects are required. The latest tool found is AWS, in which previously trained AWS AI services provide ready-to-use intelligence in applications and workflows. AI services are easily integrated with applications to address common use cases, such as creating custom recommendations, upgrading a call center, improving security, and increasing customer involvement.

TABLE V. PURPOSES FOR USING A CHATBOT FOR COVID-19 DIAGNOSIS

Purpose of using chatbot	Papers	Qty. (%)
Availability|	[1] [2] [8] [9] [10] [12] [16] [22] [25] [26] [27] [37] [39] [42] [44] [46] [49] [51] [52] [53] [55] [57] [65] [69] [70] [72] [73] [74] [77] [81] [91] [92] [97] [98]	34 (68)
User experience	[2] [4] [28] [39] [43] [45] [55] [71] [97]	9 (18)
Agility	[40] [54] [65] [79] [84] [85] [89]	7 (14)

TABLE VI. MOST IMPORTANT TOOLS FOR CHATBOT DEVELOPMENT

Tools for chatbot development	Papers	Qty. (%)
Amazon Web Services (AWS)	[9] [24] [35] [36] [37] [43] [46] [67] [69] [73] [75]	11 (52.38)
Language Understanding	[8] [20] [21] [25] [51] [74]	6 (28.57)
Watson	[34] [98]	2 (9.52)
Dialog Flow	[51] [100]	2 (9.52)

According to Pavel Smutny [105], the most important tools for the development of chatbots are DialogFlow together with Watson and Language Understanding, which are capable of understanding natural language for establishing fluid dialogues with users.

Q5 - What are the most widely used languages in the development of a chatbot?

Through the results of the SLR, we found the four most widely used programming languages in the development of chatbots, as listed in Table VII.

TABLE VII. MOST WIDELY USED PROGRAMMING LANGUAGES IN CHATBOT DEVELOPMENT

Most widely used programming languages	Papers	Qty. (%)
Python	[5] [6] [7] [8] [15] [20] [23] [28] [29] [48] [52] [58] [71] [76] [95]	15 (68.18)
Java	[22] [28] [71] [77] [99]	5 (22.73)
JavaScript	[28] [71]	2 (9.09)
PHP		0 (0)

The most widely used programming language in the development of chatbots is Python because it is one of the preferred languages used on the server side to create a back-end; it also has many interesting features, among them, an interpreter and a high-level object-oriented programming language. The second most widely used programming language found in our research is JavaScript, which is ideal for handling high user traffic applications and events where hundreds of thousands of interactions are to be sent every second. Java is another programming language that is widely used to create applications and processes in a great number of devices and allows executing the same program in different operating systems. The last programming language is PHP, which is used to develop web applications, favoring the connection between the servers and the user interface. This programming language is widely used because it is open source, which means that anyone can make changes to its structure.

According to Eleni Adamopoulou [106], these programming languages are the most recognized in the development of chatbots because they are open source and effective for high user traffic applications. In addition, these languages support all functionalities and guarantee quick and easy access to information and application services for users.

V. CONCLUSION AND FUTURE RESEARCH

This study has managed to identify the impacts and purposes of Covid-19 diagnosis through chatbots, the areas where chatbots are most frequently applied, the methods used for their development, and the tools and languages needed for their implementation. The 101 papers reviewed for this analysis were chosen based on exclusion criteria and a quality evaluation. The results obtained indicate that the most widely used methodology for the development of chatbots is next sentence prediction. The area where chatbots are mostly used is in retail owing to the high availability they offer to users. The most important purpose for using chatbots in the diagnosis of covid-19 is the agility of the response it provides to patients. The biggest impact generated is the satisfaction of the patients is by obtaining a fast response in their diagnoses. The most important tool for the development of a chatbot was found to be Watson Assistant owing to its easy integration with different communication channels. In addition, the most widely used language for the development of chatbots is Python owing to its easy integration and implementation. This review has certain limitations owing to the novelty of the topic; however, thanks to the motivation of the researchers regarding the topic because it is at a global level, a large number of papers were found to achieve a quality SLR.

The systematic review of the literature identified a new trend in the diagnosis of Covid-19, not only in the use of chatbot, but also on other technologies related to artificial intelligence, such as automatic and deep learning, which can contribute more efficiently to the diagnosis of Covid-19.

REFERENCES

[1] S. Poojary, «Role of Bioinformatics, Computational Biology and Computer Technologies in Combating COVID-19 Virus-a Review,» Journal, p. 5, 2020.

[2] A. Michal Podpora, «Human Interaction Smart Subsystem—Extending Speech-Based Human-Robot Interaction Systems with an Implementation of External Smart Sensors,» Journal, p. 16, 2020.

[3] N. Arli Aditya Parikesit, «Application of Artificial Intelligence-Based Computation in the Health Sciences to Ward off the COVID-19 Pandemic,» Journal, p. 9, 2020.

[4] I. Milan Ivkov, «Are Future Professionals Willing to Implement Service Robots? Attitudes of Hospitality and Tourism Students towards Service Robotization,» Journal, p. 16, 2020.

[5] C. Oladapo Oyebode, «COVID-19 Pandemic: Identifying Key Issues using Social Media and Natural Language Processing,» Journal, p. 12, 2020.

[6] A. Nourchène Ouerhani, «Towards a mobile conversational agent for COVID-19 post quarantine psychological assistance,» Journal, p. 8, 2020.

[7] Y. David Oniani, «A Qualitative Evaluation of Language Models on Automatic Question-Answering for COVID-19,» Journal, p. 8, 2020.

[8] M. Martin Müller, «COVID-TWITTER-BERT: A NATURAL LANGUAGE PROCESSING MODEL TO ANALYSE COVID-19 CONTENT ON TWITTER,» Journal, p. 6, 2020.

[9] R. María Consuelo Sáiz-Manzanares, «Effectiveness of Using Voice Assistants in Learning: A Study at the Time of COVID-19,» Journal, p. 20, 2020.

[10] T. Nguyen, «Artificial Intelligence in the Battle against Coronavirus (COVID-19): A Survey and Future Research Directions,» Journal, p. 14, 2020.

[11] R. Maleeha Naseem, «Exploring the Potential of Artificial Intelligence and Machine Learning to Combat COVID-19 and Existing Opportunities for LMIC: A Scoping Review,» Journal, p. 11, 2020.

[12] H. Nachit, «Digital Transformation in Times of Covid-19 Pandemic: The Case of Morocco,» Journal, p. 17, 2020.

[13] P. Bernhard Knapp, «Diagnostic Accuracy of Web-Based COVID-19 Symptom Checkers: Comparison Study,» Journal, p. 8, 2020.

[14] M. MOUSA, «AI-based Surveying the Impact of Environmental, Climatic, Economic and Demographic Conditions on the Pandemic Outbreak Rate of COVID-19,» Journal, p. 9, 2020.

[15] P. Justinas Mišeikis, «Lio-A Personal Robot Assistant for Human-Robot Interaction and Care Applications,» Journal, p. 8, 2020.

[16] D. Mhlanga, «Industry 4.0 in Finance: The Impact of Artificial Intelligence (AI) on Digital Financial Inclusion,» Journal, p. 14, 2020.

[17] J. Alistair Martin, «An artificial intelligence-based first-line defence against COVID-19: digitally screening citizens for risks via a chatbot,» Journal, p. 6, 2020.

[18] A. Ali Ahmad Malik, «Man, machine and work in a digital twin setup: a case study,» Journal, p. 25, 2020.

[19] R. Rajvikram Madurai Elavarasan a, «Restructured society and environment: A review on potential technological strategies to control the COVID-19 pandemic,» Journal, p. 18, 2020.

[20] A. Nourch`ene Ouerhani, «Smart Ubiquitous Chatbot for COVID-19 Assistance with Deep learning Sentiment Analysis Model during and after quarantine,» Journal, p. 9, 2020.

[21] J. Toby Jia-Jun Li, «Multi-Modal Repairs of Conversational Breakdowns in Task-Oriented Dialogs,» Journal, p. 14, 2020.

[22] B. Israel Edem Agbehadji, «Review of Big Data Analytics , Artificial Intelligence and Nature-Inspired Computing Models towards Accurate Detection of COVID-19 Pandemic Cases and Contact Tracing,» Journal, p. 16, 2020.

[23] V. Ahmed Al-Rawi, «Bots as Active News Promoters: A Digital Analysis of COVID-19 Tweets,» Journal, p. 13, 2020.

[24] J. Muhammad Ajmal Azad, «A First Look at Privacy Analysis of COVID-19 Contact Tracing Mobile Applications,» Journal, p. 11, 2020.

[25] M. Oliver Baclic1, «Challenges and opportunities for public health made possible by advances in natural language processing,» Journal, p. 8, 2020.

[26] L. Bickman, «Improving Mental Health Services: A 50 Year Journey from Randomized Experiments to Artificial Intelligence and Precision Mental Health,» Journal, p. 49, 2020.

[27] S. A, «Digital Orthopaedics: A Glimpse Into the Future in the Midst of a Pandemic,» Journal, p. 6, 2020.

[28] L. Sagnick Biswas1, «GO-COVID: AN INTERACTIVE CROSS-PLATFORM BASED DASHBOARD FOR REAL-TIME TRACKING OF COVID-19 USING DATA ANALYTICS,» Journal, p. 15, 2020.

[29] A. Bstract, «WHAT TYPES OF COVID-19 CONSPIRACIES ARE POPULATED BY TWITTER BOTS?,» Journal, p. 25, 2020.

[30] D. Subhash Chandir, «Impact of COVID-19 pandemic response on uptake of routine immunizations in Sindh, Pakistan: An analysis of provincial electronic immunization registry data,» Journal, p. 10, 2020.

[31] D. Sarat Kr. Chettri, «Leveraging Digital Tools and Technologies to Alleviate COVID-19 Pandemic,» Journal, p. 10, 2020.

[32] O. Agim Eliezer Chukwuyere, «Librarians' Use of Social Media in Disseminating Health Information on COVID-19,» Journal, p. 12, 2020.

[33] D. Cirillo, «Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare,» Journal, p. 11, 2020.

[34] D. Christoph Gross, «"CAir", a Telemonitoring and Hybrid Virtual Coaching Solution for Patients Suffering from Chronic Obstructive Pulmonary Disease: Protocol for a Randomized Controlled Trial,» Journal, p. 24, 2020.

[35] M. Guy Fagherazzi, M. Catherine Goetzinger, M. Mohammed Ally Rashid, M. P. Gloria A Aguayo y M. P. Laetitia Huiart, «Digital Health Strategies to Fight COVID-19 Worldwide: Challenges, Recommendations, and a Call for Papers,» Journal, p. 10, 2020.

[36] D. Julia M. Puaschunder, «The Future of Healthcare around the World: Four indices integrating Technology, Productivity, AntiCorruption, Healthcare and Market Financialization1,» Journal, p. 23, 2020.

[37] S. Joshua D. Fisher, «LiveHint: Intelligent Digital Support for Analog Learning Experiences,» Journal, p. 10, 2020.

[38] E. Davide Golinelli, «How the COVID-19 pandemic is favoring the adoption of digital technologies in healthcare: a rapid literature review,» Journal, p. 12, 2020.

[39] M. Marwah Hassounah, M. F. M. Hafsa Raheel y M. S. M. F. Mohammed Alhefzi, «Digital Response During the COVID-19 Pandemic in Saudi Arabia,» Journal, p. 14, 2020.

[40] H. Oh, «Utilizing Bots for Sustainable News Business: Understanding Users' Perspectives of News Bots in the Age of Social Media,» Journal, p. 16, 2020.

[41] C. I Ching Hsu, «Integrating machine learning and open data into social Chatbot for filtering information rumor,» Journal, p. 15, 2020.

[42] M. Hussein, «Real-time credible online health information inquiring: a novel search engine misinformation notier extension (SEMiNExt) during COVID-19-like disease outbreak,» Journal, p. 23, 2020.

[43] H. Kevin Jenkinsab, «Digital Platforms in Aotearoa,» Journal, p. 15, 2020.

[44] S. Jiang, «Usable Security for ML Systems in Mental Health: A Framework,» Journal, p. 9, 2020.

[45] A. Timothy J. Judson, «Implementation of a digital chatbot to screen health system employees during the COVID-19 pandemic,» Journal, p. 6, 2020.

[46] M. Ali Khaleghi, «New Ways to Manage Pandemics: Using Technologies in the Era of COVID-19, a Narrative Review,» Journal, p. 7, 2020.

[47] M. Pathak Laxmi Kumari, «Artificial Intelligence: Changing the scenario of COVID-19,» Journal, p. 8, 2020.

[48] A. Thomas Kuruvilla, «Computerized Healthcare System Embedded with Machine Learning,» Journal, p. 6, 2020.

[49] A. Lydia W. Li, «Digital health for patients with chronic pain during the COVID-19 pandemic,» Journal, p. 4, 2020.

[50] E. Iacopo Pozzana, «Measuring Bot and Human Behavioral Dynamics,» Journal, p. 11, 2020.

[51] D. Urmil Bharti, «Medbot: Conversational Artificial Intelligence Powered Chatbot for Delivering Tele-Health after COVID-19,» Journal, p. 6, 2020.

[52] A. Khuzani, «COVID-Classifier: An automated machine learning model to assist in the diagnosis of COVID-19 infection in chest x-ray images.,» Journal, p. 12, 2020.

[53] A. Andreu, «Digitalisation and COVID-19: The Perfect Storm,» Journal, p. 24, 2020.

[54] T. Yigitcanlar, «Contributions and risks of artificial intelligence (AI) in building smarter cities: Insights from a systematic review of the literature,» Journal, p. 34, 2020.

[55] D. Mhlanga, «Industry 4.0 in finance: the impact of artificial intelligence (ai) on digital financial inclusion,» Journal, p. 14, 2020.

[56] D. Yoneoka, «Early SNS-based monitoring system for the covid-19 outbreak in Japan: A population-level observational study,» Journal, p. 9, 2020.

[57] G. Czifra, «COVID-19 AND INDUSTRY 4.0,» Journal, p. 10, 2020.

[58] L. Sekhar, «The Future of Skull Base Surgery: A View Through Tinted Glasses,» Journal, 2020.

[59] C. Prentice, «Engaging and retaining customers with AI and employee service,» Journal, p. 13, 2020.

[60] K. Li, «How Should We Understand the Digital Economy in Asia? Critical Assessment and Research Agenda,» Journal, p. 41, 2020.

[61] M. Tayarani-N., «Applications of Artificial Intelligence in Battling Against Covid-19: A Literature Review,» Journal, p. 61, 2020.

[62] R. Madurai Elavarasan, «Restructured society and environment: A review on potential technological strategies to control the COVID-19 pandemic,» Journal, p. 18, 2020.

[63] S. Nomura, «An assessment of self-reported COVID-19 related symptoms of 227,898 users of a social networking service in Japan: Has the regional risk changed after the declaration of the state of emergency?,» Journal, p. 11, 2020.

[64] C. Sohrabi, «World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19),» Journal, p. 6, 2020.

[65] D. Zhang, «From Networking to Mitigation: The Role of Social Media and Analytics in Combating the COVID-19 Pandemic,» Journal, p. 6, 2020.

[66] M. Ali, «Digital Technologies Applications in the Provision of Library and Information Services in Health Crises,» Journal, p. 11, 2020.

[67] P. Henman, «Improving public services using artificial intelligence: possibilities, pitfalls, governance,» Journal, p. 14, 2020.

[68] Y. Zheng, «A Learning-Based Model to Evaluate Hospitalization Priority in COVID-19 Pandemics,» Journal, p. 10, 2020.

[69] A. EZE, «ICT Solutions And R&D Based on Big Data Analytics in The Fight Against COVID-19 Pandemic: African Innovations and Opportunities,» Journal, p. 23, 2020.

[70] D. Subhash Chandir, «Impact of COVID-19 pandemic response on uptake of routine immunizations in Sindh, Pakistan: An analysis of provincial electronic immunization registry data,» Journal, p. 10, 2020.

[71] L. Sagnick Biswas1, «GO-COVID: AN INTERACTIVE CROSS-PLATFORM BASED DASHBOARD FOR REAL-TIME TRACKING OF COVID-19 USING DATA ANALYTICS,» Journal, p. 15, 2020.

[72] M. Stefano A. Bini, «Digital Orthopaedics: A Glimpse Into the Future in the Midst of a Pandemic,» Journal, p. 6, 2020.

[73] L. Bickman, «Improving Mental Health Services: A 50 Year Journey from Randomized Experiments to Artificial Intelligence and Precision Mental Health,» Journal, p. 49, 2020.

[74] M. Oliver Baclic, «Challenges and opportunities for public health made possible by advances in natural language processing,» Journal, p. 8, 2020.

[75] J. Muhammad Ajmal Azad, «A First Look at Privacy Analysis of COVID-19 Contact Tracing Mobile Applications,» Journal, p. 11, 2020.

[76] V. Ahmed Al-Rawi, «Bots as Active News Promoters: A Digital Analysis of COVID-19 Tweets,» Journal, p. 13, 2020.

[77] B. Israel Edem Agbehadji, «Review of Big Data Analytics, Artificial Intelligence and Nature-Inspired Computing Models towards Accurate Detection of COVID-19 Pandemic Cases and Contact Tracing,» Journal, p. 16, 2020.

[78] E. Ian A Scot, «Can AI help in the fight against COVID-19?,» Journal, p. 5, 2020.

[79] L. Tan Yigitcanlar, «Can Building "Artificially Intelligent Cities" Safeguard Humanity from Natural Disasters, Pandemics, and Other Catastrophes? An Urban Scholar's Perspective,» Journal, p. 20, 2020.

[80] D. Sarat Kr. Chettri, «Leveraging Digital Tools and Technologies to Alleviate COVID-19 Pandemic,» Journal, p. 10, 2020.

[81] O. Agim Eliezer Chukwuyere, «Librarians' Use of Social Media in Disseminating Health Information on COVID-19,» Journal, p. 12, 2020.

[82] E. Davide Golinelli1, «Adoption of Digital Technologies in Health Care During the COVID-19 Pandemic: Systematic Review of Early Scientific Literature,» Journal, p. 23, 2020.

[83] M. ComprehensiveReview, «Managing COVID-19 Global Pandemic With Managing COVID-19 Global Pandemic WithHigh-Tech Consumer Wearables: A ComprehensiveReview,» Journal, p. 7, 2020.

[84] L. Tan Yigitcanlar, «Can Building "Artificially Intelligent Cities" Safeguard Humanity from Natural Disasters, Pandemics, and Other Catastrophes? An Urban Scholar's Perspective,» Journal, p. 20, 2020.

[85] G. Sambasivarao Yaragalla1, «COVID-19 Pandemic: A Comprehensive Updated Review with an Artificial Intelligence (AI),» Journal, p. 10, 2020.

[86] O. Shahid, «Machine Learning Research Towards Combating COVID-19: Virus Detection, Spread Prevention, and Medical Assistance,» Journal, p. 17, 2020.

[87] M. Jedrek Wosik, «Telehealth transformation: COVID-19 and the rise of virtual care,» Journal, p. 6, 2020.

[88] S. Maria Vogiatzaki, «Enhancing City Sustainability through Smart Technologies: A Framework for Automatic Pre-Emptive Action to Promote Safety and Security Using Lighting and ICT-Based Surveillance,» Journal, p. 20, 2020.

[89] A. Surabhi Vermaa, «Investigating the emerging COVID-19 research trends in the field of business and management: A bibliometric analysis approach,» Journal, p. 9, 2020.

[90] J. Veen, «Job criteria for successful use of AI,» Journal, p. 12, 2020.

[91] M. Siddique Latif, «Leveraging Data Science To Combat COVID-19: A Comprehensive Review,» Journal, p. 21, 2020.

[92] J. Michael Y. Uohara, «The Essential Role of Technology in the Public Health Battle Against COVID-19,» Journal, p. 7, 2020.

[93] S. Yuta Tanouea, «Mental health of family, friends, and co-workers of COVID-19 patients in Japan,» Journal, p. 3, 2020.

[94] S. Mohammad Haider Syed, «An Artificial Intelligence and NLP based Islamic FinTech Model Combining Zakat and Qardh-Al-Hasan for

Countering the Adverse Impact of COVID 19 on SMEs and Individuals,» Journal, p. 14, 2020.

[95] S. Raihan Hamid Suraperwata, «Language Modeling for Journalistic Robot based on Generative Pretrained Transformer 2,» Journal, p. 6, 2020.

[96] S. Nitya Singhala, «A FIGHT AGAINST COVID-19: MAJOR IT TRENDS,» Journal, p. 6, 2020.

[97] M. Iqbal H. Sarke, «Mobile Data Science and Intelligent Apps: Concepts, AI-Based Modeling and Research Directions,» Journal, p. 19, 2020.

[98] S. Sohiki Sarbadhikari, «The Global Experience of Digital Health Interventions in COVID 19 Management,» Journal, p. 18, 2020.

[99] J. Klein, «A First Look at Android Applications in Google Play related to Covid-19,» Journal, p. 36, 2020.

[100] P. Chonnatee Rodsawang, «Designing a Competent Chatbot to Counter the COVID-19 Pandemic and Empower Risk Communication in an Emergency Response System,» Journal, p. 7, 2020.

[101] A. Vishali Gupta, «Evolving consensus on managing vitreo retina and uvea practice in post-COVID 19 pandemic era,» Journal, p. 12, 2020.

[102] Kitchenham, «Guidelines for performing Systematic Literature Reviews in Software Engineering,» p. 65, 2007.

[103] A. Alaa Ali Abd-Alrazaq, «Effectiveness and Safety of Using Chatbots to Improve Mental Health: Systematic Review and Meta-Analysis,» Journal, p. 17, 2020.

[104] L. Adam S. Miner, «Chatbots in the fight against the COVID-19 pandemic,» Journal, p. 4, 2020.

[105] P. Pavel Smutny, «Chatbots for learning: A review of educational chatbots for the Facebook Messenger,» Journal, p. 11, 2020.

[106] E. Moussiades, «An Overview of Chatbot Technology,» Journal, p. 11, 2020.

[107] G. Battineni, «AI Chatbot Design during an Epidemic Like the Novel Coronavirus» Journal, p. 9, 2020.

[108] Ch. Rodsawang, «Designing a Competent Chatbot to Counter the COVID-19 Pandemic and Empower Risk Communication in an Emergency Response System» Journal, p. 7, 2020.

[109] A. Martin, «An artificial intelligence-based first-line defence against COVID-19: digitally screening citizens for risks via a chatbot» Journal, p. 7, 2020.

[110] W. Erazo, «Chatbot Implementation to Collect Data on Possible COVID-19 Cases and Release the Pressure on the Primary Health Care System» IEEE, p. 6, 2021.

[111] L. Wynants, «Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal» BMJ, p. 16, 2020.

[112] F. De Felice, « Coronavirus disease (covid-19): A machine learning bibliometric analysis» NIH, p. 5, 2020.

An Enhanced Genetic Algorithm (EGA)-based Multi-Hop Path for Energy Efficient in Wireless Sensor Network (WSN)

Battina Srinuvasu Kumar[1], S.G. Santhi[2], S. Narayana[3]
Research Scholar, Department of Computer Science & Engineering, Annamalai University-608002, India[1]
Associate Professor, Department of Computer Science & Engineering, Annamalai University-608002, India[2]
Professor, Department of Computer Science & Engineering, Gudlavalleru Engineering College, Gudlavalleru-521356, India[3]

***Abstract*—Wireless Sensor Networks (WSNs) encounter a number of issues in terms of performance. In WSN, the majority of the energy is utilized to transfer data from sensor nodes to a central station or hub (BS). There have therefore been many different types of routing protocols devised to help with the distribution of data in WSNs. Large-scale networks have been designed with minimal power consumption and multipurpose processing due to recent improvements in wireless communication and networking technology. For the time being, sensor energy remains a restricted resource for constructing routing protocols between sensor nodes and the base station, despite advances in energy collection technologies. For wireless sensor networks with far-flung cluster heads and base stations, direct transmission is a critical component since it impacts the network's efficiency in terms of power consumption and lifespan. A new approach for identifying an effective multi-hop routing between a source (CH) and a destination (BS) is investigated in this study in order to decrease power consumption and hence increase the life of a network (OMPFM). The suggested technique utilizes a genetic algorithm and a novel fitness metric to discover the best route. For selecting CHs and enhancing the speed and quality of created chromosomes, they suggest two pre-processes, which they call CH-selection and Chromosome Quality Improvement (CHI). The proposed method is evaluated and compared to others of its kind using MATLAB simulator. It has been found that using the proposed method outperforms other methods such as LEACH, GCA, EAERP, GAECH, and HiTSeC in terms of the first node die metric by 35%, 34%, 26%, 19% and 50%, respectively. It also outperforms other methods by 100% in terms of the last node die metric.**

Keywords*—Cluster head; energy efficient; multi-hop path; enhanced genetic algorithm; wireless sensor network (WSN)

I. Introduction

The development of micro electromechanical system technology has made compact, low-cost nodes practical in networks. The nodes are responsible for sensing, processing, and communicating wirelessly. A lot of the nodes in wireless sensor networks (WSNs) run on batteries and are connected to a base station (BS). It acts as a middleman between the end user and the other nodes[1] like figure1 represents the WSN's basic architecture. In any of the supervised areas, WSNs ensure continuous monitoring (CM) or engage in event detection. WSNs need non-rechargeable batteries to power the sensor nodes, therefore modelling an energy-efficient routing algorithm is required. We need an energy-efficient routing strategy in places where battery recharging and replacement aren't an option. Energy is a big worry and as a result, there are several heuristics, however they all take longer to come into play. When it comes to WSN energy efficiency, clustering stands out as a viable option[2].

This probability-based clustering produces CHs, which can only communicate with other nodes in their immediate vicinity. So, multichip (MH) routing helps the network to route traffic beyond the communication range, which is limited by the energy factor. Multichip routing as an alternative, delay is reduced while energy consumption is increased, therefore the routing is more efficient. As a result, the researchers are working to create a better routing protocol that uses less energy[3]. MH transmission and direct transmission are the two most common conventional routing options (DT). Nodes in DT connect with the BS directly, but intermediate nodes in MH communicate with each other via subdivided communication distances, resulting in little energy and delay. Short-hop and long-hop routing are terms used to describe MH transmissions based on the number of hops. There are flat and hierarchical algorithm routing protocols in the literature that work to reduce the energy used by sensor nodes. This can be explained by comparing flat and hierarchical routing: in flat routing the communication takes place between nodes within the same communication range, while in hierarchical routing the communication takes place between a node and its nearest clustering hierarchy. Although hierarchical routing conserves energy, it places a heavy burden on the cluster's central hubs. To communicate outside of the regular radio range, all sensor nodes clamp themselves to a CH. This has an impact on QoS and performance[4],[33].

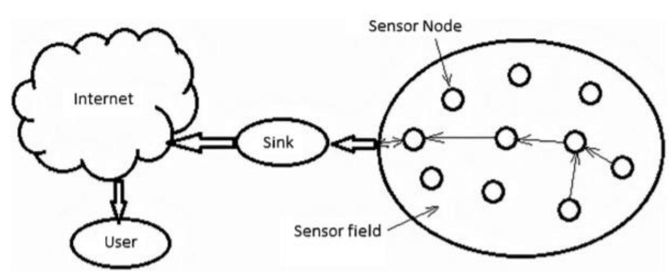


Fig. 1. The WSN's Basic Architecture.

A WSN is a network of micro-sensor-equipped nodes with a large number of spatially scattered communication infrastructures. These tiny sensors collect and interpret data from a specified area by sensing it and working together. A central station receives the data, which is subsequently processed. The relevance of WSNs in IoT-based goods and services is rising, but they have been around for a long. Things like health care, environmental monitoring, and smart cities will all need them in the future, due to the Internet of Things. [2, 3,5] It's easy to see that sensor networks have several orders of magnitude more nodes than ad hoc networks. Despite their extensive dispersion, sensor nodes have limited computing, memory, and power capabilities. The development of energy-efficient routing algorithms is a major problem in WSNs. It's impossible to delete or change sensor nodes after they've been set up in the network. Nodes' power usage and management have a strong impact on network performance. As the most energy-efficient systems in WSNs in recent years, hierarchical routing protocols have been frequently used [6]. Cluster heads operate as a go-between for the member nodes and the base station itself rather than having individual base stations for each member node [7]. Among the cluster's nodes, the CH utilises the least amount of network resources overall when gathering data, aggregating it, and transmitting it to the (BS).

It is the first hierarchically built protocol to use clustering for information routing (LEACH) [8]. When data are transported straight from the CHs to the BS, it is known as a single-hop routing protocol (SHR). It does, however, have significant drawbacks. In the first place, CHs positioned far from the BS lose more energy than those located nearby. As a result, populations in the furthest reaches of the galaxy are at greater risk of dying young. Large-scale networks can't employ LEACH's direct transmission method since it disrupts the distribution's load balancing and reduces the network's overall operating lifespan. It also depends on the relevance of the detected information at each point in time how often data is transferred between nodes. Due to the difference in activity, more active nodes die sooner than less active ones, causing an energy imbalance in the network[9].

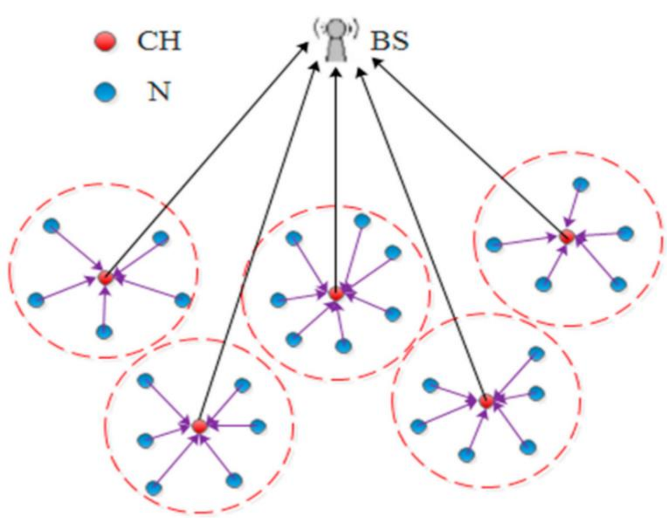


Fig. 2. The Wireless Sensor Network Architecture uses the Low Energy Adaptive Clustering Hierarchy (LEACH) Protocol.

Transmitting sensing data to the BS for analysis is critical in WSNs since this data helps the network meet its objectives. Making judgments that are in keeping with the network's aims is made easier with this strategy. With direct transmission, the LEACH protocol drains CHs quickly, especially if the distance between two CHs delivering data to BS is large. The additional energy used to transport data in a CH will have a negative influence on the network's overall lifetime factor performance. By determining the optimum route from a root CH to the BS, this research presents an improved way for increasing network lifetimes while using less power than the direct transmission process. The best transmission path is found using a genetic algorithm (GA). The proposed approach maximises the network's stability period during which all nodes are still operational. It also extends the network's lifespan in the event that all nodes fail. Maximizing the fitness function and improving the selection of cluster heads are two ways to accomplish this. The following is the paper's structure: It begins with a brief history of GA, followed by a discussion of similar works in Section 2, a detailed presentation of the proposed approach in Section 3, a quantitative assessment of the proposed method in addition to previous protocols in section 4, and a conclusion in Section 5.

II. RELATED WORK

This section describes the LEACH and GA methods. Before discussing into GA optimization, let's take a look at the LEACH approach.

Wireless sensor networks have lately seen a slew of hierarchical routing methods [8,10,11–19]. To make the network last longer, several of them used multi-hop inter-cluster connection. WSN protocol LEACH [8,11] includes two phases: configuration and steady state operation. As a node is being set up, the distributed process employs a probabilistic method to assess whether or not it should be designated as a CH. Non-CH node will join the cluster based on the strength of the CH signal. The CH delivers the TDMA schedule to all cluster nodes at once to guarantee that it is received by all of them. Continuously sending one hop of data to the BS during the steady state phase is required. LEACH uses a CDMA coding to minimise inter-cluster interference. To maintain a fair distribution of burdens, the round concludes with a random rotation of CH. However, there are a number of disadvantages to using LEACH. During the CH election process, residual authority or position is not taken into consideration. Due to greater energy consumption and the death of remote CHs induced by direct data transfer from the sink node to the CH, this single-hop routing approach is unsuitable for large networks. With multi-hop communication between CHs and the BS, energy conservation and scaling were demonstrated to be higher [12, 20]. So, Zhang et al. [13] created the LEACH-WM WSN protocol, an intra-cluster multiple hop routing mechanism. In addition to the setup phase, LEACH contains an optimization phase. The BS receives data from the CH through a weight relay node during the final stage (WR). Because of its remaining energy and proximity to the backend server, this relay node is selected as a cluster member. Clusters located far away from the BS use their energy more quickly than clusters near to it, resulting in an uneven distribution of energy across the network.

Alnawaf et al. [14] employed a multi-hop approach to improve LEACH (MHT). CHs may be divided into two categories according on how far apart they are: internal and exterior. The CHs in the first group send data straight to the BS, whereas the CHs in the second group create a custom routing table to control data flow. CHs were distributed over two tiers while utilising the Improved MHT-LEACH protocol (IMHT). You can tell if you look at the distance between one CH level and the next by looking at how far one CH level is from BS. Despite the fact that these methods were developed to be as energy-efficient as possible, the majority of them solely consider the distance element when determining routes. Because data is always routed through the nodes nearest to the BS, those nodes use up their energy faster than the others. Defended routes cannot be changed for the network's whole operational life, affecting the network's ability to maintain a balanced load. Routing methods for WSNs will enhance load balancing by taking energy and distance into account when selecting the next hop. One way to increase the network's lifespan is to provide efficient data to the base station, as described by Biradar[15] et al. Each cluster has two means to connect with the others on the network, split into several clusters. Instead of using a backend server (BS), you may use a central hub (CH) to collect data from all cluster nodes and deliver it directly to the backend server instead (CHs). Second, the CH must use an inter-cluster channel to interact with both the BS and the CS, because it is located[16,17] in a remote location. When this occurs, it uses other CHs with a low hop count to deliver data to the BS. The optimal paths to the BS may be found using a multi-hop LEACH (MH-LEACH) protocol, according to Neto et al. [17]. Depending on the intensity of the received signal, each CH begins by sending an advertising message to create its routing table. However, the CH inspects the routes beforehand to make sure they won't cause a loop or send the incorrect way for individuals to go.

Wireless Sensor Networks (WSNs) may make use of EAACA, an ant colony algorithm that is energy-conscious and utilises the ACO to determine the most efficient route to the storage system's back end (sink node). When deciding on the next node, three factors are taken into account: the distance to BS, the remaining energy of the previous node, and the path's average power. Backward ants are used in the method to produce response packets in response to the legal routes built by forward ants. Using a lifespan aware routing method for wireless sensor networks, Mohajerani et al. developed a network topology where energy dissipation is uniformly spread among nodes (LTAWSN). When installing an energy-saving pheromone operator, it makes use of a phoney ant colony. When calculating the next step in the journey, a location function takes the distance to the destination into account, as well as nodes that are closer to the goal. Because they consume all of the network's energy, ACO's fat routing protocols require more energy than previous ACO versions and can quickly deplete the network's energy supply. In addition, they use two-step transmission systems, moving forward to investigate the route and returning to leave pheromones behind. As a result of increased transmission costs, nodes consume more energy.

Using the LEACH-C (LEACH-C) methodology was suggested in [14,20]. Centralized protocols have all choices made by a single node and then relayed to all sensor nodes. Some of their duties include electing the network's chief executives (CH) and supervising clusters. A well-balanced distribution of nodes inside clusters is one of the goals of the LEACH-C protocol. Lifespan is the goal of the LEACH-Deterministic Cluster Head Selection (LEACH-DCHS) technique [15,34]. As a result, two changes have been recommended. To begin, the threshold equation must be adjusted to account for the residual energy in the sensor nodes when selecting the network's CHs. As part of the second update, new network lifetime standards will be established. To better understand the Threshold-LEACH (T-LEACH) approach, [16,21] provided an overview of the basic theory. According to this protocol, the network's CHs are selected based on their energy thresholds. The CHs can only be utilised in a certain number of rounds at a time. As long as the CH's residual energy is less than the threshold energy, this situation will persist. That would necessitate the appointment of a new House Speaker. For a one-hop connection, a technique known as Unequal Clustering LEACH has been suggested by [17,22]. (U-LEACH). One hop delivers all of the CH's data directly to the BS. The old method utilises more energy in CHs placed distant from the BS because of their location. As we travel away from the BS, the U-LEACH predicts that cluster sizes will become more unequal and then drop. There were imbalanced clusters in the original LEACH protocol, thus the LEACH-B protocol was developed to address that issue. For the purpose of choosing CHs and creating balanced clusters, this protocol takes into account not just a desired percentage of cluster members, but also the residual energy those members have left over once the election process is complete. The MHT-LEACH protocol was proposed in [11, 23] as a way of reducing network transmission distances. It divides all of the CHs into two groups of equal size. For starters, there's the internal level, which includes all the CHs that aren't too far from the BS. Also included in this level are any CHs that are located outside of a certain threshold distance from the base station (BS). A suitable CH from the internal level receives data from any external CH and routes it to the BS[24,35].

A. The Leach Protocol

The LEACH protocol's goal is to reduce the amount of power used by WSNs while increasing their efficiency [14,21,25,26]. To that end, a random factor has been used in a rotation strategy for CH selection. Each round of the LEACH procedure [14,20,27] has two stages, the first of which is the steady-state set-up. Along with the CHs selection, clusters are also established in the first stage. Random numbers between 0 and 1 are generated by nodes in each cluster to choose a CH [14,28]. In order to compare the generated numbers to, we use T(n). As soon as the reference number is greater than the randomly generated number, the CH node becomes the active node for further calculations. Using Eq. (1), we can determine the value of the constant T(n).

$$T(n) = \begin{cases} \frac{P}{1-P*(r mod \frac{1}{p})} & : if n \in G \\ 0: if n \in G \end{cases} \quad (1)$$

To illustrate, consider the graph below, where G represents a group of nodes that have not yet become CHs in the preceding 1/P rounds, and P represents the proportion of those nodes that have become CHs. All nodes in the cluster receive equal amounts of energy because of this. LEACH's power consumption is balanced across all nodes because the CH role is rotated, yet the protocol has significant drawbacks regardless of this benefit. The residual energy of a node, for example, is not taken into account by LEACH when choosing a CH. Equation (1) says that a node can become a CH at any time. All nodes have the same chance of becoming CHs after a certain number of rounds. Consequently, both high- and low-energy nodes are equally likely to become the CH. Another issue is the use of single-hop communications to transmit directly from a CH to a BS. This causes energy holes to form between CH and BS, making it impossible for isolated nodes to exchange data. When this is the case, network performance suffers as a result. As a result, an improvement to LEACH is proposed in this study to improve efficiency in the CH selection process and data transmission, as provided in the proposed mechanism, to address the issues raised above.

B. Genetic Algorithm Optimization Technique

GA is a solution-finding algorithm that mimics natural evolution by employing the same mechanisms as natural selection. GA is used to identify the best global solution to optimization and search problems. In GA, the most crucial techniques are selection [17,30], crossover and mutation. Starting with an initial sample of individuals, a random population is produced and each of those individuals is assessed according to their fitness level. Individuals with the best fitness values will be chosen from the initial population, and they will take part in the following generation [29,31]. An important part of GA is the objective function (fitness), which is used to assess the quality of individuals and pick the best contributors for future generations. By mixing the children of the parents, the crossover operator creates new ones, and these new ones will have a mix of qualities from both parents [18]. The new children are subjected to random swaps in the mutation operator, resulting in new solutions. When a certain condition is met, the algorithm comes to an end [8,19,25,26].

The crossover operation is demonstrated in Table 1. Pre-crossover chromosome presentation, followed by post-crossover chromosome presentation: that's how it works. The third position is designated as the crossing point. After that, a mutation operation depending on the mutation rate will be performed on one of the children, as detailed in the next section.

Given Table 2, when the mutation occurs at position 6, gene F in first part of table is not included in the old path (A D E F) and the new path is shown in the second part. The second offspring will be subjected to the same technique, and the fitness of the two new chromosomes will be determined. When the new offspring and their parents are compared to each other, a greedy selection is carried out to determine which has the highest fitness value. Chromosome fitness values are used in the selection process for future generations. Some are chosen for future generations based on the outcomes. In order to maintain a steady population size, chromosomes with lower fitness values will be picked for future generations rather than those with higher values. When the number of generations reaches a certain value, the process comes to a conclusion.

Using a suggested OMPFM, crossover and mutation algorithms have been tweaked to prevent the formation of erroneous routes by efficiently choosing the CHs that are participating in the GA while also ensuring that they are easily available in the network. By adopting binary representation to address the OMPFM's multi-hop route problem, the CH will never be duplicated along the way. Using GA and other evolutionary algorithms to maximise network longevity, we'll look at routing protocols that have been utilised to improve the LEACH protocol going forward.

TABLE I. CROSSOVER OPERATION

Position	1	2	3	4	5	6	7
	A	B	C	D	E	F	G
Chr1	1	1	0	0	1	0	1
Chr2	1	0	0	1	0	1	1
Crossover Point↑							
Position	1	2	3	4	5	6	7
	A	B	C	D	E	F	G
Offs1	1	0	0	0	1	1	0
Offs2	1	0	0	1	1	0	1

TABLE II. MUTATION OPERATION

Position	1	2	3	4	5	6	7
	A	B	C	D	E	F	G
Offs2	1	1	0	0	1	0	1
Position	1	2	3	4	5	6	7
	A	B	C	D	E	F	G
Offs1	1	0	0	0	1	1	0

III. PROPOSED METHODOLOGY

The GA technique is used by OMPFM (optimal multi-hop path finding methodology) to discover the shortest path between two network nodes: the source CH and the BS. Because of its simplicity, the GA is utilized in this study as an optimization strategy to find the optimal solution. The sections that follow include a lot more details on this technique. With the new approach, each chromosome identifies which CH source gave rise to BS. Chromosomes can vary in length depending on how many CHs are involved, but no more than 100 CHs should be used in a single cycle. The newly suggested approach uses binary representation to represent each gene on a chromosome as a CH in the route. Routing pathways include CHs with gene values of 1 by default. Otherwise, their gene expression levels are zero, as seen in the table below. The source CH is the CH with ID 1, and there are seven CHs in total, each with an ID from 1 to 7. As shown in Table 3, there is a direct link between the chromosomal in issue and the BS. Table 3 shows the IDs of the CHs in the first row, and the path's CHs are listed in the second row. To illustrate, let's get from CH1 to the BS via the following path.

TABLE III. THE SELECTED PATH

CH1	CH2	CH3	CH4	CH5	CH6	CH7
1	1	0	0	1	0	0

CH1 → CH2 → CH5 → BS

The proposed technique reduces energy consumption while increasing network lifespan by sending data from a source CH to a destination BS. A fitness function determines the optimum course of action. With the OMPFM protocol, you may send only the data you want from one CH to another. For example, it may identify which CHs have the most energy to follow a particular path and then utilize important attributes to narrow the pool of choices. It was shown that using two pre-processing techniques, the GA optimization strategy performed better than the fitness function's proposed parameters for determining the optimum direct path from a source CH to BS. Both examine the GA's chromosomal quality and the suggested fitness function, as stated in the next subsections.

A. Filter Pre-processing

In the GA's first generation, the quality of the chromosomes used will be taken into consideration in the second revision of the formula. Before GA operations begin, the best CHs will be picked based on their energy levels via a filtering pre-process. We'll start by calculating the global average CH energy. CHs with higher energy levels will be needed to help create the second generation when the first is finished. By starting with the best chromosomes, the GA will be more efficient since it will have better chromosomes to work with. Additionally, by reducing the chromosomal length, this change enhances the GA's speed. When four CHs have more energy than normal, they will take part in the GA, and the chromosome will only be three instead of six in length (save for the source CH). The following paragraphs go into detail on how energy affects chromosomal quality. Due to optimal route discovery being an important component, it is necessary in the proposed transmission technique to take into account how many CHs there are in transmission path. There will be less average energy spent to transport data from one CH to another as the number of CHs in the transmission channel increases Eq. (2) gives the average amount of transmission-related CHs energy utilized.

$$\mathrm{Avg_{CET}} = (\textstyle\sum_{i=1}^{N} \mathrm{E_{Tx}}(i))/(N) \quad (2)$$

To convey data from one CH to another through the transmission channel, you need a certain amount of energy called $\mathrm{E_{Tx}}$, whereas $\mathrm{Avg_{CET}}$ measures the average amount and N stands for the number CH involved. $\mathrm{E_{Tx}}$ can be calculated using the following formula [26].

$$E_{Tx} = E_{elec} * k + E_{amp} * k * d^m \quad (3)$$

Were n having a value of 2 in open space and a value of 4 in enclosed places where E_{elec} is the electronic energy and E_{amp} is the amplifier energy.

The same approach is used to determine the average amount of CHs consumed during the receiving process, which may be summarised as follows:

$$Avg_{CER} = (\textstyle\sum_{i=1}^{N} E_{Rx}(i))/(N) \quad (4)$$

Avg_{CER} is equal to the average energy necessary to receive data when N CHs are present in the transmission channel, where E_{Rx} is the average energy required to receive data from each CH. The energy model is used to determine E_{Rx} [26]:

$$E_{Rx} = E_{elec} * k \quad (5)$$

As a result of the calculations, the average CHs energy is as follows:

$$Avg_{RE} = (\textstyle\sum_{i=1}^{N} E_r(i))/(N) \quad (6)$$

Avg_{RE} is the average amount of remaining energy after all CHs have been spent in the current round? At the moment, the CH has an energy level of Er. The effects of energy on data transmission to the BS are shown in Equations (2), (4), and (6), which are all found in the source code. This means that reducing participation in the GA by CHs will make the transmission process more efficient. The following section delves deeper into the fitness function that has been proposed.

With the suggested method, the major objective is to reduce the amount of energy used in the data transfer procedure from a CH to BS. Several fitness function factors are taken into consideration in order to reach this objective. One of the parameters of the fitness function is the distance from the BS to each participant CH. As the number of CHs increases, so does the fitness function. As a result, another fitness function parameter is the ratio of proposed CHs to total CHs on the path. Number of members in a cluster is a critical fitness function parameter since it impacts how much energy cluster members consume. The number of participants in each suggested CH is the final fitness function parameter. It's true that the efficiency of transmission is maximised when the distance between the source and destination is as short as possible, but this comes at the cost of increased energy consumption.

Using the fitness function below, we can minimise the fitness function's values to discover the best route:

$$F(i) = D(CH_s, BS) + \left(\frac{N_{par}}{N}\right) + NO_{ofpart} + members \quad (7)$$

With respect to which, F(i) signifies the number of homologous regions on the i^{th}, average distance between the origin CH and the BS is represented by chromosomal D (CHs, BS)., as indicated in Chromosome homologs are represented by the letters N, P, and O, and the number of parts they play in the transmission process is represented by the letters No, P, and O, as described in the section on how they participate in the process [24,32].

In the first phase of the fitness function, the energy consumption is distributed across all CHs based on the distance between the source CH and the BS. It's important to note that this second section shows what percentage (really) of participants' CHs are actually employed in the transmission process. Because the suggested fitness function is a minimization function, when the percentage is low, the path's optimality improves. The optimality diminishes in direct proportion to the value. A CH that has engaged in the transmission process more than once would soon lose its energy, thus the third section shows the total number of CHs that have taken part in the transmission process along the way. In other words, when its value increases, the fitness function will become more valuable, and an inefficient path will result. The number of nodes in each clusteris the final consideration.

This is especially crucial because of its impact on energy usage.

IV. Result and Discussion

The decision was made to use a MATLAB simulation tool and the simulation parameters in Table 4 to test the fitness function. The effectiveness of the approach was evaluated by comparing the findings of the proposed technique with those of the LEACH process [26] and other comparable methods. Several tuning processes are used to specify the crossover and mutation rates after which the GA parameters provided in this paper are listed in Table 4.

The LEACH protocol is the most frequently used clustering method in the WSN community. As a consequence, referring to it as a guide is critical. On the other hand, the proposed OMPFM is pitted against several other protocols that are developed on top of LEACH. Before comparing the proposed OMPFM with the current methods, an assessment of the efficiency of the improved GA should be conducted in terms of chromosomal sizes and running time. This will indicate that the better GA is more efficient. A single round of comparison is shown in Figures 3 and 4 to demonstrate how conventional GA (NGA) compares to improved GA (EGA). For every additional round, the chart becomes increasingly jumbled; on the other hand, for every additional round less than ten, the comparison values get increasingly muddled. There will be a total of 10 rounds. Because all nodes in the selected rounds are still alive, the running time is precise and predictable, and the ideal chromosomal sizes have been attained. The chosen rounds are also the first ones. This is followed by ten tests in which the standard GA is compared to its upgraded counterpart (see Figures 3 and 4). Time required for 10 conventional GA rounds is shown in Figure 3, whereas time required for enhanced GA is shown in Figure 2. Figure 2 depicts the average chromosomal size throughout the first ten cycles. There is only one way to get the average chromosome size from the BS to each CH, thus the total number of CHs divided by the number of CHs each round gives the average chromosome size. Figure 3 demonstrates that the enhanced GA requires less time each round than the regular GA does. To improve efficiency, we used the suggested pre-processing steps. Furthermore, as seen in Figure 4, enhanced GA shrinks the size of the chromosomes when compared to regular GA. The graph below illustrates the average chromosomal size during the experiment. The improved version of GA's smaller chromosome sizes increases the effectiveness of the GA by finding the optimal solution by utilising short chromosomes, thanks to the suggested pre-process that picks the appropriate CHs to be included in the GA.

Figure 3 depicts how much faster the improved GA runs, on average, than the standard GA. (Number of intermediate CHs) is less for each run in the improved GA than it is for each run in the standard GA (Fig. 4). In terms of running time and chromosomal size, the results show that the proposed GA's pre-processes are more efficient than the standard GA. Various scenarios were developed in order to assess the OMPFM's feasibility. The simulations were run a total of 20 times to provide a fair comparison. Each sensor node in the network has its own array data structure that is used to calculate the fitness value of the suggested strategy. To illustrate the array data structure, see Table 5.

TABLE IV. Parameters and Values used in the Simulation

Representation	Declaration	Value
A	Size of the network	100×100
BS(i,j)	Base station location	area center (50×50)
N	The total no. of nodes	100
Etx	Transmission energy	51 nJ/bit
Eo	Initial energy	0.5 Joule, 1 Joule
Eelec	Consumption of electronic resources	51 nJ/bit
Erx	Energy for reception	51 nJ/bit
Eamp	amplified signal transmission	12 pJ/bit/m2
Eda	energy used for data collection and analysis	6 nJ/bit
Rmax	Number of rounds allowed as a maximum	5000
Pack/CtrPack	Packet size for data and control	528 bytes/50 bytes
Indiv	No. of Individuals	50
Iter	No. of iterations	120
Cr	Ratio of changeover	0.9
Mr	Change rate	0.05

Fig. 3. Time Spent Running in the First Ten Rounds.

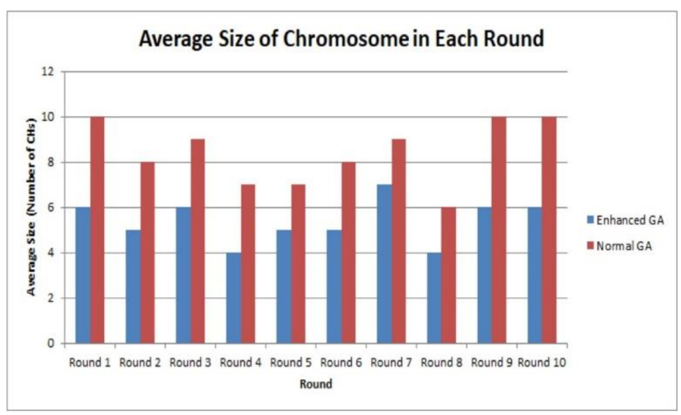


Fig. 4. The Average Chromosomal Size in the First Ten Rounds.

TABLE V. ARRAY DATA STRUCTURE

Fields	xd	yd	G	Status	No. of Parameters	Members	ToBs	Types	E
1	99.83	63.79	0	1	26	0	140.35	'C'	-0.0059
2	59.61	51.53	0	1	16	0	143.78	'C'	-0.0037
3	66.36	39.48	0	1	36	0	156.37	'C'	-0.0031
4	56.56	69.30	0	1	29	0	125.86	'N'	-9.81e-05
5	9.62	90.97	0	1	31	0	111.58	'N'	-5.33e-04
6	51.71	19.61	0	1	29	0	175.38	'N'	-1.80e-04
7	61.13	15.54	0	1	32	0	179.79	'N'	-1.62e-04
8	73.02	72.88	0	1	0	0	124.26	'A'	-0.007
9	75.32	33.67	0	1	29	0	163.30	'N'	-4.177e-04
10	85.66	75.07	0	1	46	0	125.11	'N'	-5.14e-05
11	80.81	5.71	0	1	31	0	191.78	'N'	-2.87e-04
12	44.25	6.17	0	1	28	0	188.91	'N'	-2.26e-05
13	32.27	3.22	0	0	0	0	192.60	'A'	-0.0030

Table 6 and Fig. 5 show the outcomes of using the proposed technique in comparison to the comparable methods [5,14,16,22,36]. When compared to techniques based on the number of dead nodes, the recommended strategy leads in longer network lifetimes, as shown by the findings above. The recommended approach outperforms the LEACH protocol by 28%, the GCA by 25%, the EAERP by 18%, and the GAECH by 11% in dead node percentages of 10% compared to the LEACH protocol. As compared to the standard LEACH protocol, this one outperforms it by 35%, as well as the GCA (34%), EAERP (26%), and GAECH (19% when including HND, or 50% of dead nodes). The new approach beat the previous ones by a factor of 101 percent when compared to LEACH and a factor of 99 percent when compared to GCA and EAERP. Longevity gains are seen when employing the recommended technique, and these gains can be ascribed to using energy-efficient intermediate CHs in the selection process for the fitness function.

To simulate the second case, the starting energy of each node was reduced to 0.5 Joule from 1 Joule as per [11]. Table 7 shows the FND and LND of the proposed OMPFM. Table 7 shows the average values of the FND and LND parameters over tens of thousands of runs, which are 1695 and 3946, respectively. Figure 5 depicts the [11]'s FND and LND. Table 6 and Table. 7 [11] reveal that our OMPFM has superior FND and LND values than previous techniques. A FND of 1695 characterises the suggested OMPFM, while a FND of 800 characterises the techniques outlined in [11]. The LND in the OMPFM is 3946, but in the approaches in [11], it's roughly 2000. In terms of running time, the recommended OMPFM performs better thanks to better parameter selection in the fitness function and better utilisation of the pre-processing.

TABLE VI. PERCENTAGE OF DEAD NODES AND NUMBER OF ROUNDS

Dead nodes%	No. of Rounds				
	LEACH	CGA	EAERP	GAECH	OMPFM
10%	1898	2163	2244	2440	2646
20%	2013	2146	21782	2331	2713
40%	2057	22602	2334	2366	2644
60%	2201	2277	2356	2478	2994
80%	2225	2274	2366	2522	3244
100%	2255	2354	2409	2490	4418

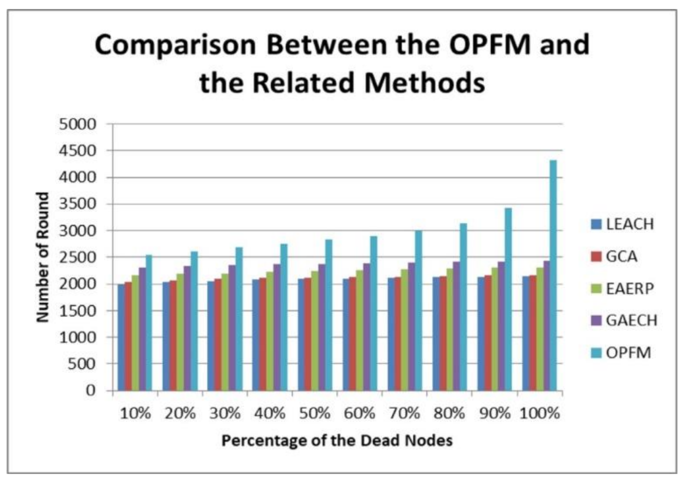


Fig. 5. Results of the Suggested Method's Simulation Compared to Similar Techniques.

TABLE VII. PROPOSED MODEL OMPFM FOR LND AND FND

FND	LND
1633	4155
1778	4422
1781	3754
1752	3889
1614	4167
1622	3923
1689	3855
1798	4109
1789	3755
1699	3666

V. CONCLUSION AND FUTURE SCOPE

Data transmission in WSNs has a significant impact on a network's lifetime performance. It's difficult to use direct transmission in WSNs because of the increased energy usage. This is especially true when the source CH is far from the BS. Furthermore, choosing the right CHs is a problem that has a big impact on the network's lifespan. These issues have been the subject of a slew of investigations. A novel approach to the direct transmission problem, based on a modified GA, is given in this work. The settings of the boundary variables for the cluster head selection function can be altered to improve results based on the specific application. A change is recommended to the CHs selection threshold. Since the suggested technique is half as good as the LEACH protocol and other comparable methods presently in use in terms of network lifetimes and power consumption, this would lead to longer run times. Eventually, a new evolutionary algorithm will be implemented to speed up the processing time. Additionally, future mobile WSNs will consider a multi-hop routing method.

REFERENCES

[1] Sankaralingam, S.K., Nagarajan, N.S. and Narmadha, A.S. (2020) Energy aware decision stump linear programming boosting node classification based data aggregation in WSN. Comput. Commun.

[2] Maheshwari, P., Sharma, A.K. and Verma, K. Energy efficient cluster based routing protocol for WSN using butterfly optimization algorithm and ant colony optimization. Ad Hoc Netw., 110, 102317.

[3] Singh, R. and Verma, A.K. (2017) Energy efficient cross layer based adaptive threshold routing protocol for WSN. Int. J. Electron. Commun., 72, 166–173.

[4] Sood, T. (2020) LUET: a novel Lines-of-Uniformity based Clustering protocol for Heterogeneous-WSN for multipleapplications. J. King Saud Univ. Comput. Info. Sci.

[5] GouseBaig Mohammad, S Shitharth, Puranam Revanth Kumar (2021). Integrated Machine Learning Model for an URL Phishing Detection. International Journal of Grid and Distributed Computing, vol. 14, Issue 1, pp. 513-529.

[6] Sundararaj, V. (2019a) Optimal task assignment in mobile cloud computing by queue based Ant-Bee algorithm. Wirel. Pers. Commun., 104, 173–197.

[7] Sundararaj, V. (2019b) Optimised denoising scheme via opposition-based self-adaptive learning PSO algorithm for wavelet-based ECG signal noise reduction. Int. J. Biomed. Eng. Technol., 31, 325–345.

[8] Sundararaj, V., Muthukumar, S. and Kumar, R.S. (2018) An optimal cluster formation based energy efficient dynamic scheduling hybrid MAC protocol for heavy traffic load in wireless sensor networks. Comput. Secur., 77, 277–288.

[9] Sundararaj, V., Anoop, V., Dixit, P., Arjaria, A., Chourasia, U., Bhambri, P., Rejeesh, M.R. and Sundararaj, R. (2020) CCGPAMPPT: Cauchy preferential crossover-based global pollination algorithm for MPPT in photovoltaic system. Prog. Photovolt., 28, 1128–1145.

[10] Rejeesh, M.R. (2019) Interest point based face recognition using adaptive neuro fuzzy inference system. Multimed. Tools Appl., 78, 22691–22710.

[11] Rejeesh, M.R. and Thejaswini, P. (2020) MOTF: Multiobjective Optimal Trilateral Filtering based partial moving frame algorithm for image denoising. Multimed. Tools Appl., 79, 28411–28430.

[12] Singh, A. and Nagaraju, A. (2020) Low latency and energy efficient routing-aware network coding-based data transmission in multi-hop and multi-sink WSN. Ad Hoc Netw., 107, 102182.

[13] Sharma, B.M., Neelam, S. and Singh, K. (2020) QoS-based energy-efficient protocols for wireless sensor network. Sustain. Comput. Inform. Syst., 100425.

[14] Samara, G. and Aljaidi, M. (2019) Efficient energy, cost reduction, and QoS based routing protocol for wireless sensor networks. arXiv, preprint arXiv: 1903.09636.

[15] Karunanithy, K. and Velusamy, B. (2020) Cluster-tree based energy efficient data gathering protocol for industrial automation using WSNs and IoT. J. Ind. Inf. Integr., 100156.

[16] Baradaran, A.A. and Navi, K. (2020) HQCA-WSN: high-quality clustering algorithm and optimal cluster head selection using fuzzy logic in wireless sensor networks. Fuzzy Set. Syst., 389, 114–144.

[17] Dhand, G. and Sheoran, K. (2020) Protocols SMEER (Secure Multitier Energy Efficient Routing Protocol) and SCOR (Secure Elliptic curve based Chaotic key Galois Cryptography on Opportunistic Routing). Mater. Today.

[18] Jeske, M., Rosset, V. and Nascimento, M.C.V. (2020) Determining the trade-offs between data delivery and energy consumption in large-scale WSNs by multi-objective evolutionary optimization. Comput. Netw., 107347.

[19] Stephan, T., Al-Turjman, F., Joseph, K.S., Balusamy, B. and Srivastava, S. (2020) Artificial intelligence inspired energy and spectrum aware cluster based routing protocol for cognitive radio sensor networks. J. Parallel Distrib. Comput.

[20] Nivedhitha, V., Gopi Saminathan, A. and Thirumurugan, P. (2020) DMEERP: a dynamic multi-hop energy efficient routing protocol for WSN. Microprocess. Microsyst., 79, 103291.

[21] Seyyedabbasi, A. and Kiani, F. (2020) MAP-ACO: an efficient protocol for multi-agent pathfinding in real-time WSN and decentralized IoT systems. Microprocess. Microsyst., 103325.

[22] Devi, V., Seedha, T.R. and Baghavathi Priya, S. (2020) Cluster based data aggregation scheme for latency and packet loss reduction in WSN. Comput. Commun., 149, 36–43.

[23] Shi, Y. (2011) Brainstorm Optimization Algorithm. In Proc. 2nd Int. Conf. Swarm Intelligence, pp. 303–309.

[24] Shi, Y. (2015) Brain Storm Optimization Algorithm in Objective Space. In Proc. IEEE CEC.

[25] Osborn, A.F. (1957) Applied Imagination. Scribner, New York. [26] Arai, K., Fukuyama, Y., Iizaka, T. and Matsui, T. (2019) Optimal operational planning of energy plants considering uncertainty of renewable energy outputs by global-best brain storm optimization. IFAC-Papers OnLine, 52, 153–158.

[26] S. Mann , S. Singh , Energy-efficient hierarchical routing for wireless sensor networks: a swarm intelligence approach, *Wireless Pers. Commun.* 92 (2017) 785–805.

[27] S. Mudundi , H. Ali , A new robust genetic algorithm for dynamic cluster formation in wireless sensor networks, in: Proceedings of the 7th IASTED International Conferences on Wireless and Optical Communications (WOC '07), Montreal, Canada, June 2007, pp. 360–367.

[28] A. Norouzi , H. Zaim , Genetic algorithm application in optimization of wireless sensor networks, *Sci. World J.* (2014).

[29] P. Ragavan , K. Ramasamy , Optimized routing in wireless sensor networks by establishing dynamic topologies based on genetic algorithm, *Clust. Comput.* (2018) 1–7.

[30] S. Sivanandam , S Deepa , Genetic algorithm implementation using MATLAB, *Introd. Genet. Algorithm.* (2008) 211–262.

[31] A .K. Sohal , A .K. Sharma , N. Sood , A study on energy efficient coverage aware clustering protocols in wireless sensor networks, J. Netw. Secur. 6 (3) (2019) 9–17.

[32] Y. Song , C. Gui , X. Lu , H. Chen , B. Sun , A genetic algorithm for energy-efficient based multipath routing in wireless sensor networks, *Wirel. Pers. Commun.* 85 (4) (2015) 2055–2066.

[33] R. Urena , G. Kou , Y. Dong , F. Chiclana , E. Herrera-Viedma , A review on trust propagation and opinion dynamics in social networks and group decision making frameworks, *Inf. Sci.* 478 (2019) 461–475.

[34] M. Warrier , A. Kumar , An energy efficient approach for routing in wireless sensor networks, *Proc. Technol.* 25 (2016) 520–527.

[35] H. Zhang , P. Chen , S. Gong , Weighted spanning tree clustering routing algorithm based on leach, in: Proceedings of the 2nd International Conference on Future Computer and Communication (ICFCC), 2010, pp. 223–227. May.

[36] R. Zhou , M. Chen , G. Feng , H. Liu , S. He , Genetic clustering route algorithm in WSN, in: Proceedings of the 2010 6th International Conference on Natural Computation, ICNC, 2010, pp. 4023–4026.