

Indoor Pollutant Classification Modeling using Relevant Sensors under Thermodynamic Conditions with Multilayer Perceptron Hyperparameter Tuning

Percival J. Forcadilla

Department of Computer Science, College of Science
University of the Philippines Cebu
Lahug, Cebu City, Philippines

Abstract—Air pollutants that are generated from indoor sources such as cigarettes, cleaning products, air fresheners, etc. impact human health. These sources are usually safe but exposure beyond the recommended standards could be hazardous to health. Due to this fact, people started to use technology to monitor indoor air quality (IAQ) but have no capability of recognizing pollutant sources. This research is an improvement in building a classification model for recognizing pollutant sources using the multilayer perceptron. The current research model receives four data parameters under warm & humid and cool & dry conditions compared to nine parameters of the previous literature in detecting five pollutant sources. The classification model was optimized using GridSearchCV to obtain the best combination of hyperparameters while giving the best-fit model accuracy, loss, and computational time. The tuned classification model gives an accuracy of 98.9% and a loss function value of 0.0986 under the number of epochs equal to 50. In comparison with the previous research, the accuracy was 100% with the number of epochs equal to 1000. Computational time was greatly reduced at the same time giving the best-fit accuracy and loss function values without incurring the problem of overfitting.

Keywords—Indoor air pollutants; pollutant sources; indoor air quality; IAQ; sensors; multilayer perceptron; classification modeling; gridSearchCV; hyperparameter tuning

I. INTRODUCTION

In the last several years, scientific evidence has been produced regarding indoor air being more harmful than the air outdoors. That is why, people at homes, offices, and schools desire to have indoor air to be fresh and pleasant, not harmful to health to promote productivity. Air pollutants are one of the contributory factors in harming indoor air which comes from different sources and significantly impacts the health of the people in the long run. With this threat present, relevant organizations develop Indoor Air Quality (IAQ) standards and guidelines to eliminate or prevent people from the potential health risk of these indoor pollutants. These organizations gave threshold values for concentration levels and exposure time in maintaining the optimal IAQ level.

Aside from the IAQ standards and guidelines made by different organizations around the world, the research community contributed a lot to the development of advanced technologies that will help monitor and keep IAQ at the optimum level. Research themes like performance assessments on industry and consumer sensors and devices, factors affecting indoor air pollution in residential and commercial spaces,

machine learning algorithms, strategies, and techniques were also included giving new contributions to the existing body of knowledge. But with all these research trends, it was found that limited contribution was given to IAQ research on pattern recognition techniques. This gave the researcher motivation to further study the topic revolving around machine learning using pattern recognition techniques.

One of the studies that were considered was from Saad Shaharil Mad et al. [19]. In the study, nine parameters were used to develop pollutant recognition based on pattern recognition techniques. With the motivation of exploring the topic of pattern recognition techniques, the researcher formulated a research question to further explore the topic. The question was - How effective is an algorithm if the dataset parameters are reduced from nine to targeted parameters in classifying indoor air pollution sources in terms of model accuracy and loss? The research question was then supported with another study which gives this research a starting point for further investigation. That study was from Demanega Ingrid et al. [19] and was all about assessing the performance of low-cost monitors and not classifying indoor pollutants. The result of the study shows positive detection outcomes for all sensors and devices using particulate matter (PM), carbon dioxide (CO₂), total volatile organic compounds (TVOC), indoor air temperature, and relative humidity as the parametric specifications considered in the research. Most of the low-cost sensors in the study responded to the simulated indoor pollutant sources with varying parametric readings. Also, the pollutant sources in the study of Demanega Ingrid et al. [11] and Saad Shaharil Mad et al. [19] shows to have relatively the same sources used as indoor pollutants. On that note, a hypothesis was then created to predict that selecting key parameters out of the nine parameters that were considered in the study of Saad Shaharil Mad et al. [19] can still be able to recognize or classify indoor air pollutants. With this study, the sensors to use were purely dependent on consumer-grade monitors available in the market which has no ability to recognize sources. The selected parameter that was considered were PM, CO₂, TVOC, and formaldehyde (HCHO). These parameters were collected in a controlled room setup where indoor air temperature and relative humidity were set as one parameter while generating an environment simulation commonly encountered in many indoor climates around the world. These indoor climates are set as two thermodynamic conditions, warm & humid and cool & dry conditions which were adopted in the research

of Demanega Ingrid et al. [11]. The multilayer perceptron classification model algorithm was used as the only algorithm for this study since it performs to have the highest classification accuracy in the previous study. Lastly, to further improve the previous work and set an additional contribution, optimization was done through hyperparameter tuning.

This paper starts by introducing what Indoor Air Quality is followed by the significance of the study, and the different research trends that bring motivation to implementing this research. Validation of the research question was done through the review of the literature followed by a brief description of the methodology. The methodology starts with data collection of the target parameters under two thermodynamic conditions, processing the data through a pattern recognition technique (Multilayer Perceptron) leading to the classification of pollutant sources. Afterward, a preliminary analysis was done to show the parameter's categorical correlation followed by the model's accuracy and loss results. Finally, an optimization process through hyperparameter tuning was also performed.

II. LITERATURE REVIEW

A. Indoor Air Quality: Description, Sources and Effects to Human Health

The quality of air inside a building or structure like homes, schools, and offices which promotes good health and comfort for occupants is called indoor air quality (IAQ). It is also the extent to which human requirements in terms of indoor air are met. In this case, it is the desire which air should be fresh and pleasant, have no harmful effect on health, promotes good working conditions in the context of building offices or homes, and productivity in terms of learning at schools [1]. IAQ has been recognized as an important factor in people's health and comfort in indoor environments because 90% of people's time is spent indoors [2]. Also, studies have shown that the occupants are 100 times more exposed to indoor air pollutants than outdoor air pollutants and the concentration of air pollution indoor is seen to be 2 to 4 times higher than that of outdoor [3].

When indoor air quality is not taken into consideration for buildings and structures, possible health concerns due to high indoor pollutant exposure may be experienced. According to United State Environmental Protection Agency (EPA), there are many sources of indoor air pollution. Sources like fuel-burning combustion appliances, tobacco products, products for home cleaning and maintenance, building materials, furnishings, a product like air fresheners, and many more [4]. Over the years, these sources have been producing gaseous pollutants which are chemicals tagged as hazardous. These include radon, ozone, nitric oxides, sulfur dioxide, carbon monoxide, diatomic carbon, and VOCs [5]. Aside from these sources, particulate matter from combustion and cleaning activities, heavy metals from fuel consumption and building materials, airborne particles, pest control chemicals, and biological contaminants are recognized as air pollutants harmful to human health [6].

With the threat of pollutant exposure leading to significant risks to human health, relevant organizations developed different IAQ standards and guidelines. Among these organizations are the World Health Organization (WHO) and the United

States Environmental Protection Agency (USEPA) which contributed to the making of indoor air quality standards and guidelines. The guidelines serve as a database reference to eliminate or prevent people from too much exposure to indoor air pollution and the potential health risks that may be brought to the human population [7]. Aside from WHO and USEPA, other recognized organizations like ASHRAE [22], National Health & Medical Research Council in Australia, and many more around the world have set guidelines and threshold values to maintain an optimal IAQ [8].

B. IAQ Current Research Trends

With the standards and guidelines set by the different organizations and the development of sensor technologies in a network and systems, different research trends have been developed through time. Most of this literature fall under advanced technologies for monitoring, performance assessments of cheap and high-end monitoring device or systems, factors affecting indoor air pollution in residential and commercial spaces, forecasting, and pattern recognition techniques.

1) *Advanced technologies for monitoring IAQ:* The development of mobile technologies and the internet of things (IoT) has brought great capability toward improving IAQ monitoring systems. Air quality monitoring nowadays can easily be done through smartphones by simply accessing the web. In the study of M. Tastan et al. the research proposed an IoT-based real-time e-nose measuring system using low-cost electronic sensors [9]. The system includes sensors such as MH-Z14A for Carbon Dioxide, MICS-4514 for Nitrogen Oxide and Carbon Monoxide sensor, GP2Y1010AU as a dust sensor, and DHT22 for Temperature and Humidity sensor together with an ESP32 microcontroller with built-in Wi-fi, used to process the information provided by the sensors. Pollutant concentration data is thrown into the Blynk cloud, and an android-based mobile user interface was developed for users to access the data in a form of digital or graphical monitoring.

In another study by Giacomo Chiesa et al. [10], a system was developed based on several sensor data to model the IAQ which serves as input in controlling the ventilation system. The system is connected to an app that features management of the device, real-time data visualization, and statistical data [10]. Device management allows the user to create a list of installed devices and set desired ventilation time to report indoor conditions in threshold value for each device. Real-time data visualization includes the quality level of indoor air and each of the parameters because of the sensor devices, and lastly, statistical data which gives users daily or weekly graphs for each sensor. The parameters that were considered for this research were carbon dioxide, TVOC, pressure, humidity, and temperature. Raspberry Pi serves as the backend platform where it handles device management, sensor-to-microcontroller interface, sensor data communication, control algorithms, data storage, and aggregation. The research data platform used for IoT needs was MongoDB. It is a source-available cross-platform document-oriented database program.

2) *Performance assessment on IAQ monitoring device and systems:* Another research trend from the existing body of knowledge revolves around an IAQ theme that was based on IAQ monitoring devices and system performance assessment

and evaluation. Performance assessment is needed since there are different sensor devices and systems in the market and these sensors can either be low-cost or industry-grade sensors that influence the sensor reading's accuracy. Sensor devices and systems were tested to know their performance reliability or prove if these devices can competently measure a target parameter. In the study of Ingrid Demanega et al. consumer environmental monitors available in the market together with low-cost single variable sensors were assessed to know their reading reliability in thermodynamic conditions [11]. Different sources were introduced inside a room chamber like candle burning, mosquito coil burning, wood lacquer drying, room deodorant injection, essential oil heating, carpet vacuuming, popcorn cooking, and carbon dioxide injection. The parameters to be evaluated were the particulate matter (1.0, 2.5, and 10), carbon dioxide, and TVOC. Particulate Matter Monitor miniWRAS, LI-COR 850 Bioscience gas analyzer for carbon dioxide, and GrayWolf AdvancedSense Pro as TVOC monitor were all used as the industry-grade reference monitor to evaluate the consumer-grade monitors and sensors.

In another research, Zhiqiang Wang et al. [12] tested the performance of low-cost IAQ monitors for PM 2.5 and 10 [12]. The low-cost monitors are Air Quality Egg 2018 version, IQAir Airvisual Pro, Awair 2nd Edition, Kaiterra Laser Egg 2, PurpleAir Indoor, and Ikaair with reference measurement systems together with professional-grade particulate monitors. The test chamber used in this research is a room of 120m³ with three external walls, two doors, and raised ceiling. Several sources were used in multiple experiments, sometimes with variations that include measuring indoor concentrations of infiltrated outdoor PM_{2.5} were collected to evaluate the performance of the low-cost devices.

3) *Factors influencing indoor air pollution:* There are different factors that contribute to the level of indoor air pollution inside buildings and structures. Several researchers have spent time and effort discovering such factors with the aim to improve air quality inside buildings and structures. In the study of Wonho Yang et al. [13], the IAQ level was investigated in different schools in Korea with the intention to know the correlation with respect to the age of the buildings [13]. Air samples were taken indoors and outdoors with parameters to consider such as carbon monoxide, carbon dioxide, particulate matter (PM₁₀), total microbial count, total volatile organic compounds, and formaldehyde. Results show that the factors which contributed to indoor air pollution in schools are unsatisfactory ventilation and chemical emissions from building materials or furnishings. Significant high concentrations of carbon dioxide, TVOCs, and HCHO were also found at schools constructed within one year.

A relevant factor that influences the variation of indoor pollutants inside the buildings and structures points to the different seasons. In the study of Corinne Mandin et al. [14], a European project OFFICAIR was made to gain more knowledge with regard to IAQ in modern office buildings. It was found that significantly higher concentrations of formaldehyde and ozone were measured in summer whereas benzene, α -pinene, D-limonene, and nitrogen dioxide were significantly high during winter. Another study focused on the variation in the concentration of pollutants at different locations in India. The study of Arindam Datta et al. [15] focuses on

the indoor air quality of non-residential buildings which is scarce in India [15]. It was found that among different non-residential buildings, a lower concentration of pollutants was recorded in the educational building compared to the two office buildings. A ductless air-conditioning system with poor air circulation and active air filtration contributed to the higher concentration of PM_{2.5}. In Doha, Qatar, another study was carried out to investigate the level of PM_{2.5} and PM₁₀ in office environments [16]. It was found that the cause for significant concentrations of particulate matter inside the office is ventilation, faulty envelopes, and windows.

Different factors influencing indoor air quality have been a trend in the research community. Aside from the identified factors and themes that were done by the researchers above, the study of Mehzabenn Mannan et al. [8] summarizes some of the factors that were gathered by the researcher's related literature which influenced indoor air pollution. The research identified factors like different indoor building materials and few surface finishes and appliances, nearby construction activity, indoor movement, tobacco smoke, and computer operation, high benzene concentration was observed in the lower-level classrooms and school carpet was seen to be responsible for higher PM. Another factor indicated to be the source of air pollution was the concrete additive in an office while comparing two office settings (Beijing and Stockholm) and the contribution of indoor air pollution with respect to newly built and refurbished office buildings. Factors like occupant behavior, the role of humans, respiratory emissions from human beings, and the reaction of ozone to skin lipids are seen to be giving a contribution to indoor air pollution.

4) *Machine learning and statistical modeling in IAQ:* Building a system that can forecast the concentration level of pollutants to characterize indoor air quality has been a long important topic in the community and domain of indoor environment and health science [17]. In real-occupied environments, statistical modeling has great potential to explore and predict indoor air pollution concentration levels [18]. Statistical modeling on IAQ can use forecasting techniques to predict the level of IAQ and pattern recognition techniques which allow the system to recognize certain types of smell [19].

In the study of Wenjuan Wei et al. [18], a summary of common machine learning and statistical modeling methods was collected through a literature review. Methods were compared based on strengths and weaknesses while discussing how and where the methods have been used in the field of IAQ. A summary of machine learning algorithms used in IAQ are based on either supervised or unsupervised learning with a different type of response variable and linearity of the model.

a) *Forecasting techniques:* A study about forecasting indoor concentration levels in an office space using machine learning was made by Johanna Kallio et al. [20]. The research contribution gives the body of knowledge a comprehensive dataset covering a full year with the applicability of four machine learning methods which include ridge regression, decision tree, random forest, and multilayer perceptron. Accuracy was evaluated with respect to the different methods of prediction, history window time frames, and the impact of multiple sensor modalities. In another study, Shisheng Chen et al. [23] use a machine learning approach to predict CO₂, TVOC, and HCHO. Five Classrooms at the National University

of Singapore were used to perform the continuous parametric collection of data. The dataset was trained and tested using Support Vector Machine, Gaussian Processes, M5P, and a backpropagation neural network. According to Wenjuan Wei et al. [18], an artificial neural network (ANN) is the most popular method for the prediction of IAQ based on the researcher's findings. Dwellings, offices, schools, and subway stations are the different sites where ANN modeling was used to predict several IAQ parameters [18].

b) IAQ pattern recognition techniques: Forecasting techniques prove to be relevant and saturated in the field of indoor air quality. In the case of pattern recognition techniques, little literature has been found to have a relevant contribution to the IAQ research community. One of which was a study that uses pattern recognition techniques to recognize specific types of pollutants by Saad Shaharil Mad et al. The authors of this topic publish two papers in 2015 and 2017. The first paper classifies pollutant sources with the use of one pattern recognition technique, ANN. The other study was an enhancement to the previous paper which utilizes different supervised machine learning algorithms like multilayer perceptron, k-nearest neighbors, and linear discrimination analysis. Nine (9) parameters were used to classify five (5) different pollutant sources. These parameters were nitrogen oxide, carbon dioxide, ozone, carbon monoxide, oxygen, VOCs, particulate matter, temperature, and humidity [19].

C. IAQ Parameter and Sensors

The research topic regarding indoor air quality was narrowed down to the context of classifying pollutant sources using pattern recognition techniques. To classify indoor air pollutant sources, sensors must collect different parametric data to generate the IAQ dataset. Choosing the target parameters with the corresponding sensors serves as one of the important points to make this research a success.

1) Relevant IAQ parameter consideration: In the study of Saad Shaharil Mad et al. [19], sensors were used to target the different parameters like CDM4161 for carbon dioxide, TGS5342 for carbon monoxide, TGS2602 for VOC, MiCS2610 for Ozone, MiCS2710 for nitrogen dioxide, KE25 for oxygen, HSM20G for Temperature and Humidity, and GP2Y1010AUOF for Particulate Matter (PM10). These sensors were used to respond to the five indoor pollution sources, such as the ambient air, combustion activity, chemical presence, fragrance product, and food & beverages [19].

In another study, H. Zhang et al. [21] developed a low-cost IAQ multi-pollutant monitoring system using raspberry pi. Different sensors with specifications from different manufacturers together with their prices were carefully considered to be included in the Low-Cost Air Quality System or LCAQS. Sensors that measure Relative Humidity, Temperature, Particulate Matter (PM_{2.5/10}), Nitrogen Dioxide, Sulfur Dioxide, Carbon Dioxide, Carbon Monoxide, Ozone, and Total Volatile Organic Compound (TVOCs) was used to develop the system [21].

To measure the indoor air quality level, parameters were divided into categories: physical condition, chemical contaminants, biological contaminants, and other common IAQ parameters. Using these categories, the study of Saad Shaharil

Mad et al. divided the sensors into three types: gas sensor, particle sensor, and thermal sensor [19]. The two studies above both have common thermal and particle parameters (Temperature, Relative Humidity, and Particulate Matter) with different IAQ research themes. Both studies above utilize nine (9) parameters, and most of the gas parameters were common to both studies except for oxygen and sulfur dioxide. In the study of Demanega Ingrid et al. [11], four parameters were only utilized to assess low-cost environmental monitors and single sensors which were used to respond to different indoor pollution sources. These parameters were temperature, relative humidity, carbon dioxide, and particulate matter [11].

Also, it can be seen in the study of Demanega Ingrid et al. [11] that the results of using the different simulated activities for indoor pollution sources show positive detection outcomes for all sensors and devices used in the study. Different particle sensor responds to indoor pollution sources such as candle burning, mosquito oil burning, and popcorn cooking. This reflects that the current study has the basis to use only target parameters to classify pollutant sources.

2) Thermodynamic conditions: To simulate an environment that is commonly encountered indoors in many climates around the world, thermodynamic conditions should be considered which helps assess the performance based on the standard thermal comfort zone [11,22]. This methodology was also adopted in this research in using thermodynamic conditions – warm & humid (26 +/- 1 C, 70 +/- 5%) and cool & dry (20 +/- 1 C, 30 +/- 5%)[11].

D. IAQ Research Motivation and Summary

The exhaustive search done on the internet through online research databases proves that the topic of indoor air quality revolves mostly around advanced technologies for monitoring, performance assessments of cheap and high-end monitoring devices or systems, factors affecting indoor air pollution in residential and commercial spaces, and research about machine learning specifically forecasting techniques. On that note, a limited contribution was found regarding pattern recognition techniques in IAQ research, thus, this research takes that route related to pattern recognition. One of the pattern recognition studies that were reviewed came from Saad Shaharil Mad et al. which became the major motivation for doing the current research. The research done by Saad Shaharil Mad et al. considers nine parameters to classify pollutant sources. This pique the interest of the researcher to formulate a research question in optimizing the previous strategy. A strategy to select a few parameters out of the nine, to classify pollutant sources. In selecting parameters to be included in the classification process, the study of Demanega et al. shows results that were relevant in choosing the parameter of this research. To add another layer of parameter, the thermodynamic condition was also adopted to simulate two indoor climates commonly encountered.

III. METHODOLOGY

This chapter discusses the key design choices, concepts, and procedures in attaining the classification of Indoor Air Pollutant Sources using targeted Pollutant Parameters based on Machine Learning's Pattern Recognition Techniques for Indoor

Air Quality (IAQ) Systems. Fig. 1 below shows the outline of the methodological process on how to attain the research objectives.

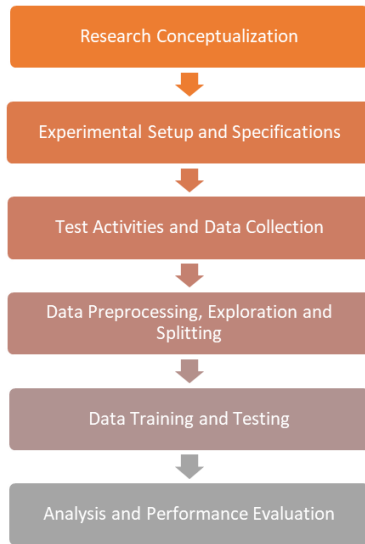


Fig. 1. Research framework flowchart

After establishing the research novelty, relevance, and feasibility through the review of related literature, the development procedure of this research starts with the implementation stage through research conceptualization. This defines the research considerations regarding the approach, initial strategies, the participants involved, and the research setting. Experimental setup and specifications were the next procedure which provides details regarding the room setup, thermodynamic conditions, sources of indoor air pollutants to consider, IAQ parameters, and what device monitor to use. When the different setup and conditions were met in the second stage, test activities and data collection follows. This procedure provides details as to how the data was collected in the room setup, the simulation activity for the source of pollutants, and the timeframe needed for the device to collect data. When all the necessary data under different conditions were collected, the raw data undergoes the stage of preprocessing and data splitting. Raw data was secured to be clean, and normalized, making the data classification-ready, and undergoes data splitting for the training proper. After the previous procedure, data training follows using a pattern recognition technique identified in this study. The trained data was then validated based on its accuracy and performance evaluation was done through the given statistical methods involved in this research.

A. Research Conceptualization

The ground theory of the previous literature builds the foundation of this research. The literature of this study revolves around (1) the significance of why IAQ is needed and the possible health impact on humans, (2) the indoor air quality research trends, giving details to literature who were well-studied, researched, and studies with limited resource contribution, (3) IAQ parameters to consider using the sensor results and readings under different thermodynamic conditions,

and (4) different pattern recognition techniques based on the machine learning research trends.

In the last several years, indoor air quality has been a well-studied area in the environmental research community. Different environmental organizations have taken steps to establish standards and guidelines to address issues and potential risks to human health. Technological advancements are also given attention to the research community and IAQ industry on intelligent systems, IoT, machine learning, etc. During the literature review, limited research contribution was seen in the area of indoor air quality which involves pattern recognition techniques. This identified gap motivated the researcher to conduct further study on the topic. The study of Saad Shaharil Mad et al. [19], was the main literature that gave motivation to this research. This study is an improvement to the study of Saad Shaharil Mad et al. [19] by taking into consideration the use of targeted parameters under different thermodynamic conditions.

1) *Research approach and initial strategies:* The study's general objective was to classify five (5) indoor air pollutants using thermal, particle, and gas parameters under two (2) thermodynamic conditions using multilayer perceptron pattern recognition technique. Thus, the research approach should be quantitative. Another consideration in selecting what research approach to use was based on the typical approach used by the related literature in this research. Also, the nature of the research objectives should clearly define what strategy to utilize. Since the research is about the classification process in machine learning then the strategy to pursue should be the predictive strategy.

2) *Participants and setting:* The usual strategy for the collection of data regarding air quality uses environmental chambers and other controlled room setups. In the study of Shisheng, Chen et al. [23] regarding time series prediction of CO₂, TVOC, and HCHO, they made use of the rooms with the air-conditioning unit (ACU) inside the campus of the National University of Singapore (NUS) as the location for the air quality data collection [23]. In the same way, this research used the facility of the University of the Philippines – Cebu as the location for collecting air parameters in classifying pollutant sources. Specifically, this research was conducted in the Department of Computer Science Conference Room, 3rd floor, Room 313 of the Arts and Sciences (AS) building University of the Philippines - Cebu.

B. Experimental Setup and Specifications

Before acquiring the data for preprocessing and data training, needed preparations were done to achieve organized procedures regarding the simulated test activities. This subpart talks about the setup and conditions inside the room, pollutant sources and parameters to consider, and the IAQ device monitor to use.

1) *Room setup and conditions:* Fig. 2 below shows the room's details, fixtures, and specifications. The figure was the floorplan venue for the collection of indoor air parameters. The room has the dimension of 5.33m x 3.10m x 3.05m, it has one door, three windows, one window-type air-conditioning unit (ACU), and fixtures like couches (small and big), a conference table with office chairs, and small cabinet.

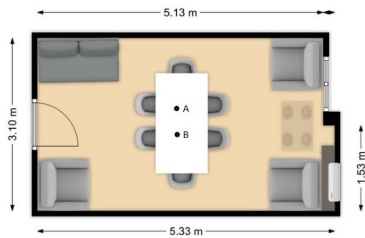


Fig. 2. Room specifications using floorplanner.com

2) *Sources of indoor air pollutants:* The research also adopted one of the indoor air pollutants activities being used in the study of Demanega, Ingrid, et al. [11] which was candle burning for combustion activity. Other source pollutants adopted came for the study of Saad, Shaharil Mad et al. [19] such as cleaning agents like “Lysol”, rotten cooked fish, and the ambient condition. This research also included rubbing alcohol as one of the sources which is currently a big contributor to indoor air pollutants because of the pandemic. Table I shows the summary of the source pollutants included in this research with the activity description, data collection interval, and span. Each pollutant undergoes two thermodynamic conditions cool & dry and warm & humid. Data was then collected every 1-minute interval and a collection span of 16 hours or 960 samples will be collected across each of the source pollutants considering the thermodynamic condition. An 8-hour day plan was decided in consideration of the university’s working hours. Gathering samples for the 5 single source pollutants for each thermodynamic condition gives us a total of 9600 samples or 160 hours of collection time. Given an 8-hour day plan, 160 hours is equal to 20 days collection process.

TABLE I. ACTIVITY DESCRIPTION, DATA COLLECTION INTERVAL OF THE SOURCE POLLUTANTS

Source Pollutants	Thermodynamic Condition	Activity Description	Data Collection Interval
Ambient	Cool and Dry	Observe the ambient condition reading of the surrogate sensors and device reference.	Collect data every 1 minute. A total of 960 samples should be collected.
	Warm and Humid		
Combustion	Cool and Dry	Light up two candles in the room.	
	Warm and Humid		
Chemical	Cool and Dry	Place half full glass of isopropyl alcohol in the middle of the room.	
	Warm and Humid		
Fragrance	Cool and Dry	Automatic Spray using Lysol for every 18 minutes in the room.	
	Warm and Humid		
Rotten Food	Cool and Dry	5 days old rotten food(leftovers) will be placed in the middle of the room	
	Warm and Humid		

3) *Indoor air quality parameter and device monitor:* In the study of Saad Shaharil Mad et al. [19], nine parameters were used to develop a pollutant recognition based on pattern recognition techniques, but the study of Demanega Ingrid et al. [11] paved way to consider only using a few target parameters in classifying indoor pollutants. That same

study gave results showing positive detection outcomes for all sensors and devices. Most of the sensors responded to the simulated indoor pollutant sources which gave way to adopting the previous studies’ parameters. The pollutant parameters that were considered were air temperature and relative humidity for thermal parameters, particulate matter (PM2.5/10) for particle parameters, and carbon dioxide as a gas parameter. Thus, this study considers the parameters that were studied in the previous related work.

The study of Demanega Ingrid et al. [11], provides an assessment of low-cost monitors, research, and professional-grade IAQ systems used as reference monitors. These systems are expensive, yet the reading accuracy is top-notch with respect to the different IAQ parameters. Not just expensive, most of these systems detect a single parameter only unlike the consumer-grade IAQ monitors which are available in the market. But in that same study, some of the consumer-grade monitors were assessed and exhibited good performance grading in detecting the different parametric values for pollutant sources. On that note, the IAQ monitor that was used in this study was low-cost and consumer-grade monitors targeting the relevant parameters that were considered in the previous study. Two IAQ handheld devices were considered in implementing this research. Temptop M2000 2nd Generation was use to collect CO2 and particulate matter. BR-SMART-123SE was used to collect TVOC and HCHO. These devices were then positioned in the middle of the room, specifically on the top of the table.

4) *Working with thermodynamic conditions:* According to the study of Xiangguo et al. [24], conventional all-air central air-conditioning (AC) systems can control the temperature and humidity through cooling, reheating, and humidifying equipment but AC systems which were commonly seen in small and medium-sized buildings have no specific dehumidifying equipment to deal with moisture. In the Philippines, relative humidity is high because of the surrounding body of water. It is said on the PAGASA website that the average monthly relative humidity varies between 71 percent in March and 85 percent in September. In a requirement to have fresh air indoors, good ventilation by opening the windows can help but will greatly influence the moisture level. In working with thermodynamic conditions and achieving the cool & dry and warm & humid room setup, doors and windows were sealed to avoid moisture air influencing the indoor humidity and temperature. Also, dehumidifiers and humidifier equipment were placed inside to control the humidity level of the room while an AC unit was used to control the room temperature with a significant contribution to the relative humidity. The dehumidifier and humidifier equipment has automatic control which directly reacts with the varying humidity levels in the room.

C. Data Collection and Test Activities

Before the collection procedure, the two thermodynamic conditions should be first met. In the case of cool and dry, the ACU was turned on together with the dehumidifier for some time until the conditions were met for an activity to happen. For relative humidity control, only the dehumidifier was turned on since the room space was not completely sealed. A humidity level of 30% was never achieved in consideration with the study of Demanega Ingrid et al. [11]. Only 45% +/- 5% was

achieved in this research. Both the ACU and dehumidifier have a control device that shuts off when a certain temperature or humidity value is reached. For warm and humid, ACU will be turned off and both the dehumidifier and humidifier were turn on to control the humidity level of the room. In the same way, the ACU, dehumidifier, and humidifier were connected to a control device that establishes the right thermodynamic condition. When the conditions were set, different source pollutants were injected. Also, the device monitors which were placed in the middle of the room were turned on at least 1 hour before each activity allowing the sensors to have enough time to stabilize and the collection procedure commences as soon as the desired thermodynamic conditions were met.

Table I shows the order in collecting the source pollutants data with the desired activity description. A warm and humid condition will be implemented first to be followed by a cool and dry condition. The start and end of the collection procedure will be manually timestamped to map the desired data activity. PM2.5, CO2, TVOC and HCHO was collected while considering thermodynamic conditions. Thus, giving this research a total of 5 parameters.

D. Data Preprocessing, Exploration and Splitting

The raw dataset that was then generated from the device monitor and was extracted through USB in excel format. The manual timestamp was defined and divided into different data based on the thermodynamic conditions and different pollutant sources. After organizing the dataset, data cleaning was followed. This was where incorrect data type or format, missing values, and data duplication in the dataset were either modified, replaced, or deleted before the data normalization/standardization process and choosing other data preprocessing techniques. Preprocessing is important because it may aid in the success of pattern recognition performance [19].

After the procedures above, data exploratory analysis follows. This process will investigate the data values and explore meaningful insights. One process for data exploration is through descriptive statistics which gives this research the first insight for interpretation or an overview of what the clean raw data look like. This process will also give insight into the potential outlier readings inside the dataset which can be deleted. After doing descriptive statistics, this research will explore some of the questions identified below.

1. What are the relationships between the collected parameters?
2. Are the collected values different considering two thermodynamic conditions?

Answering these questions through the process of data exploratory analysis gives the researcher an initial understanding of what the dataset looks like. The initial analysis taken through the dataset was then generated using visual representation. Data exploratory analysis use python software to generate the visual results.

After the data exploratory analysis, data splitting was followed by dividing the cleaned and explored raw data into a training dataset and a testing dataset. The training dataset was used for the classification training and the testing dataset was used to check how the current generated model was working.

If the model was not performing well, an iterative process of training to develop a pattern recognition model followed by testing dataset checking occurs. The testing dataset will be used to evaluate the classification model's performance. 60% of the data will be taken for training and 40% will be taken for the testing procedure.

E. Data Training and Testing

According to the study of Saad Shaharil Mad et al. [19], an artificial neural network (ANN) was more suitable to be embedded in an IAQ system. The algorithm does not need large storage space unlike the other counterparts and is easy to embed because it requires a less complicated formula. ANN are parallel information processing approaches that are applied for data processing, process analysis, control, fault detection, pattern recognition, defining the complex and non-linear relationship, and employing a number of input-output training patterns from the experimental data [25]. The commonly used neural network architecture is the multilayer feed-forward neural network known as Multilayer Perceptrons or MLP networks which are based on a backpropagation algorithm and comprise multiple hidden layers and neurons. Adding one or more hidden layers creates another set of synaptic connections and more neural interactions which leads to the improvement of the network's accuracy. Fig. 3 shows the general architecture of the multilayer feed-forward neural network for prediction.

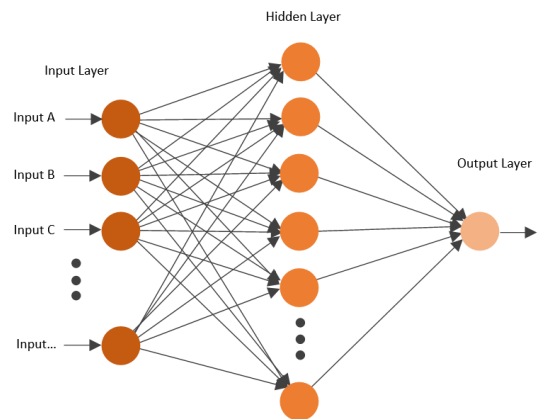


Fig. 3. Sample multilayer perceptron network architecture

Hyperparameters in the study of Saad Shaharil Mad et al. [19] were adopted in this study. One hidden layer was used with three (3) neurons. Vector array normalization was used as feature scaling and an assumption of using stochastic gradient descent was used as the optimizer algorithm since there was no mention in the previous study. Other hyperparameters like the input layer, batch size, kernel initializer, activation function in the last layer, etc. are all defined based on the nature of the classification problem. Using the hyperparameters, training of the dataset will follow to generate the multilayer perceptron model accuracy and loss function. Optimization was then carried out to improve the model with the consideration of overfitting and underfitting. GridSearchCV was then used to find the values of the optimal hyperparameters.

F. Performance Evaluation and Software Details

Evaluation of the performance of the classification model was done through model training and validation metric of accuracy while the loss function was through categorical cross-entropy.

The high-level programming language that was used to implement the methods regarding data cleaning, exploration, normalization, splitting, training, testing, and evaluation was done through Python language. This research utilized the integrated development environment (IDE), Spyder, as the open-source platform for scientific programming in the python language. During the implementation using the platform, software libraries imported such as pandas, sci-kit-learn, Keras, and TensorFlow, etc.

IV. RESULTS AND DISCUSSION

In this study, the IAQ handheld devices were set up in the middle of the room to collect PM2.5, CO2, HCHO, and TVOC. Secondly, temperature and humidity equipment were prepared to achieve two thermodynamic conditions commonly encountered in many climates around the world. During the preparation of the experimental room setup, a specific limitation was found before the implementation of the data collection process. Before injecting a source pollutant, the two thermodynamic conditions must be met first. In the case of the cool and dry condition, room temperature should be at 20C +/- 1C and humidity of 30% +/- 5%. It was found that the room condition can reach the required temperature for the cool and dry condition but can never reach the humidity value of 30% +/- 5%. This limitation was brought about by the experimental room setup which was not completely sealed. This setup simulates the regular room setup where pollutant sources can be found. The final humidity value of 45% +/- 5% was set for cool and dry condition. Finally, adding the limitation above, the simulation of different indoor air pollutants was administered while collecting the data values generated by the devices. This section reports the results and findings in relation to the research question and hypothesis of this study while providing meaning, importance, and relevance of the result.

A. Preliminary Analysis

Initially, data cleaning was performed to the collected raw dataset to ensure correctness and improve data quality. The cleaned raw dataset was processed through preliminary analysis by describing key features of the data. Each parameter for all categorical conditions was correlated to provide insights into the parameter's relationship. Table II shows the individual categorical correlation.

Examining the table above shows that for ambient conditions, it was observable that HCHO has a very high positive correlation with respect to the thermodynamic condition and has a value of 0.92. For combustion conditions, PM2.5 and TVOC has negative high relationship with respect to thermodynamic conditions. For chemical conditions, HCHO has a very high positive correlation with the thermodynamic condition

TABLE II. CATEGORICAL PEARSON CORRELATION

Class 0	Ambient Condition				Class 1	Combustion Condition					
	PM2.5	CO2	HCHO	TVOC		PM2.5	CO2	HCHO	TVOC		
PM2.5	1.000				PM2.5	1.000					
CO2	0.400	1.000			CO2	0.310	1.000				
HCHO	0.087	0.560	1.000		CO2	-0.170	0.055	1.000			
TVOC	0.220	0.480	0.340	1.000	CO2	0.450	0.200	0.300	1.000		
CD/WH	-0.023	0.480	0.920	0.049	1.000	CO2	-0.790	-0.077	0.220	-0.710	1.000
Class 2	Chemical Condition				Class 3	Fragrance Condition					
	PM2.5	CO2	HCHO	TVOC		PM2.5	CO2	HCHO	TVOC	CD/WH	
PM2.5	1.000				PM2.5	1.000					
CO2	0.062	1.000			CO2	0.600	1.000				
HCHO	0.370	0.079	1.000		CO2	0.590	0.690	1.000			
TVOC	0.390	0.320	0.670	1.000	CO2	0.200	0.429	0.680	1.000		
CD/WH	0.390	0.084	0.930	0.590	1.000	CO2	0.430	0.670	0.800	0.830	1.000
Class 4	Rotten Food Condition										
	PM2.5	CO2	HCHO	TVOC	CD/WH						
PM2.5	1.000										
CO2	0.410	1.000									
HCHO	-0.130	0.020	1.000								
TVOC	0.110	0.190	0.960	1.000							
CD/WH	-0.200	-0.059	0.960	0.910	1.000						

and a moderate correlation value of 0.59 with respect to TVOC. For the fragrance conditions, most of the parameters has a moderate to very high positive correlation with respect to each other. Lastly for rotten food conditions, HCHO and TVOC has a high positive correlation with thermodynamic conditions.

The categorical dataset was then merged to provide insight to have the overall Pearson correlation of the dataset's parameters including the categorical conditions. Fig. 4 shows the heatmap generated through seaborn in python.



Fig. 4. Overall correlation heatmap

Firstly, the figure above shows that PM2.5 has a high positive correlation value with respect to CO2 which measures a positive correlation of 0.77 which verifies the result in most of the IAP conditions having a significant correlation value. The same positive correlational value was also observed for HCHO and TVOC parameters. Strong correlation was found for HCHO and TVOC since HCHO was part of the parameters to be collected in TVOC device. Secondly, CO2 has negligible correlational value with respect to the thermodynamic condition and has a very minimal negative correlation with respect to TVOC and HCHO with the value of 0.37 and 0.2 respectively. The result of CO2 is also verified using the correlation table of the individual IAP condition. Thirdly, the thermodynamic condition has a low correlational value to HCHO of positive 0.32 which verifies as well in the individual IAP condition correlation table. Lastly, the overall correlation of the IAP conditions has a very low negative correlation with respect to PM2.5 and CO2 with negligible correlation to other variables.

B. Accuracy and Loss Comparison between Distinct Scaling Technique and Optimizer Algorithm

After the preliminary analysis, model training was initiated. The study of Saad Shaharil Mad et al. [19] model training algorithm motivates this study in adopting multilayer perceptron as the only machine learning algorithm and adopted some of the hyperparameters used in the previous study. Table III shows the adopted hyperparameters and using additional hyperparameters to carry out this study.

TABLE III. MULTILAYER PERCEPTRON HYPERPARAMETERS

Adopted Hyperparameters	
Hidden Layer	1
Neurons	3
Output Layer	5
Learning Rate	0.01
Momentum Constant	0.5
Epochs	1000
Additional Hyperparameters	
New Input Layer	5
Batch Size	64 (based on a paper)
Feature Scaling	VAN, Standardization
Optimizer Algorithm	SGD, Adam
Hidden Layer Activation	Relu (default on Keras)
Kernel Initializer	glorot_uniform (default on Keras)
Output Layer Activation	Softmax
Metric	Accuracy
Loss Function	Categorical Cross Entropy

This study addresses the hypothesis by using only 5 IAQ parameters in predicting the different classes of indoor air pollution, thus, setting the input layer equal to 5. The hidden layer, neurons, output layer, learning rate, and momentum constant were directly adopted in generating the model accuracy and loss. Additional hyperparameters were set to successfully train the data using Multilayer Perceptron. This study uses a fixed batch size of 64, in reference to the recommendation of the study of Kandel, Ibrahim et al [26]. For feature scaling, Vector Array Normalization was seen to be the best performer in the previous research, and standardization was not included for comparison. This study has drawn insights into the difference in results between using VAN and Standardization (STD) technique. Optimizer Algorithm was never mentioned in the previous research; thus, this study assumes and adopted the use of stochastic gradient descent (SGD) as the optimizer for this research. Also, this study has explored the Adam optimizer since this optimizer is always compared with SGD in many papers. For this comparison setup, this study has used 'relu' for the hidden layer activation, 'glorot_uniform' for the kernel initializer, and 'softmax' for the output layer activation function. The two former variables were considered as default hyperparameter of Keras for model training. 'Softmax' was used as the output layer activation function since this is an example of a multiclass classification problem.

After setting up the hyperparameters, model training starts. Accuracy and loss function results was compared using the combination of two feature scaling and two optimizer algorithms.

Given Fig. 5, in using VAN as the feature scaling and SGD as the optimizer, it can be observe that the accuracy output

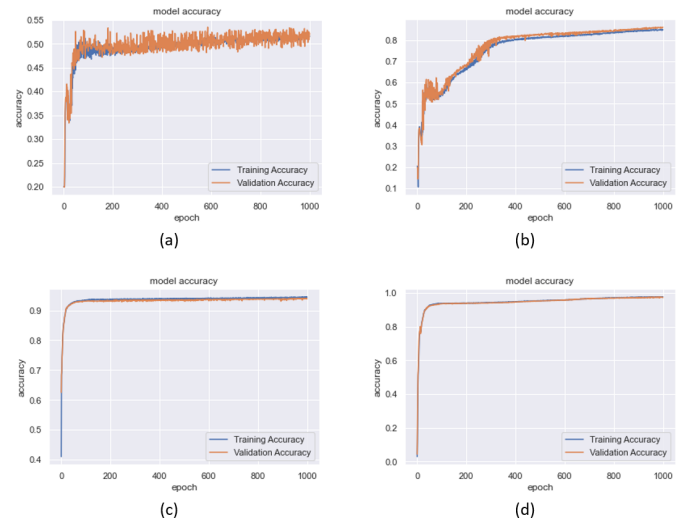


Fig. 5. Model training and validation accuracy (a) VAN-SGD (b) VAN-Adam (c) STD-SGD (d) STD-Adam

was noisy, and the training process was slow reaching to a validation accuracy of 52% for 1000 epochs. Using the same feature scaling while changing SGD to Adam optimizer, the accuracy results are less noisy, and the validation accuracy improves reaching 86% for 1000 epoch. Using standardization (STD) as the feature scaling while varying the optimizer, the result greatly improves. With SGD as the optimizer, the validation accuracy incurs a value of 93.8% compared to Adam optimizer with a value of 97% for 1000 epochs.

Fig. 6 shows the model loss comparison. Model loss in VAN feature scaling was observed to be bigger compared to the feature scaling of standardization after 1000 epochs. VAN-Adam has better loss value compared to VAN-SGD with 0.34 and 0.99 validation loss respectively. Both optimizers using STD as the feature scaling reaches a loss value of less than 0.2 in 1000 epochs. STD-SGD incur a validation loss of 0.16 compared to STD-Adam with a validation of 0.07.

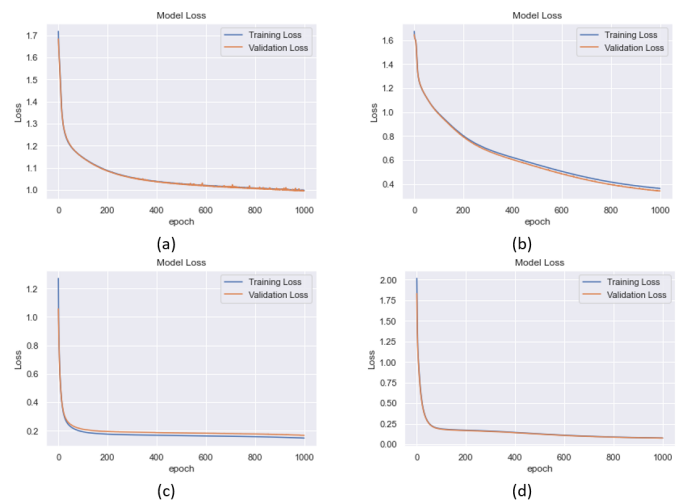


Fig. 6. Model training and validation loss (a) VAN-SGD (b) VAN-Adam (c) STD-SGD (d) STD-Adam

C. Hyperparameter Tuning

Referring to the comparison results in the above section, the accuracy in using STD and Adam can be seen as good fit already. Even if the result was good, it can be observed that some of the hyperparameters was not tuned with the hypothesis of gaining the optimized accuracy and loss value. To further improve the results, tuning of different hyperparameters was investigated. Hyperparameters like number of input layers, number of hidden layers, number of neurons, number of output layers, output layer activation function, metrics and loss function were preset because of the nature of the dataset. The result in the previous section in using standardization as the feature scaling and Adam as the optimizer algorithm was also included as a preset value for the tuning process. Momentum was excluded in this process because of the new optimizer. It was also observed that in using STD as feature scaling, 1000 epochs was too much of a training period and eventually just limits the number of epochs to 50. Table IV shows the validation accuracy and loss function for the different epochs. The table shows that the use of standardization (STD) feature scaling is superior compared to vector array normalization (VAN) in this dataset. It is observable that even if the epoch was 50, using standardization still exhibits a good accuracy and loss value.

TABLE IV. VALIDATION RESULT PER EPOCHS

	Validation Result					
	1000 epochs		100 epochs		50 epochs	
	Accuracy	Loss	Accuracy	Loss	Accuracy	Loss
VAN-SGD	0.5216	0.9952	0.4794	1.1598	0.4711	1.2210
VAN-Adam	0.8617	0.3420	0.5706	1.0109	0.4599	1.1576
STD-SGD	0.9383	0.1169	0.9284	0.2206	0.9102	0.2902
STD-Adam	0.9711	0.0750	0.9362	0.1837	0.9255	0.2475

Taking into consideration the execution time of model training, hyperparameters like learning rate and batch size was varied. Other hyperparameters like hidden layer activation and kernel initializer was also varied to find the best possible combination of hyperparameters. Table V shows the summary of hyperparameters.

TABLE V. NEW SET OF HYPERPARAMETERS

Hyperparameters	
Input Layer	5
Hidden Layer	1
Neurons	3
Output Layer	5
Learning Rate	0.1, 0.01, 0.001
Epochs	50
Batch Size	32, 64, 128
Feature Scaling	Standardization (STD)
Optimizer Algorithm	Adam
Hidden Layer Activation	Tanh, Relu, Sigmoid
Kernel Initializer	glorot_uniform, uniform, he_uniform
Output Layer Activation	Softmax
Metric	Accuracy
Loss Function	Categorical Cross Entropy

Given all the hyperparameters above, model training was initiated and using python's gridsearchCV to exhaustively

search for the best fit combination of hyperparameters. Cross validation (CV) for this process was set to 5.

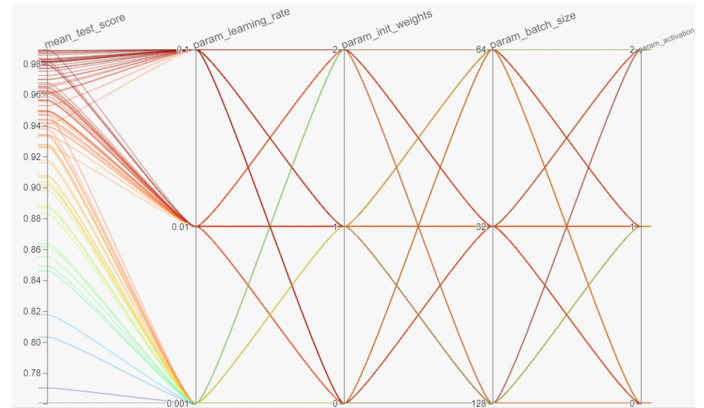


Fig. 7. GridSearchCV hyperparameter combinations

Fig. 7 shows the different hyperparameter combinations with the mean test score value where mean test score is the mean accuracy of the classification. Using STD and Adam optimizer as the new hyperparameter, it is shown that all hyperparameter combination exhibits the value of mean test score greater than 70%. It also shows that there are hyperparameter combinations that reaches more than 99% of mean test score.

Table VI shows top 10 performing hyperparameter combinations with respect to the mean test score with mean fit time

TABLE VI. TOP PERFORMING HYPERPARAMETER COMBINATION BASED-ON MEAN TEST SCORE

Activation	Batch Size	Initial Weights	Learning Rate	Mean Test Score	Mean Fit Time
Relu	32	be_uniform	0.01	0.98924	15.98910
Sigmoid	64	glorot_uniform	0.1	0.98854	8.60077
Relu	64	Uniform	0.1	0.98819	8.35799
Sigmoid	32	glorot_uniform	0.1	0.98901	15.08160
Sigmoid	32	be_uniform	0.1	0.98698	14.59695
Sigmoid	32	Uniform	0.1	0.98681	15.14623
Tanh	128	glorot_uniform	0.1	0.98559	4.98741
Relu	128	glorot_uniform	0.1	0.98333	4.68067
Sigmoid	128	glorot_uniform	0.1	0.98264	4.75186
Relu	128	be_uniform	0.1	0.98264	4.82595

It was observable that rank 1 combination garnered a mean test score of 98.9 percent while having a mean fit time of 15.878 seconds. Succeeding combinations have good mean test score values but has longer mean fit time. The mean fit time is the average time of training between cross validation folds. The correlation figure below shows the relationship between mean test score and mean fit time with the varying hyperparameters.

TABLE VII. CORRELATION OF VARYING HYPERPARAMETERS WITH MEAN TEST SCORE AND MEAN FIT TIME

	Activation	Batch Size	Initial Weights	Learning Rate
Mean Test Score	-0.1963	-0.2720	0.0035	0.5749
Mean Fit Time	0.0212	-0.9222	0.0194	-0.0091

Table VII shows the correlation of varying hyperparameter results. Batch size can be seen to have a high negative correlation with respect to the mean fit time which gives an idea that training period is greatly influence by the value of batch size. The bigger the batch size, the lesser the mean fit time and the lower the batch size, the bigger the mean fit time. Negligible relationship can be seen through learning rate, activation function and initial weights with respect to the mean fit time. Moderate positive correlation value was found from mean test score with respect to learning rate and very low to negligible relationship can be seen from mean test score with respect to hidden layer activation and batch size and initial weight.

D. Accuracy and Loss using the Optimal Hyperparameter based on Mean Test Score

Using the optimal value of hyperparameters based on mean test score, model training was again initiated to get the accuracy and loss function. Fig. 8 shows the model training and validation results.



Fig. 8. Model training and validation using the optimal hyperparameter(a) accuracy (b) loss

Taking into consideration Table IV and Fig. 8, it is observable that the accuracy and loss function in 50 epochs greatly improves using the tuned hyperparameters in comparison to old hyperparameters. Accuracy of the old hyperparameters in 50 epochs was 92.55% while the new hyperparameter incurred a value 98.9%. Loss function for old hyperparameter in 50 epochs was 0.2475 in comparison of the new hyperparameter with a value of 0.0986.

V. CONCLUSION

This research presents the process of classifying indoor air pollutant sources using relevant parameters under two thermodynamic conditions. The research question that was formulated at the start of this study was finally addressed using this research. Model accuracy and loss function values were seen to be in good fit which is comparable to the previous study. From nine parameters, this research streamlines the number of parameters into five considering CO₂, PM_{2.5}, TVOC, HCHO, and Thermodynamic Conditions. Using the dataset generated in this study feature scaling was compared using normalization and standardization. Standardization scaling was shown to be superior in relation to the vector array normalization scaling which was used in the previous study. The hyperparameters

used in the previous study were also taken into consideration for this research to have the optimal values of hyperparameters. It can be observed that hyperparameters should be tuned to gain optimal result value. Computational time was greatly reduced at the same time giving the best-fit accuracy and loss function values without incurring the problem of overfitting.

The current study classifies indoor pollutants with a minimal number of sensors but can only classify one pollutant source at a time. The previous study collected a dataset on a single source and tested the acquired model on mixed pollutant sources but relatively gains poor performance. Considering this limitation, further dataset gathering and investigation on mixed pollutant sources should be examined while taking into consideration hyperparameter tuning.

ACKNOWLEDGMENT

The researcher would like to thank Central Visayas Studies Center (CVSC) of University of the Phillipines Cebu for funding this research. Also, special thanks to Lovelie and Elle for being an inspiration.

REFERENCES

- [1] P. O. Fanger, *What is IAQ?*, DOI: 10.1111/j.1600-0668.2006.00437.x, 2006
- [2] R. Pitarma, G. Marques and B. R. Ferreira, *Monitoring Indoor Air Quality for Enhanced Occupational Health*, J Med Syst 41, 23 (2017). <https://doi.org/10.1007/s10916-016-0667-2>
- [3] M. J. Jafari, A. A. Khajevandi, S. IM. Najarkola, M. S. Yekaninejad, M. A. Pourhoseingholi, L. Omidi, S. Kalantary, *Association of Sick Building Syndrome with Indoor Air Parameters*, 2015; 14(1):55-62. PMID: 26221153; PMCID: PMC4515331.
- [4] United States Environmental Protection Agency (USEPA), *emphIndoor Air Quality*, <https://www.epa.gov/indoor-air-quality-iaq/introduction-indoor-air-quality>
- [5] J. Saini, M. Dutta, G. Marques, *A comprehensive review on indoor air quality monitoring systems for enhanced public health*, Sustain Environ Res 30, 6 (2020). <https://doi.org/10.1186/s42834-020-0047-y>
- [6] V. V. Tran, D. Park, Y. C. Lee, *Indoor Air Pollution, Related Human Diseases, and Recent Trends in the Control and Improvement of Indoor Air Quality*, Int J Environ Res Public Health. 2020;17(8):2927. Published 2020 Apr 23. doi:10.3390/ijerph17082927
- [7] World Health Organization, Regional Office for Europe, *Air quality guidelines for Europe, 2nd ed.. World Health Organization*, <https://apps.who.int/iris/handle/10665/107335>, 2000
- [8] M. Mannan, S. G. Al-Ghamdi, *Indoor Air Quality in Buildings: A Comprehensive Review on the factors Influencing Air Pollution in Residential and Commercial Structure*, Int. J. Environ. Res. Public Health 2021, 18(6), 3276; <https://doi.org/10.3390/ijerph18063276>
- [9] M. Tastan, H. Gokosan, *Real-time Monitoring of Indoor Air Quality with Internet of Things- Based E-nose*, Appl. Sci. 2019, 9(16), 3435; <https://doi.org/10.3390/app9163435>
- [10] G. Chiesa, S. Cesari, N. Garcia, M. Issa, S. Li, *Multisensor IoT Platform for Optimising IAQ levels in Buildings Through a Smart Ventilation System*, Sustainability 2019, 11(20), 5777; <https://doi.org/10.3390/su11205777>
- [11] I. Demanega, I. Mujan, B. C. Singer, A. S. Andelkovic, F. Babich, D. Licina, *Performance assessment of low-cost environmental monitors and single sensor under variable indoor air quality and thermal conditions*, 2021, Building and Environment 187 (2021) 107415
- [12] Z. Wang, W. W. Delp, B. C. Singer, *Performance of low cost Indoor air quality monitors for PM2.5 and PM10 from residential sources*, <https://doi.org/10.1016/j.buildenv.2020.106654>
- [13] W. Yang, J. Sohn, J. Kim, B. Son, J. Park, *Indoor air quality investigation according to age of the school buildings in Korea* Journal of Environmental Management 90 (2009) 348e354

- [14] C. Mandin, M. Trantallidi, A. Cattaneo, N. Canha, V. G. Mihucz, T. Szigeti, R. Mabilia, E. Perreca, A. Spinazzè, S. Fos-sati, Y. Kluzenaar, E. Cornelissen, I. Sakellaris, D. Saraga, O. Hänninen, E. O. Fernandes, G. Ventura, P. Wolkoff, P. Car-ner, J. Bartzis, *Assessment of indoor air quality in office build-ings across Europe – The OFFICAIR study*, Science of The Total Environment, Volume 579, 2017, Pages 169-178, ISSN 0048-9697, <https://doi.org/10.1016/j.scitotenv.2016.10.238>.
- [15] A. Datta, R. Suresh, A. Gupta, D. Singh, P. Kulshrestha, *Indoor air quality of non-residential urban buildings in Delhi, India*, International Journal of Sustainable Built Environment, Volume 6, Issue 2, 2017, Pages 412-420, ISSN 2212-6090, <https://doi.org/10.1016/j.ijse.2017.07.005>.
- [16] D. Saraga, T. Maggo, E. Sadoun, E. Fthenou, H. Hassan, V. Tsiouri, S. Karavoltos, A. Sakellari, C. Vasilakos, K. Kakosimos, *Chemical Characterization of Indoor and Outdoor Particulate Matter (PM_{2.5}, PM₁₀) in Doha Qatar*, <https://doi.org/10.4209/aaqr.2016.05.0198>
- [17] M. Castelli, F. M. Clemente, A. Popovič, S. Silva, L. Van-neschi, *A Machine Learning Approach to Predict Air Quality in Cal-ifornia*, Complexity, vol. 2020, Article ID 8049504, 23 pages, 2020. <https://doi.org/10.1155/2020/8049504>
- [18] W. Wei, O. Ramalho, L. Malingre, S. Sivanantham, J. C. Little, C. Mandin *Machine learning and statistical models for predicting indoor air quality*. Indoor Air. 2019 Sep;29(5):704-726. doi: 10.1111/ina.12580. Epub 2019 Jul 25. PMID: 31220370.
- [19] S. M. Saad, A. M. Andrew, M. Shakaff, M. Dzahir, M. Hussein, M. Mohamad, Z. A. Ahmad, *Pollutant Recognition Based on Supervised Machine Learning for Indoor Air Quality Monitoring Systems* Appl. Sci. 2017, 7, 823. <https://doi.org/10.3390/app7080823>
- [20] J. Kallio, J. Tervonen, P. Räsänen, R. Mäkynen, J. Koivusaari, J. Peltola, *Forecasting office indoor CO₂ concentration using machine learning with a one-year dataset*, Building and Environment, Volume 187, 2021, 107409, ISSN 0360-1323, <https://doi.org/10.1016/j.buildenv.2020.107409>.
- [21] H. Zhang, R. Srinivasan, V. Ganesan, *Low-Cost , Multi-Pollutant Sensing System Using Raspberry Pi for Indoor Air Quality Monitoring*, Sustainability 2021, 13, 370. <https://doi.org/10.3390/su13010370>
- [22] American Society of Heating Refrigerating and Air-Conditioning En-gineers - ASHRAE, *Thermal environmental conditions for human occu-pancy*, ANSI/ASHRAE Stand. - 55. 7 (2017) 6.
- [23] S. Chen, K. Mihara, J. Wen, *Time series prediction of CO₂, TVOC and HCHO based on machine learning at different sampling points*, Building and Environment, Volume 146, 2018, Pages 238-246, ISSN 0360-1323, <https://doi.org/10.1016/j.buildenv.2018.09.054>.
- [24] X. Xiangguo, S. Chao, D. Shiming, *Using two parallel connected compressors to implement a novel control algorithm for improved indoor humidity at a low cost* Building Services Engineering Research and Technology. 2013;34(3):349-354. doi:10.1177/0143624412446092
- [25] A. Rostami, M. Anbaz, H. Gahrooei, M. Arabloo, A. Bahadori, *Accurate estimation of CO₂ adsorption on activated carbon with multi-layer feed-forward neural network (MLFNN) algorithm*, Egyptian Journal of Petroleum, Volume 27, Issue 1, 2018, Pages 65-73, ISSN 1110-0621, <https://doi.org/10.1016/j.ejpe.2017.01.003>.
- [26] I. Kandel, M. Castelli, *The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset*, ICT Express, Volume 6, Issue 4, 2020, Pages 312-315, ISSN 2405-9595, <https://doi.org/10.1016/j.ict.2020.04.010>.