Beyond the Horizon: A Meticulous Analysis of Clinical Decision-Making Practices

Bilal Saeed Raja[1], Sohail Asghar[2]

Department of Computer Science

COMSATS University Islamabad, Pakistan

Abstract—**Clinical advancements are one of the major outcomes of the technological phase shift of data sciences. The significance of information technology in medical sciences by utilizing the Clinical Decision Support System (CDSS) has opened the spillways of exponentially improved predictive models. Utilizing the latest norms of classification algorithms on clinical data are widely incorporated for prognostic assessments. Medical experts have to make decisions that are crucial in nature and if the research can develop a mechanism that assists them in evolving solid reasoning, infer the knowledge and clearly express their clinical decision by justifying their assertions made, it will be a win-win situation. However, this field of science is still an unknown world for clinicians despite the fact that the enormous amount of medical data cannot be exploited to its maximum without invoking the technological support. The objective of this research is to introduce the clinicians and policymakers of the medical domain with the renowned computer-based methodologies employed to construct a clinical decision support system. We expect that gaining the technical insight into the medical domain by the stakeholders will ensure commissioning the accurate and effective CDSS for improved healthcare delivery.**

Keywords—*Decision support system; clinical decision support; classification; clustering; association rule mining; multi-objective evolutionary optimization*

I. INTRODUCTION

Decision Support Systems (DSS) are the most studied areas of data sciences and their widespread adoption has earned standing in multiple domain such as education sector [48], customer relationship management [34], fraud detection in financial matters [33], detection of eavesdroppers and intruders in networks [39] and health care [6] including genetic programming [19]. Technical advancements have given a new dimension to the automated decision-making capability and health care is no exception to this due to its importance and critical nature as humans are the direct beneficiary or affectees of the outcome. Automation has given great ease to handle and query the huge volumes of data however it is an uphill task to process this data manually due to its size and non-homogeneity. Clinical Decision Support Systems (CDSS) are the most researched models as compared to any other science domain due to their effective decision-making capability. CDSS can be described as an "extraction of implicit potentially useful and novel information from medical data to improve accuracy, decrease time and cost, construct decision support system with the aim of health promotion" [15]. Due to the subjective nature decisions made by the domain experts may add inaccuracies that can be minimized by utilizing the technology to increase effective decision-making [67]. Furthermore, technological induction in health care has the strength to

extract relationships within variables, identify the factors that may cause various risks and further impart fresh knowledge to yield befitting precision augmented with a convincing reasons [29]. This can only be achieved if the policymakers and clinicians have a deep insight into the technical strengths and weaknesses of computer-based methodologies being employed to construct CDSS. These decision support systems must be assessed for their performance evaluation so that quality care be provided with high precision value [36]. Data gathering and pre-processing is always a tough choice in different fields of science but it becomes much harder when it comes to health care due to its critical nature [11]. The medical data is not only numeric but may include images, temporal & combination of these which makes it a tough choice for automated decision-making. Medical data has huge volumes and its handling and organization in a manner that is understandable to the clinical decision support model is another daunting task. Therefore, the selection of appropriate classification schemes must be opted to handle multifaceted data for better throughput. In a broader scene, clinical and temporal data are the major categorizations in the health care domain [25]. This data is either quantitative (numeric), qualitative (non-numeric), temporal data (based on timestamp) or time-series based.

The major contribution of this research is to highlight the latest and most widely used decision-making models and techniques that enlightens the reader to earmark the pros and cons of these predictive models at the early stage of development. Furthermore, the research effort is to educate technical as well as non-technical domain specific audience that makes it an interesting academic resource by imparting better understanding of decision support methodologies that can be applied from analysis, design, development and deployment phases of CDSS. The SWOT analysis covers the socio-behavioural aspect of commissioning these models in medical domain.

Because of the non-homogeneous and varied characteristics of clinical data and keeping in view the importance of clinical decisions, selection of suitable and befitting methodology must be adopted by the stakeholders for assistive CDSS. An additional yet very crucial deliberation must be pursued on the selection of CDSS based on the operational characteristics that are as follows [16]:

- Trigger-based which are mainly used for drug prescription.

- Data Repository such as patient records & patient's pathological results.

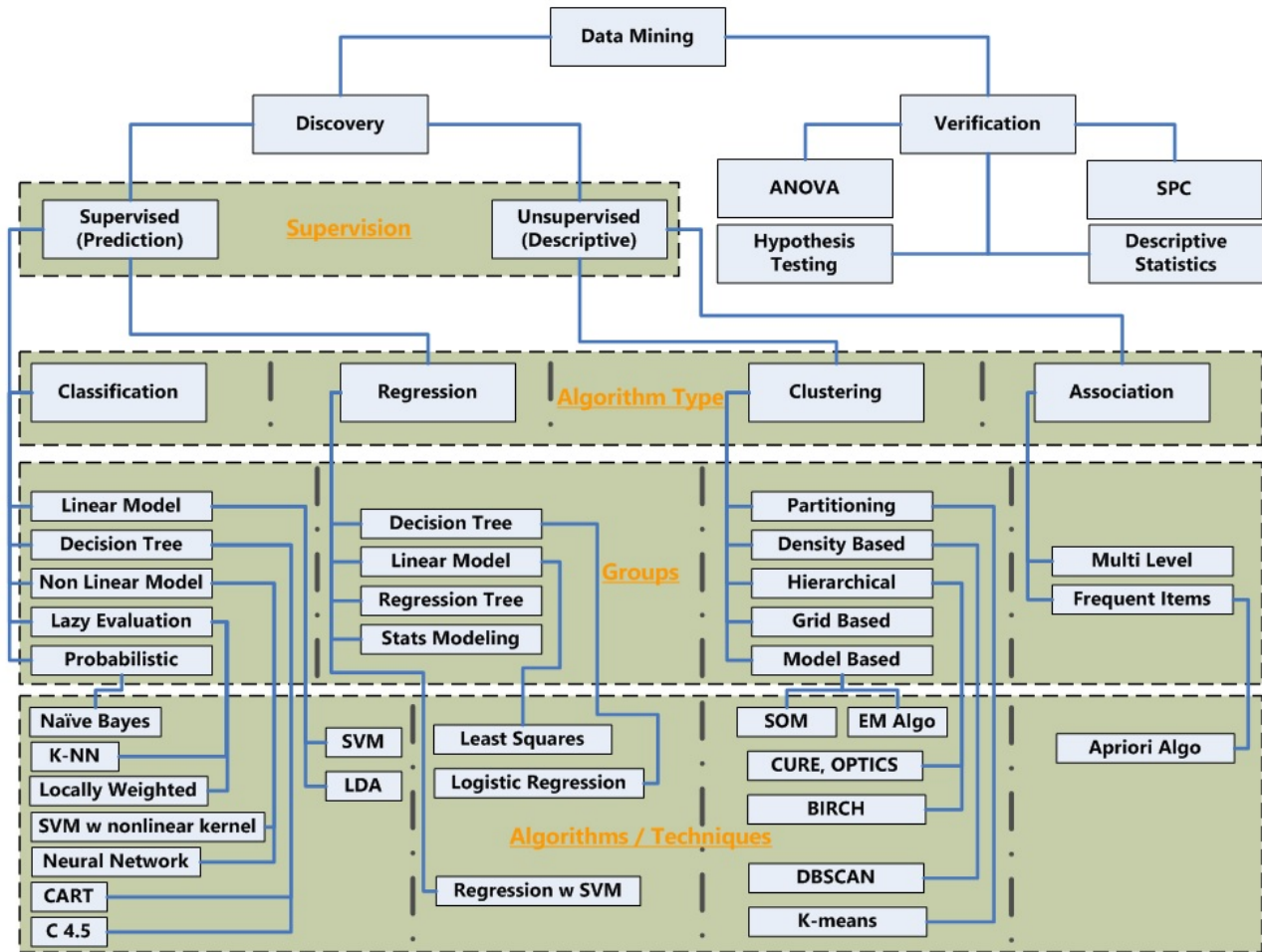- Interventions systems that send alert messages to

Fig. 1. Taxonomy of data mining algorithms

clinicians.

- Offered Choices in prescribing treatments and medication.

The automatic discovery of intelligent knowledge from the clinical decision support system commissions the major science of information technology that is data mining. Data mining is the branch of business intelligence that has two major branches one is the discovery and the other is verification. Disease classification and clinical decision support systems rely on data mining algorithms and these algorithms play a vital role in harnessing accurate and interpretable clinical decisions. Based on the available literature, taxonomy is presented in Figure 1.

II. METHOD OF LITERATURE REVIEW

Substantial research has been carried out so far to acquire clinical decisions based on a multi-faceted web of knowledge. Most relevant and current research content was extracted from the web of knowledge. The search term used to search the literature was "clinical decision support systems, disease prediction, decision making, and classification algorithms". The search revolved around the most relevant journals that were primarily focusing on the medical and bioinformatics

domain (Artificial Intelligence in Medicine Elsevier, Computational and Structural Biotechnology Journal Elsevier. BMC Neurology Springer. Journal of Biomedical Informatics, Health Policy and HPT, Elsevier, etc.) from the period 2000 to date.

In quest of the resolution to the research questions in mind, relevant research articles on disease prediction/classification were collected from Google Scholar, IEEE Xplore, Springer-Link, ACM, DBLP ISI Web of Knowledge to name a few. Initially, we selected 209 research articles from which we studied and included the research contribution of 56 most relevant articles. The major focus was to include the latest research of the domain and with few exceptions, we confined our survey to the year 2010. This comprehensive review is carried out to shortlist and include the most recent and suitable articles of clinical sciences with high-quality research. The methodology of the literature review and count of research paper that are studied for compiling this research is presented in Figure 2.

To conduct this survey, a systematic literature review methodology is adopted. Furthermore, a brief description of the disease prediction/classification scheme is presented followed by strengths and weaknesses. This methodology is adopted by keeping one aspect in mind that the reader should have a clear understanding of the pros and cons of the classification
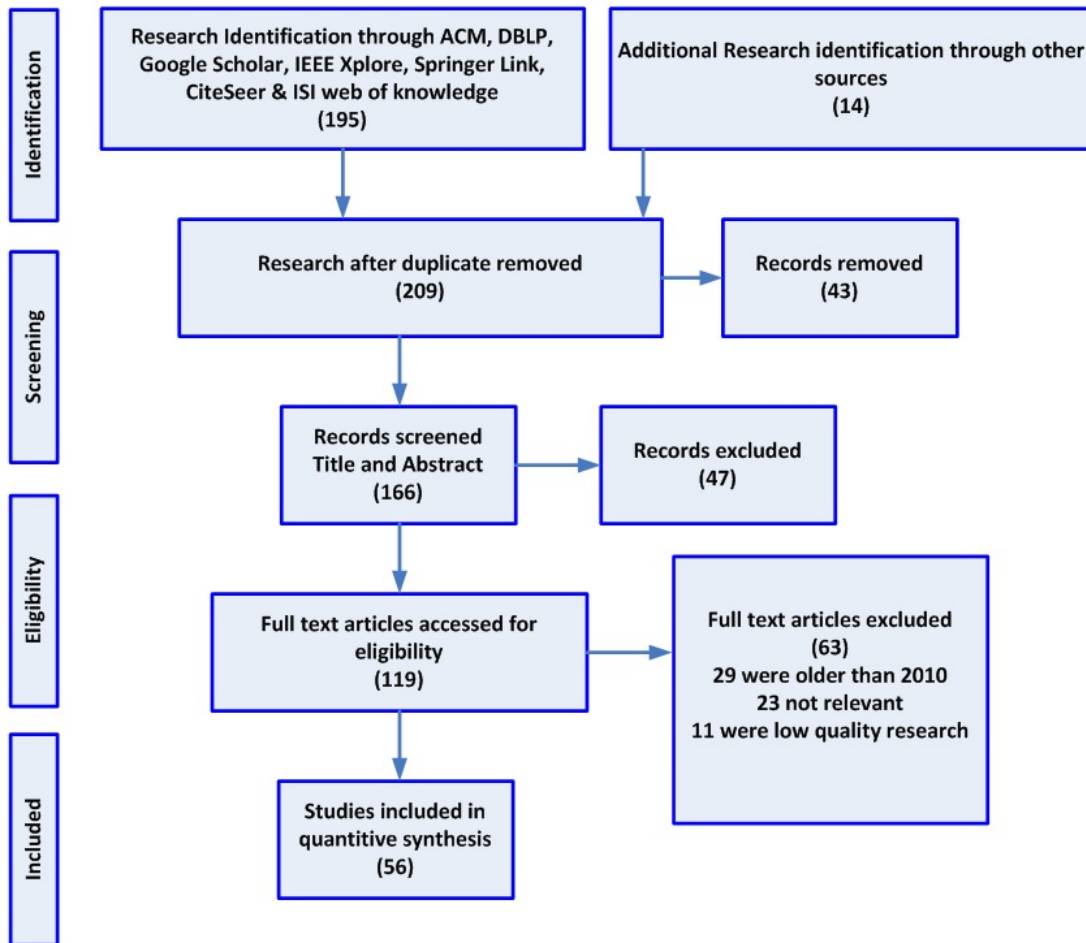
Fig. 2. Methodology of Literature Review

scheme. By doing so, policymakers can reckon very easily about the classification scheme being used during the development of CDSS.

A. Research Questions

Following questions were kept in mind while conducting this survey:

- What are the existing disease prediction techniques/algorithms?

- What kind of methodology is adopted in these techniques?

- What are the effect/repercussions of the findings regarding techniques is going to have while developing new CDSS?

- What are the major shortcomings/bottlenecks limitations of these techniques?

- What are the latest development methodologies in place to evolve effective CDSS?

In the upcoming sections, we will explain the various techniques and methodologies like machine learning, knowledge

representation, text mining and multi-objective optimization techniques by which we can evolve an effective and efficient disease classification and clinical decision support system. It should also be considered that these methodologies can be used to evolve any type of CDSS and disease classification system elaborated in the research above. However, the selection of suitable and optimum classifier is a very important and crucial decision. Since all of the methodologies cannot be elaborated in detail so most relevant and famous are explained. The rest of the paper is organized to dovetail relevant methodologies in their categorical class and comprehensive yet most recent literature is organized in a fashion that suits almost all kinds of audience. This will not only help the technical team developing the CDSS but will also help the clinicians and decision-makers of health care facility to have prior wisdom in selecting appropriate solution at the very initial stage.

III. MACHINE LEARNING

Machine learning is primarily based on the learning curve extracted from the data [20]. The main objective applied in machine learning is data cleansing in which missing values are identified and outliers are removed. After the data cleansing process training of classifier is carried out on the dataset. The prime objective of this training is to make the classifier intelligent enough to perform requisite prediction and classification
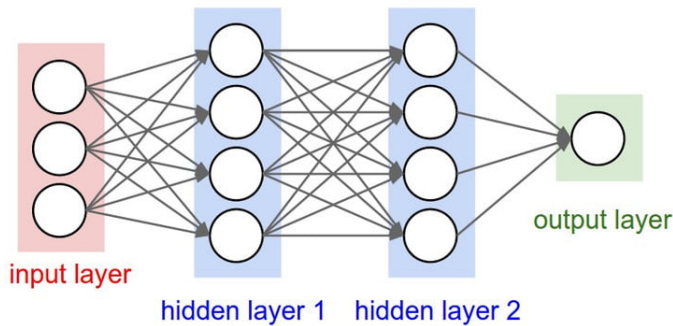
Fig. 3. Artificial Neural Network

Fig. 4. Support Vector Machine

on clinical data. The efficiency and accuracy of the classifier depend on the training parameters. The system needs to be trained in a manner that it should not be over or under trained. As there is no cookbook and one solution fit for all kinds of things, the training should be carried out wisely. There is every likelihood that for one clinical problem a classifier works well but for another, the same may fail to converge in a desirable manner. Once the training phase is over new dataset is given to the classifier for prediction and classification purposes. In the next subsection, we will elaborate on the most common type of machine learning methodologies.

A. Artificial Neural Networks (ANNs)

Artificial neural networks are mathematical models that mimic the human brain in the learning process [22]. To find the hidden pattern in the data and classification ANN uses a complex network of artificial neurons that are interconnected. ANN has an input layer and an output layer and most important of all, the hidden layer (one or more). The input layer takes some input and the output layer generates one or more than one outputs while the hidden layer trains itself to learn the hidden patterns from the data. The whole scenario is depicted in Figure 3.

The latest research on the validation and discovery of Alzheimer's disease was carried out in [68], where an artificial neural network pipeline was used for the said task. The research is a major contribution to the domain of dementia, which has no established cause and cure. Pipeline analyzes a public dataset with a continuous and categorical algorithm and further infers through network inference to generate novel markers for disease prediction. However, the research lacked performance and interpretability indices due to overfitting and noisy data. Another research [3] predicts the Dengue virus by implementing an artificial neural network on patient data for disease classification. Five performance objectives were employed including mean absolute error and accuracy. The research outnumbers the CART algorithm, however, improvement can be acquired if iterative non-parametric imputations are applied. ANN-based research was produced in [7] that used an artificial neural network (ANN) and Fuzzy analytic hierarchy process (Fuzzy_AHP) for an integrated decision support system. This hybrid approach got limitation for the automatic calculation of weights for the attributes which are used to train the ANN classifier.
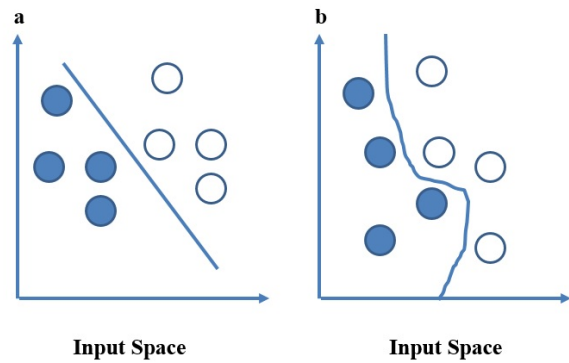
Artificial neural network like other computer methodologies has pros and cons. The ANN algorithm does not require rigorous training and is simple to implement with an ability to detect a non-linear relationship between variables. However, their hidden layer is not transparent to the user and has the tendency to over-fit when there are some noise and error in the data.

B. Support Vector Machine (SVM)

The basic theme behind the support vector machine resembles artificial neural networks, but SVM is much more powerful and effective when the classification of a complex task is required. The values from the dataset under classification are presented in the form of a point in a solution space [32]. The whole idea behind the support vector machine is that the data is projected in the solution space in the form of point and the point belongs to a designated class. For example, a dataset of disease classification will depict the presence or absence of a particular disease when the data is mapped on the same solution space. The data value will become part of one of the classes based on which side of the hyperplane it falls in. Figure 4 (a) depicts a linear classifier in which ANN can perform well, however, 4 (b) depicts a nonlinear classifier in which ANN cannot perform well and the best choice to solve such problem is support vector machine.

The novel research was presented in [31] where Parkinson's disorder was classified by diffusing proton spectroscopy, tensor imaging, and photometric data to obtain quantitative markers for the consumption of SVM. To achieve high accuracy a new graph-based technique was commissioned but the research lacked its full potential due to the fact that small data set with classifiable Parkinsonism was considered. Another research [56] classified healthy and Asthmatic individuals with the help of electronic nose through gas emission in exhaled breath by applying SVM in feature extraction and classification.

The system showed 78.8% accuracy when non-linear binary SVM was used instead of linear with a high rate of sensitivity. Adaptive SVM was commissioned for the first time to diagnose chest disease with high precision value by computing the appropriate bias term value to SVM [66]. In [58] a support vector machine and radial base function network structure is presented to predict the heart disease in the patients but a uni classifier is used in the research. [40] Proposed an

automatic logistic regression and support vector machine-based prediction and classification for Parkinson's disease. The support vector machine along with Radial Basis Function (RBF) kernel achieves more accurate classification however results can further be improved if the ensemble framework is commissioned instead of uni-classifier. [62] Used various classification techniques for breast cancer prediction but the classification of support vector machine outnumbered all of them in performance. But this research has the same limitation of using a single classifier.

SVM's perform well when less training data is available and also suitable for multi-dimensional data which is un-balanced in nature. The algorithm hides the internal details of the working methodology but performs well by using mathematical modeling on the unstable dataset.

IV. Knowledge Representation

Knowledge Representation is principally based on vocabulary and in the clinical domain it generates a clinical knowledge descriptive language. This vocabulary is comprehensible and exploitable by the computer system and an amalgamation of an automated reasoning and inference system is formed [8]. In the medical domain patient's data can constitute vocabulary that can be automatically reasoned using clinical practicing guidelines to infer about patient's health.

A. Fuzzy Logic

Fuzzy logic is based on a probabilistic model that enumerates the human reasoning in approximate values [63]. The results generated are not just true and false but also enlightens the end-user with the degree of truth and false. Thus the results of fuzzy rules are more accurate because of their non-discrete nature. The fuzzy rule set is simple if-then-else rules that can be apprehended from natural language. Fuzzy logic has a Knowledge base that is built on the combination of Rule-base and database and a Fuzzy Inference Engine (FIE) as shown in Figure 5. Following lines express the rules for the fuzzy system:

$$R_i : \textbf{IF } x_1 \textit{ is } F_1 \textbf{ AND } x_2 \textit{ is } F_2$$
$$\textbf{AND } x_m \textit{ is } F_m \textbf{ THEN } y = C_i \quad (1)$$
$$\textbf{\textit{WHERE i=1,2,3,4, ...,M}}$$

Very important research carried out in [35] integrates the fuzzy standard additive model (SAM) with a genetic algorithm (GA), called GSAM was adopted and wavelet transformation was employed to extract discriminative features for high-dimensional datasets. GA was used to optimize the number of fuzzy rules before supervised learning. GSAM dominates PNN, SVM & ANFIS on classification accuracy but has disadvantage regarding computational cost compared to these competitive methods. Fuzzy logic was amalgamated with the modular neural network to diagnose the risks of hypertension [30]. The age, risk, and blood pressure were a major deriving force of the research and modular neural network utilized three modules, one looking after heart rate and the remaining two looking after systolic and diastolic readings. Two fuzzy inferences were incorporated for heart rate and night profiling of the subject. High accuracy and interpretability can be
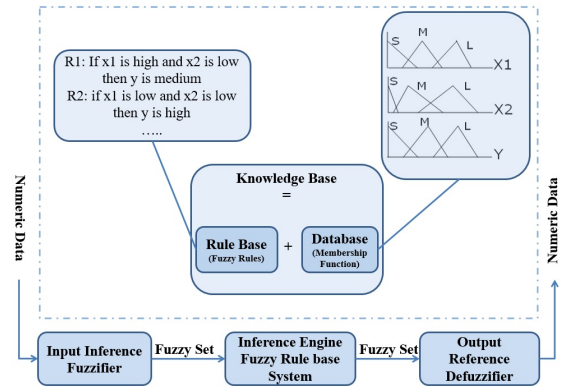

Fig. 5. Fuzzy Inference Engine

achieved if meta-heuristic models are applied along with type 2 fuzzy inferences. The befitting model of fuzzy logic was presented with a name PreFurGe that has the capability to predict the chances of invitro fertilization to help gynecologists and embryologist [14]. The model can be further improved by generating better rules and by ingesting GA. A low cost and accurate framework is presented in [46] that used a matrix-oriented fuzzy rule-based predictive model for heart disease. In [60] a learning membership function that uses neural networks in addition to fuzzy logic systems is proposed. However, this framework lacked desirable accuracy that would have been achieved if the multi-layer approach can be used instead of a single layer approach.

Fuzzy logic is an excellent representation of providing linguistic variables into computing with the liberty of dealing non-discrete/non-linear and imprecise problems. This power of fuzzy logic makes it a perfect choice for the predictive systems that require high accuracy. However, there is a handicap of tuning the membership function and other parameters manually that might cause inconvenience in terms of time and effort. Fuzzy logic cannot scale well for large problem set but still are a lucrative choice for medical domain predictive models.

V. Text Mining

Text mining is an emerging field of decision mining that adopts statistics, machine learning and linguistic techniques to extract high-quality information from unstructured data repositories [18]. Text mining has shown some remarkable results in medical data classification as the data generated by the medical domain is diversified and carries huge volumes. The most important text mining methodologies include information retrieval and natural language processing. The next subsection explains natural language processing in brief.

A. Natural Language Processing (NLP)

The large corpus of medical data has emerged as a major problem area for the domain users. The free context of medical data, medical notes and reports have given a new dimension to the text mining paradigm. The information retrieval represents document under scrutiny as a collection of predefined words however natural language processing takes this a step forward
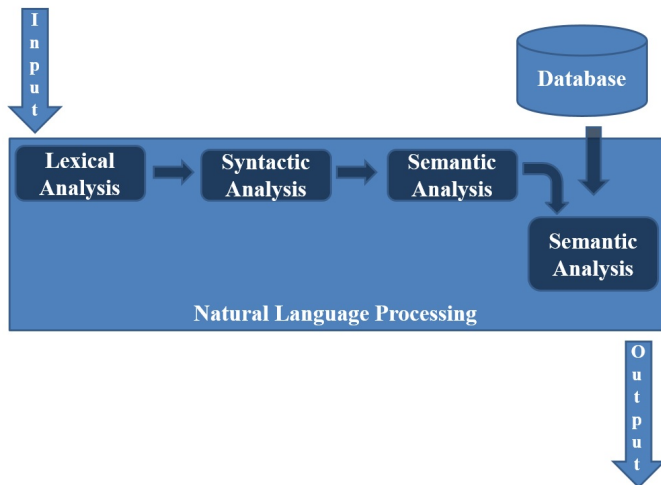
Fig. 6. Natural Language Processing

and generates meaning from natural language and human beings and is depicted in Figure 6 [21].

Latest research presented MetaMap that aimed to reduce the error rate by identifying eligibility for Intravenous Thrombolytic Therapy (IVT) in stroke patients using natural language processing [57]. MetaMap handicapped itself in the generalization of outcomes due to a small sample size and tend to acquire long processing time which makes it a hard choice for real-time large datasets. NLP was used in another research for Healthcare-Associated Infections (HAI) monitoring [61]. The major objectives were sensitivity and specificity. The major areas of medical sciences for this study were digestive, neuro and orthopedic surgery including adult intensive care. The performance factor can further be improved if semantics is applied with expert rules. Similarly, a similar kind of study was conducted on the Mayo Clinic health record for predetermined asthma criteria using NLP [65]. Natural language processing is one of the most widely used methodologies of artificial systems that involves natural language for the representation of knowledge. The main advantage of using NLP is that it is highly expressive and virtually can articulate any real-world situation, emotion, ideas, and pictures. The human intuition by default understands the semantics and vocabulary. NLP perform well when the domain in which it is applied has clarity and is narrowed to the deep understanding. However, NLP tends to have difficulties in elaborating syntax and semantics both when the domain is divergent. If not properly implied, NLP has a tendency of having little uniformity in the sentences that make the grammar ambiguous. NLP is widely used in medical decision-making and performs well where its precise and limited scope in establishing decision making or predictive analysis. Conventional disease classification algorithms perform well when the problem set is simple. With the advancement in the information technology and business intelligence methodologies, their era is converging to more complex and improved variants of disease classifications. A tabular representation of some of the conventional classification algorithms and techniques is presented in Table 1 below with their major characteristics and pros and cons elaborated in detail.

VI. Multi-Objective Optimization Techniques

In CDSS exclusive, potentially useful and original information is extracted from data set to enhance interpretability, accuracy, decrease time and cost [17]. The extracted knowledge is represented in legitimate, valuable and reasonable structures, trend and patterns. Various techniques of classification, clustering, association rule mining, and forecast deliver pronounced help to experts in the earlier detection and diagnosis of the disease [5]. The prompt evolution of data analysis techniques has empowered the production of CDSS to be much more tolerable than ever before. To solve mono and multi-objective problems, evolutionary algorithms have been evolved and their basic design procedure is depicted in Figure 7 [47]. Evolutionary algorithms are designed to allow survival of fittest theory in which the algorithm initiates itself by the selection of the population randomly. Then the method uses a sequence of generations, in which the best design point in a selected population is taken as most fit and is allowed to reproduce. Mathematical modeling is employed by simulating selection, breeding and mutation process. Multi-objective optimization evolutionary algorithms (MOEAs) are efficient with problems of two or three objectives. Evolutionary multi-objective (EMO) algorithms such as Non-Dominant Sorting Genetic Algorithm-II (NSGA-II), MOEA/D, GDE3, SPEA2, and others, have displayed remarkable results in addressing many scientific application problems of real-time nature, engineering and economics problems that pivotally cater two to five objectives. Nevertheless, if a solution to solve a greater number of objectives problem (termed as the multi-objective optimization that usually includes more than three objectives), most of the EMO algorithms fall short to unearth well spread & well converged non-dominated solutions due to the decrease in fitness evaluation function's selection pressure that results in compromised accuracy and interpretability.

Multi-objective optimization was used by employing a variant of classical SVM that is sequential minimal optimization (SMO). The base classifier was used in collaboration with an evolutionary algorithm Elephant Herding Optimization, Decomposition based MOEA and NSGA-II to constitute a framework named CEHO [44]. The system can be modified by invoking ELM and deep neural networks. [17] Presented a multi-objective optimization approach that is based on genetic fuzzy logic. The major objectives used in the research were accuracy and interpretability. However, the model lacked in determining the final non-dominated solutions with a high spread and well-balanced distribution in the objective space. It is evinced from the research resources that NSGA algorithms display better performance when the operators are selected optimally, such as random polynomial mutation to produce the offspring, differential evolution and simulated binary crossover.

Multi-objective optimization techniques have the capability and power of simultaneously dealing with a set of candidate solutions. Unlike to their counterpart, exploration of quite a few members of Pareto optimal set in a single run. The only point of concern is that they require a large number of iterative

TABLE I. COMPARISON OF CONVENTIONAL CLASSIFICATION TECHNIQUES

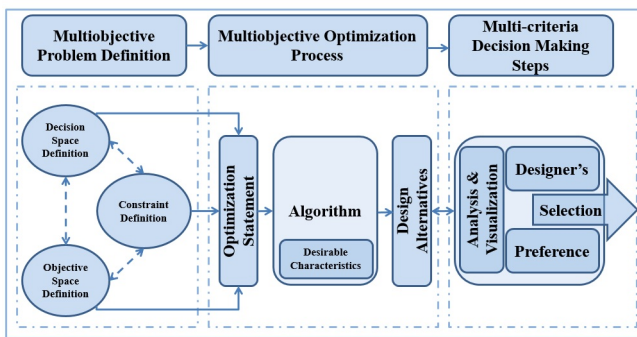| Sr.	Algorithm	Advantage	Disadvantage	Characteristics
		Classification		
1.	Decision Tree [41]	Easily understandable and resilient to replication and data with noise	Not good with inconsistent data & accuracy is also compromised	Recursive algorithm, uses greedy approach.
2.	Artificial Neural Network [50]	Resilient to data replication with ability to handle complex relationships	Susceptible to the data anomalies & hides internal details	Involves more than one layer with a precondition of holding one layer bare minimum.
3.	Rule-based Algorithms [17]	Easily understandable with resilience to inconsistent data	Compromised accuracy	Based on if-then-else rules
4.	Support Vector Machine [69]	Ideal when less training data is available and also suitable for multi-dimensional data	Hides internal details & require parameters	Unorthodox approach and involves mathematical calculations
5.	Naïve Bayes Algorithm [38]	Resilient to all sorts of data abnormalities	Compromise in accuracy & require earlier probability	Involves statistical calculations
6.	K- Nearest Neighbor [53]	Straightforward algorithm with utmost elasticity	Susceptible to replication and noise.	Lethargic approach with prediction results based on local data
		Clustering		
7.	K- Means Algorithms [27]	Straightforward algorithm with utmost processing speed	Not feasible for heterogeneous data and susceptible to data noise	Algorithm uses prototype based results
8.	Hierarchical Method [10]	No parameters required and less prone to initial value	Susceptible to data noise and involves complexity both in terms of space and time	Bottom up approach and uses graph theory
9.	Density-Based Spatial Clustering of Applications with Noise [1]	Noise resilient & ability to handle random density and size.	Involves complexity both in terms of space and time	Approach based on density
10.	Fuzzy C-Means Algorithms [49]	Straightforward algorithm with utmost processing speed	Not feasible for heterogeneous data and susceptible to data noise	Algorithm uses prototype based results
		Association Rule Mining		
11.	Apriori Algorithms [45]	Straightforward algorithm with wide acceptance	Involves complexity both in terms of input/output and time	It is a recursive approach that uses prior knowledge
12.	Dynamic Itemset Counting [55]	Ability to handle input / output complexity	Susceptible to non-heterogeneous data	Dynamic algorithm discovering lost patterns
13.	Direct Hashing and Pruning [42]	Ability to handle candidate patterns count	Cannot handle hash table anomalies	Based on the concept of hashing
14.	E-CLAT Algorithm [26]	Less complex with input /output	Cannot handle large data sets and complex with respect to space	Uses lattice theory and it's a bottom-up approach
15.	D – CLUB Algorithm [26]	Automatically adjusts to the situation and efficiently handle time and space complexity	Compromised accuracy	Best suited for no centralized databases and can support parallel processing

Fig. 7. Multi-objective Optimizing Technique

runs that require high computational effort to generate Pareto front.

In [9] an Evolutionary Algorithms - EAs are used to propose an optimal searching feature subset. This is achieved by introducing a penalty term, to minimize feature count in the selected subset without affecting classification accuracy. In order to achieve the proposed objective, various Evolutionary Computational Algorithms (ECAs) are applied using penalty-based fitness function, which evaluates the next optimal feature subset. ECAs end up with high accuracy for higher-dimensional datasets. The proposed work has been tested using dimensions up to 10,000. However, for feature subset selection, ECAs lack in reducing residual features from final selection

and they are costly too. In [28] development of a robust optimized machine learning - ML system is presented. The aim is to improve risk stratification accuracy by replacing outliers with median configuration, which is based on assumption. Concrete classes were used to accurately classify diabetes in patients. The proposed machine learning system is designed, developed and evaluated using a feature selection strategy. It is then combined with several kinds of classifiers. The proposed approach has shown stable and reliable results. It also improved the performance of existing systems by replacing outliers with median computations. With this approach, the results have become more accurate. However, the system works for only Indian diabetic medical data and its classification. The final solution is costly and time consuming for medical specialists as well as for patients. In [43] the fundamental purpose of the study is to enhance accuracy, sensitivity and specificity rates on Z-Alizadeh Sani dataset. This work proposes a hybrid method, which is highly accurate to diagnose coronary artery disease. This study achieves high performance as far as neural networks are concerned. This performance enhancement is attained by applying Practical Swarm Optimization (PSO). The proposed study helps to reduce the cost considerably along with no major side effects. Thus, coronary artery disease is detected without the need for invasive diagnostic methods, using clinical data. With the help of this approach, multi-objective of accuracy, specificity and accuracy rates on Z-Alizadeh Sani dataset are evaluated. Keeping in mind the dynamics of the problem, designing of a proper network can be a very tough task because of its dependency on problem dynamics. In

[12] Bio-inspired Multi-objective algorithm is presented. The process of gene selection is carried out using microarray data classification. Refined formulations of BA have been used along with MO search techniques and specialized operators. Variable selection is also done using binary domain called MOBBA-LS. The proposed algorithm called BA produces best subsets with lesser number of genes with the highest accuracy. These genes have excellent relevancy. Proposed work showed low performance, which needs to be enhanced. The other drawback is the increased time-complexity. Another study is based on the notion to build a simple classifier using multiclass classification strategy, which has integrated various multi-objective namely feature selection as well as its construction. It further models the intelligibility objective into a distance-based classifier [23]. The proposed approach optimizes data models by using genetic programming. This model named (M4GP) is based on an innovative stack-based program method, which makes the multi-dimensional solutions simpler to construct. This methodology gives M4GP an edge when a comparison is made with M2GP and M3GP (both of these models applied tree-based structure). The results of this model show that the final solution is interpretable and more accurate. M4GP also offers an efficient and flexible solution for providing accurate classifiers. Moreover, it also yields the best classification for small dimensional operations. Since M4GP works on the population-related domain, therefore it incurs a higher computational price. In [2] the main theme of this paper is to deal with such MO problems having high uncertainty. This uncertainty is represented by triangular fuzzy numbers. It involves solving the problem using fuzziness propagation to fitness functions. The proposed approach consists of a fuzzy Pareto dominant solution and then to apply EAs to reach to a solution. One of the advantages is that it uses transformations of other shapes using operations like projections, linguistic classifiers, and compositions. TFN can be deduced using the above operations. Regardless of a lot of suggested methodologies, there are still many open issues to be addressed for this domain. For instance, there is no real /close remedy to deal with uncertainty, which prevails as the core aspect in the area of multi-objective problem domains. In [4] the main purpose of this research is to present a model that is able to extract health indicators HI's. The aim is to keep track of various signals during operation. Thus, to study the component degradation during this process /operation. The proposed idea is derived from the usage of the feature extraction method. The selection of Health Indicators has three steps; firstly, feature extraction is done. Secondly, the selection is done and at the third stage, fusion is done by applying the BDE algorithm (multi-objective Binary Differential Evolution). This method has produced much satisfactory HI's. It has shown more satisfactory health indicators than found in other research studies. A set back of the proposed study is that unsatisfactory prognostic performances can result due to RUL. In [59] the main aim of this research is to study various entropy-based design optimization schemes and attempt to decrease the gap between them. This proposed research study is based on the notion of join entropy schemes that are independent. It is further applied to a real-life problem related to the water distribution network. Various stages include maximizing the joint entropy along with applying a penalty-free genetic algorithm with three objectives. The benefits of using such a methodology is that it presents a relatively simpler and easier way to be assimilated

using multi-objective optimization algorithms. Another aspect is its efficiency in generating results. In short, this study shows a balance between computational budgeting and flexibility. One of the major drawbacks is that a significant increase in the available feasible solutions is gained when compared to previous research studies. This increase in entropy values made infeasible solutions to be vague and distorted. The main theme of another research is to present an innovative model that is able to enlighten occurrence of asthma, and also detect two markers of allergy namely IgE antibody against common allergens, and skin prick test positivity for common allergens (SPT) [64]. The technique was based on MGGP (multi-objective grammar-based genetic programming). The medical dataset contains details of nutritional, psychosocial, socioeconomics, atmospheric and infectious factors gathered from children who were part of the study/process. MGGP model achieved higher accuracy and results were also easy to interpret. The performance of MGGP model for each iteration takes 28.1 h along with its limitation/absence to offer parallelism capability for proposed work for now. In [51] presents the optimum aspects of SPIF parameters for titanium denture plate. The present paper attempts to measure the likelihood of generating customized Commercially Pure Titanium Grade 1 (CP-Ti Gr.1) denture plate with reasonable accuracy. The proposed working aims to control some process parameters that affect the quality of the final product with the prime objective of geometrical accuracy. The proposed strategy of optimization of multi-objective is based on numerical simulation using a Multi-Objective Genetic Algorithm and the Global Optimum Determination. This is done by Linking and Interchanging Kindred Evaluators algorithm to find the optimum solutions. Minimizing sheet thickness, its ultimate along with increasing forming force were main objectives. Achieving robustness for the selection of optimum factors in SPIF is the core advantage. It also results in improving geometric errors, especially in the base area. However, a large number of errors in the part wall section needs to be addressed in further development specifically to validate the quality of the surface after forming sections. Another research applied age prediction using neuroimaging with the help of ML schemes. The core theme of this study is to improve accuracy for age prediction by Bayesian parameter optimization pertaining to age prediction [24]. Bayesian optimization is done in an iterative manner to check the sample space for various parameters to achieve accuracy in the resultant space. This approach improves the notion to distinguish young and old brain. Neuroimaging data is the basis for the whole idea. The Bayesian optimizations achieve optimum voxel size thus improves performance. Keeping in view the complexity in neuroscience because of its multi-disciplinary nature, the research analysts may not hold expertise in every area. Thus, further unbiased optimization parameters may give more benefit.

Due to their precise predictive nature, these algorithms are utilized in the medical domain to evolve CDSS by amalgamating them with fuzzy logic.

VII. CHALLENGES AND PRACTICES

As the rule of thumb, all domains of sciences have their own challenges and rigorous efforts are always underway by the researchers of respective domains to rectify those challenges. Like other domains, disease classification and CDSS

has a number of challenges, in particular, some of them are enlisted below:

- ***Critical heterogeneous data [25]:*** The data available in the medical domain is possibly the most difficult in nature including numeric, alphanumeric, pictorial, and continuous, etc forms. To handle such data there is a requirement of a comprehensive data cleansing and normalization mechanism to eradicate the heterogeneity, missing values, and outliers.

- ***Clinical Data Privacy [54]:*** Privacy and secrecy of the clinical data are one of the major concern in health care. Most of the time this data is reused by various stakeholders for the betterment of health care but the concern of this data being misused can not be ruled out. The general attitude of the stakeholders toward its reuse must be analyzed carefully before the sharing of this critical information to mitigate this threat.

- ***Compromise on Identification of risk factors [16]:*** Most of the clinicians and technical hands involved in the development of CDSS neglect the identification of risk factors in the earlier design. The risk factors are generic and some of them are specific to the domain (a type of disease, type of CDSS, etc). The earlier detection, identification, and eradication of all sorts of risk factor makes the CDSS more robust and resilient to any change and advancement in technology.

- ***Incremental sensitivity and specificity [5]:*** Sensitivity is the true positive rate and specificity is the true negative rate. Both attributes have an important role in the construction of CDSS. It is desired that the CDSS should increase the sensitivity and specificity incrementally by training and fitting its model.

- ***Effectiveness [5]:*** The CDSS can only be effective if it is easy to use, generate accurate predictions, the decisions are interpretable with high sensitivity and specificity and has a low computational cost. The CDSS should be designed by keeping all these objectives to an acceptable level.

- ***Scalability and Adaptability [5]:*** CDSS should be developed with the ideology of being scalable and adaptable; as the medical domain is an evolving field of science a tunnel approach in the design of CDSS leads the clinicians and policymakers toward an ineffective CDSS design. It is always recommended that the CDSS should be constructed by keeping in view the scalability and adaptability perspective in mind.

- ***Accuracy & Interpretability [17]:*** Accuracy is the ability to provide a clinical decision process that generates an outcome with high precision value. However, Interpretability is the feature that provides the user with condensed and comprehensible enlightenment and explanation with reason, of the proposed decision. A major concern of the medical domain is to evolve CDSS that can improve the accuracy and interpretability simultaneously and considerably reduce the computational cost.

- ***Different Knowledge Areas [37]:*** CDSS's are systems that facilitate the clinicians but are evolved from the amalgamations of various knowledge areas that include computer sciences, statistics, mathematics, bioinformatics, medical sciences, etc. To work on CDSS require multifaceted knowledge of different domains of sciences. That is the reason that clinicians and policymakers of the medical domain working on the development of CDSS find it a difficult choice.

- ***Extraction of the relationship between variables [13]:*** It is considered essential in the construction of CDSS that all the variables along with their relationship should contribute. The relationship of the variables makes the system much more comprehensible and gives the utmost interpretability of the clinical decisions.

- ***The requirement of multiple objective decisions [13]:*** Clinical decisions are critical in nature and require multiple objectives to accomplish accurate predictive models. Most of the CDSS have their focus on achieving accuracy, however, over the period of time, multiple objectives like interpretability, specificity, sensitivity, and computational complexity have evolved as a major deriving force in accessing a CDSS. So the inclusions of multi-objective decision support systems have wider acceptability and are recent research areas of clinical decision making and disease prediction.

- ***Transfer Learning [52]:*** CDSS are primarily based on various decision making computer methodologies that require time to learn different pattern over time, it is, therefore, desired that the CDSS should be capable of transferring this inferred knowledge to its inherited system.

VIII. SWOT ANALYSIS

SWOT analysis refers to the Strengths, Weaknesses, Opportunities, and Threats allied to a solution. Our survey opines that invoking a CDSS has certain pros and cons allied to it that can be sorted out for the better results in healthcare. In the upcoming section we will explain all the above mentioned important analysis factors in more detail and the same is depicted in Figure 8 below for consideration while opting CDSS invocation:

A. Strengths

CDSS has shown remarkable improvement in health care by diagnosing complex imaging results which remained unidentifiable by the human eye. The same is true for numeric data values. The major strength lies in their ever-evolving/training nature that makes them less prone to errors. As automated systems are free from sentiments/behavioral aspect that makes their performance to remain at an optimum threshold level. The results generated are not only accurate but they are comprehensible by the stakeholders. Automation brings scalability and transfers learning opportunities as part of a package that helps in an ongoing evolution of even better systems.
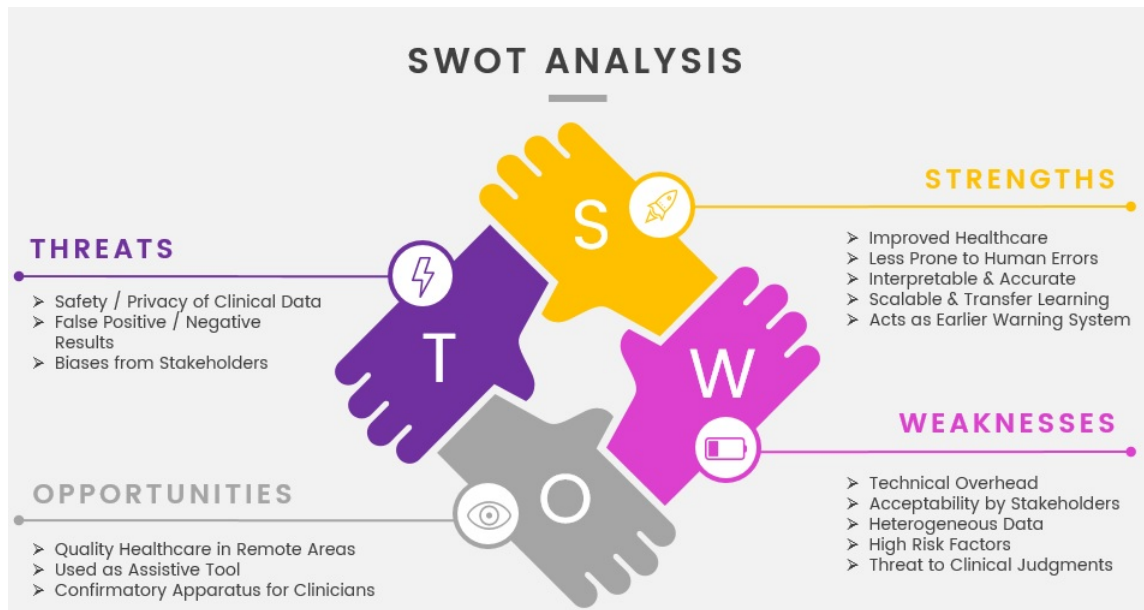
Fig. 8. SWOT Analysis

B. Weaknesses

Most of the studies discussing the commissioning of CDSS highlight the issue of acceptability from the stakeholders as they consider it to be an overhead instead of helping hand in assisting their assertions/decisions. Clinicians consider the decision support systems to be an evaluation apparatus for their decisions and prognosis. Clinicians have to perform diagnostic tasks that are time-consuming. Another very important factor that makes these systems complicated is the heterogeneous data. There is every likelihood that the outliers, missing values, and typo errors may cause daunting results that can lead to confusion and further investigation that is an overhead in terms of time and cost.

C. Opportunities

Revolution in technology brings opportunities along-with and CDSS has given great ease of use to hospital administration, health care policymakers and clinicians. The major opportunity comes in the remote areas where specialized healthcare resources in terms of man and material both are not available. These systems act as an assistive tool in justifying and conforming to the decisions made by the doctors. Hospital administration and policymakers can also monitor the true positive/negative and false positive/negative diagnosis very easily and may take measures to improve healthcare.

D. Threats

One of the major considerations that need to be catered to is the safety, privacy, and secrecy of clinical data. Clinical data contains health information of masses that can not be revealed to someone irrelevant. As the adaptability of these systems and data allied to it is growing exponentially, data safety has become a vital concern. Similarly, the behavioral aspect should also be taken as a threat as most of the mindsets don't allow change to the conventional methodologies in practice. CDSS being a relatively new methodology in the health domain may face these biases from the stakeholders.

IX. CONCLUSION

CDSS is expected to improve medical healthcare quality by assisting the doctors in making clinical decisions. The health-care data classification mechanism assists medical experts in the early identification and management of medical malfunctioning and symptoms arisen in the patient. This substantial contribution has a pivotal role in the quality enhancement of healthcare by assisting doctors in decision-making. The contribution of decision support in general fields has shown great acceptance but specific to the medical domain and disease classification their acceptance is still scarce. A large amount of investigation and research in evolving an effective CDSS for the medical domain and disease classification are studied, the mechanisms used to service this research area include classification, clustering, ensemble, artificial neural networks, evolutionary algorithms like genetic algorithms (GA) and deep learning. The major challenge of CDSS is to attain the utmost accuracy, which has the ability to provide a clinical decision process that generates an outcome with high precision value. The aim of this review is to appraise the clinician and policy-makers of the medical domain to gain the technical knowledge of renowned computer-based methodologies that are employed to construct the decision support models of medical domains. By doing so, the clinician and policymakers can give sensible and informed input during the analysis, design, development and deployment stages of CDSS. In the implementation stage, the clinicians could provide guidance on which of the method-ology described above yields better results based on the clinical problem, the type of CDSS required and the dataset. With the deep insight on the methodologies surveyed in this research, the clinicians and policymakers will have the confidence to ad-vocate the importance and significance of the system that will result in improved medical care and quality in the validation

phase. Another depiction of relatively very recent and relevant multi-objective methodologies in evolving CDSS is highlighted in this research. It is further opined that multi-objective optimization techniques have shown remarkable results especially in the field of medical decision-making and is gaining a fast reputation for their accurate and interpretable results.

REFERENCES

[1] Amineh Amini, Teh Ying Wah, and Hadi Saboohi. On density-based data streams clustering algorithms: A survey. *Journal of Computer Science and Technology*, 29(1):116–141, 2014.

[2] Oumayma Bahri, El-Ghazali Talbi, and Nahla Ben Amor. A generic fuzzy approach for multi-objective optimization under uncertainty. *Swarm and Evolutionary Computation*, 40:166–183, 2018.

[3] K Balasaravanan and M Prakash. Detection of dengue disease using artificial neural network based classification technique. *International Journal of Engineering & Technology*, 7(1.3):13–15, 2018.

[4] P Baraldi, G Bonfanti, and E Zio. Differential evolution-based multi-objective optimization for the definition of a health indicator for fault diagnostics and prognostics. *Mechanical Systems and Signal Processing*, 102:382–400, 2018.

[5] Saba Bashir, Usman Qamar, and Farhan Hassan Khan. Heterogeneous classifiers fusion for dynamic breast cancer diagnosis using weighted vote based ensemble. *Quality & Quantity*, 49(5):2061–2076, 2015.

[6] Riccardo Bellazzi and Blaz Zupan. Predictive data mining in clinical medicine: current issues and guidelines. *International journal of medical informatics*, 77(2):81–97, 2008.

[7] Thomas Bernd, Markus Kleutges, and Andreas Kroll. Nonlinear black box modelling–fuzzy networks versus neural networks. *Neural computing & applications*, 8(2):151–162, 1999.

[8] RJ Brachman, HJ Levesque, and M Pagnucco. Knowledge representation and reasoning: Knowledge representation and reasoning, 2004.

[9] Basabi Chakraborty and Atsushi Kawamura. A new penalty-based wrapper fitness function for feature subset selection with evolutionary algorithms. *Journal of Information and Telecommunication*, 2(2):163–180, 2018.

[10] Kerina Blessmore Chimwayi, Noorie Haris, Ronnie D Caytiles, and N Ch SN Iyengar. Risk level prediction of chronic kidney disease using neuro-fuzzy and hierarchical clustering algorithm (s). 2017.

[11] Krzysztof J Cios and G William Moore. Uniqueness of medical data mining. *Artificial intelligence in medicine*, 26(1-2):1–24, 2002.

[12] M Dashtban, Mohammadali Balafar, and Prashanth Suravajhala. Gene selection for tumor classification using a novel bio-inspired multi-objective approach. *Genomics*, 110(1):10–17, 2018.

[13] Kalyanmoy Deb. Multi-objective optimisation using evolutionary algorithms: an introduction. In *Multi-objective evolutionary optimisation for product design and manufacturing*, pages 3–34. Springer, 2011.

[14] José Gonzalez Enríquez, V Cid, N Muntaner, J Aroba, José Navarro, FJ Domínguez-Mayo, María José Escalona, and I Ramos. Behavior patterns in hormonal treatments using fuzzy logic models. *Soft Computing*, 22(1):79–90, 2018.

[15] Nura Esfandiari, Mohammad Reza Babavalian, Amir-Masoud Eftekhari Moghadam, and Vahid Kashani Tabar. Knowledge discovery in medicine: Current issue and future trend. *Expert Systems with Applications*, 41(9):4434–4463, 2014.

[16] Paolo Fraccaro, Panagiotis Plastiras, Chiara Dentone, Antonio Di Biagio, Peter Weller, et al. Behind the screens: Clinical decision support methodologies–a review. *Health Policy and Technology*, 4(1):29–38, 2015.

[17] Marian B Gorzałczany and Filip Rudziński. Interpretable and accurate medical data classification–a multi-objective genetic-fuzzy optimization approach. *Expert Systems with Applications*, 71:26–39, 2017.

[18] S Inzalkar and Jai Sharma. A survey on text mining-techniques and application. *International Journal of Research In Science & Engineering*, 24:1–14, 2015.

[19] Daxin Jiang, Chun Tang, and Aidong Zhang. Cluster analysis for gene expression data: a survey. *IEEE Transactions on knowledge and data engineering*, 16(11):1370–1386, 2004.

[20] Michael I Jordan and Tom M Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260, 2015.

[21] Wahab Khan, Ali Daud, Jamal A Nasir, and Tehmina Amjad. A survey on the state-of-the-art machine learning models in the context of nlp. *Kuwait journal of Science*, 43(4), 2016.

[22] Miroslav Kubat. Artificial neural networks. In *An Introduction to Machine Learning*, pages 91–111. Springer, 2015.

[23] William La Cava, Sara Silva, Kourosh Danai, Lee Spector, Leonardo Vanneschi, and Jason H Moore. Multidimensional genetic programming for multiclass classification. *Swarm and evolutionary computation*, 44:260–272, 2019.

[24] Jenessa Lancaster, Romy Lorenz, Rob Leech, and James H Cole. Bayesian optimization for neuroimaging pre-processing in brain age classification and prediction. *Frontiers in aging neuroscience*, 10:28, 2018.

[25] Nada Lavrač. Selected techniques for data mining in medicine. *Artificial intelligence in medicine*, 16(1):3–23, 1999.

[26] Jing-Song Li, Yi-Fan Zhang, and Yu Tian. Medical big data analysis in hospital information system. *Big Data on Real-World Applications*, page 65, 2016.

[27] A Malav, K Kadam, and P Kamat. Prediction of heart disease using k-means and artificial neural network as hybrid approach to improve accuracy. *International Journal of Engineering and Technology*, 9(4):3081–3082, 2017.

[28] Md Maniruzzaman, Md Jahanur Rahman, Md Al-MehediHasan, Harman S Suri, Md Menhazul Abedin, Ayman El-Baz, and Jasjit S Suri. Accurate diabetes risk stratification using machine learning: role of missing value and outliers. *Journal of medical systems*, 42(5):92, 2018.

[29] Gunjan Mansingh, Kweku-Muata Osei-Bryson, and Han Reichgelt. Using ontologies to facilitate post-processing of association rules by domain experts. *Information Sciences*, 181(3):419–434, 2011.

[30] Patricia Melin, Ivette Miramontes, and German Prado-Arechiga. A hybrid model based on modular neural networks and fuzzy systems for classification of blood pressure and hypertension risk diagnosis. *Expert Systems with Applications*, 107:146–164, 2018.

[31] Rita Morisi, David Neil Manners, Giorgio Gnecco, Nico Lanconelli, Claudia Testa, Stefania Evangelisti, Lia Talozzi, Laura Ludovica Gramegna, Claudio Bianchini, Giovanna Calandra-Buonaura, et al. Multi-class parkinsonian disorders classification with quantitative mr markers and graph-based features using support vector machines. *Parkinsonism & related disorders*, 47:64–70, 2018.

[32] Janmenjoy Nayak, Bighnaraj Naik, and H Behera. A comprehensive survey on support vector machine in data mining tasks: applications & challenges. *International Journal of Database Theory and Application*, 8(1):169–186, 2015.

[33] Eric WT Ngai, Yong Hu, YH Wong, Yijun Chen, and Xin Sun. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50(3):559–569, 2011.

[34] Eric WT Ngai, Li Xiu, and Dorothy CK Chau. Application of data mining techniques in customer relationship management: A literature review and classification. *Expert systems with applications*, 36(2):2592–2602, 2009.

[35] Thanh Nguyen, Abbas Khosravi, Douglas Creighton, and Saeid Nahavandi. Classification of healthcare data using genetic fuzzy logic system and wavelets. *Expert Systems with Applications*, 42(4):2184–2197, 2015.

[36] Federico Paoli, Ingrid Schmidt, Olivia Wigzell, and Andrzej Ryś. An eu approach to health system performance assessment: Building trust and learning from each other. *Health Policy*, 123(4):403–407, 2019.

[37] Vimla L Patel, Edward H Shortliffe, Mario Stefanelli, Peter Szolovits, Michael R Berthold, Riccardo Bellazzi, and Ameen Abu-Hanna. The coming of age of artificial intelligence in medicine. *Artificial intelligence in medicine*, 46(1):5–17, 2009.

[38] Shadab Adam Pattekari and Asma Parveen. Prediction system for heart disease using naïve bayes. *International Journal of Advanced Computer and Mathematical Sciences*, 3(3):290–294, 2012.

[39] Tadeusz Pietraszek and Axel Tanner. Data mining and machine learning-towards reducing false positives in intrusion detection. *Information security technical report*, 10(3):169–183, 2005.

[40] R Prashanth, Sumantra Dutta Roy, Pravat K Mandal, and Shantanu Ghosh. Automatic classification and prediction models for early parkinson's disease diagnosis from spect imaging. *Expert Systems with Applications*, 41(7):3333–3342, 2014.

[41] Yinsheng Qu, Bao-Ling Adam, Yutaka Yasui, Michael D Ward, Lisa H Cazares, Paul F Schellhammer, Ziding Feng, O John Semmes, and George L Wright. Boosted decision tree analysis of surface-enhanced laser desorption/ionization mass spectral serum profiles discriminates prostate cancer from noncancer patients. *Clinical chemistry*, 48(10):1835–1843, 2002.

[42] E Ramaraj et al. An efficient pattern mining analysis in health care database. 2009.

[43] P Ranjitha and Vanishri Arun. Decision making for heart disease detection using hybrid neural network-particle swarm optimization algorithm. 2018.

[44] Nalluri MadhuSudana Rao, Krithivasan Kannan, Xiao-zhi Gao, and Diptendu Sinha Roy. Novel classifiers for intelligent disease diagnosis with multi-objective parameter evolution. *Computers & Electrical Engineering*, 67:483–496, 2018.

[45] P Sambasiva Rao and T Uma Devi. Applicability of apriori based association rules on medical data. *International Journal of Applied Engineering Research*, 12(20):9451–9458, 2017.

[46] P Sambasiva Rao and T Uma Devi. Improving accuracy of fuzzy rule based mining for heart disease detection using cost matrix. *International Journal of Advanced Research in Computer Science*, 9(1), 2018.

[47] Gilberto Reynoso-Meza, Javier Sanchis, Xavier Blasco, and Sergio García-Nieto. Physical programming for preference driven evolutionary multi-objective optimization. *Applied Soft Computing*, 24:341–362, 2014.

[48] Cristóbal Romero and Sebastián Ventura. Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(6):601–618, 2010.

[49] Indira Rustempasic and Mehmet Can. Diagnosis of parkinson's disease using fuzzy c-means clustering and pattern recognition. *Southeast Europe Journal of Soft Computing*, 2(1), 2013.

[50] Oluwarotimi Williams Samuel, Grace Mojisola Asogbon, Arun Kumar Sangaiah, Peng Fang, and Guanglin Li. An integrated decision support system based on ann and fuzzy_ahp for heart failure risk prediction. *Expert Systems with Applications*, 68:163–172, 2017.

[51] M Sbayti, R Bahloul, H BelHadjSalah, and F Zemzemi. Optimization techniques applied to single point incremental forming process for biomedical application. *The International Journal of Advanced Manufacturing Technology*, 95(5-8):1789–1804, 2018.

[52] Manjeevan Seera and Chee Peng Lim. A hybrid intelligent system for medical data classification. *Expert Systems with Applications*, 41(5):2239–2249, 2014.

[53] Parul Sinha and Poonam Sinha. Comparative study of chronic kidney disease prediction using knn and svm. *International Journal of Engineering Research and Technology*, 4(12):608–12, 2015.

[54] Lea L Skovgaard, Sarah Wadmann, and Klaus Hoeyer. A review of attitudes towards the reuse of health data among people in the european union: The primacy of purpose and the common good. *Health Policy*, 2019.

[55] T Smitha and V Sundaram. Association models for prediction with apriori concept. *International Journal of Advances in Engineering & Technology*, 5(1):354, 2012.

[56] Hari Agus Sujono, Muhammad Rivai, and Muhammad Amin. Asthma identification using gas sensors and support vector machine. *TELKOM-NIKA (Telecommunication Computing Electronics and Control)*, 16(4), 2018.

[57] Sheng-Feng Sung, Kuanchin Chen, Darren Philbert Wu, Ling-Chien Hung, Yu-Hsiang Su, and Ya-Han Hu. Applying natural language processing techniques to develop a task-specific emr interface for timely stroke thrombolysis: a feasibility study. *International journal of medical informatics*, 112:149–157, 2018.

[58] Tuan Zea Tan, Chai Quek, and Geok See Ng. Ovarian cancer diagnosis by hippocampus and neocortex-inspired learning memory structures. *Neural Networks*, 18(5-6):818–825, 2005.

[59] Tiku T Tanyimboh and Anna M Czajkowska. Joint entropy based multi-objective evolutionary optimization of water distribution networks. *Water resources management*, 32(8):2569–2584, 2018.

[60] Feyzullah Temurtas. A comparative study on thyroid disease diagnosis using neural networks. *Expert Systems with Applications*, 36(1):944–949, 2009.

[61] Nastassia Tvardik, Ivan Kergourlay, André Bittar, Frédérique Segond, Stefan Darmoni, and Marie-Hélène Metzger. Accuracy of using natural language processing methods for identifying healthcare-associated infections. *International journal of medical informatics*, 117:96–102, 2018.

[62] Elif Derya Übeyli. Implementing automated diagnostic systems for breast cancer detection. *Expert systems with Applications*, 33(4):1054–1062, 2007.

[63] Fevrier Valdez, Patricia Melin, and Oscar Castillo. A survey on nature-inspired optimization algorithms with fuzzy logic for dynamic parameter adaptation. *Expert systems with applications*, 41(14):6459–6466, 2014.

[64] Rafael V Veiga, Helio JC Barbosa, Heder S Bernardino, João M Freitas, Caroline A Feitosa, Sheila MA Matos, Neuza M Alcântara-Neves, and Maurício L Barreto. Multiobjective grammar-based genetic programming applied to the study of asthma and allergy epidemiology. *BMC bioinformatics*, 19(1):245, 2018.

[65] Chung-Il Wi, Sunghwan Sohn, Mir Ali, Elizabeth Krusemark, Euijung Ryu, Hongfang Liu, and Young J Juhn. Natural language processing for asthma ascertainment in different practice settings. *The Journal of Allergy and Clinical Immunology: In Practice*, 6(1):126–131, 2018.

[66] Amani Yahyaoui and Nejat Yumusak. Decision support system based on the support vector machines and the adaptive support vector machines algorithm for solving chest disease diagnosis problems. *Biomedical Research*, 29(7), 2018.

[67] Duen-Yian Yeh, Ching-Hsue Cheng, and Yen-Wen Chen. A predictive model for cerebrovascular disease using data mining. *Expert Systems with Applications*, 38(7):8970–8977, 2011.

[68] Dimitrios Zafeiris, Sergio Rutella, and Graham Roy Ball. An artificial neural network integrated pipeline for biomarker discovery using alzheimer's disease as a case study. *Computational and Structural Biotechnology Journal*, 16:77–87, 2018.

[69] Rahmat Zolfaghari. Diagnosis of diabetes in female population of pima indian heritage with ensemble of bp neural network and svm. *Int. J. Comput. Eng. Manag*, 15:2230–7893, 2012.