Innovative Expert System for Coronary Heart Disease Diagnosis Utilizing Soft Sets

Jackson S[1]
PG and Research Department of Mathematics, V.O.Chidambaram College , Thoothukudi, Tamilnadu, India.
jacks.mat@voccollege.ac.in

Sasikala D[2]
Department of Mathematics
PSGR krishnammal College for Women
Coimbatore, Tamilnadu, India.
dsasikala@psgrkcw.ac.in

Deepa M[3]
Department of Mathematics
PSGR krishnammal College for Women
Coimbatore, Tamilnadu, India.
mdeepa@psgrkcw.ac.in

2024 Control Instrumentation System Conference (CISCON) | 979-8-3503-7548-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/CISCON62171.2024.10696680

***Abstract*— In the recent past, there has been a development of a vast array of theories seeking to explain deal with uncertainties. As a branch of artificial intelligence, medical systems supported with mathematical equations have shown great results in disease diagnosis. Soft set theory is a relatively novel field developed by Molodstov that shows a great deal of potential. In this paper, the "Soft Expert System Framework Incorporating Soft Set Theory and Fuzzy Set Theory" is proposed with the aim of providing a new mathematical tool for handling vagueness in diagnosis. We undertook the development of the Soft Expert System (SES) using medical dataset that contains patient records in order to predict coronary artery disease. In this study, the fuzzy set and soft set were utilized to assess the likelihood of the disease based on the variables including age, blood sugar, high density lipoprotein (HDL), low density lipoprotein (LDL), glycated hemoglobin (HbA1C), uric acid, and blood pressure. The methodology includes key procedures: The generic structure includes Data input, Fuzzyfication, Conversion of Fuzzy sets to soft sets, Reduction of parameters in soft sets, Formulation of soft rules and Data output. This system assumes the role of assisting medical practitioners to improve on their diagnostic abilities and will also ensure that the process takes less time as it would in the normal course.**

Keywords*— *Soft Set, Fuzzy Logic, Coronary, Disease, Diagnosis, Health Risk.

1. INTRODUCTION

Mathematics is an essential tool that can be useful in solving challenges as well as tackling issues in the medical faculty. Another area of mathematics is frequently used in conjunction with probability and statistics to estimate the efficiency of the newly developed drugs and the treatment methods to assess the probability of survival and patients' mortality rates with different diseases treated by a particular method[1]. This is especially the case given the fact that as the field of medicine develops, especially for prediction and assessment, mathematical techniques, or the use of mathematics, are slowly gaining their importance. The last decade has witnessed the appearance of such notions as fuzzy sets, soft sets, probabilistic, ambiguous, and rough set solutions to solve problems effectively[2]. These theories become useful in dealing with continuum of various types of uncertainties and ambiguities that may occur. Originally, Zadeh in 1965 proposed the theory of fuzzy sets as a generalization of conventional or classical set theory. Fuzzy sets are mainly applied to models for recognition and processing of non-probabilistic data or information, thus being useful when controlling non- linear or undefined systems[3]. It can be concluded, therefore, that as fields, disciplines and networks expand, the theory of fuzzy sets emerges as an essential tool that provides new potential for the development of various mathematical applications. Specifically; fuzzy set theory is characterized as a method of working with information that is incomplete or indeterminate[4]. Fuzzification is always performed upon input parameters as they are converted to the multi-valued logic. This approach is based on the paradigm of an element having a degree of membership in one or more sets, where the possible values of the membership function range from 0 to 1. William B. Kannel and Daniel L McGee in 1979 presented a cross-sectional association between Diabetes and Cardiovascular disease risk factors in men and women[5]. Acceptable risks are well anticipated by the fuzzy expert system [6] designed for heart disease diagnosis, which always provides accurate results to doctors to be considered as an ideal diagnostic tool. The first objective, which is most crucial, is to make the right diagnosis as early as possible. A generic fuzzy expert system acts as an expert and often utilizes fuzzy instead of Boolean reasoning[7]. It consists of the membership functions and the rule based system that is used for data reasoning and numerical computations. The approximate reasoning fuzzy expert system is defined and employed to solve problems and specify values. These systems have been over the years used to mimic the analytical ability of a doctor in the identification of diseases like lung cancer [8], kidney diseases [9], and cardiac diseases [10]. This paper presents the following: Schematic Representation of the Soft Expert System is shown in Fig. 1 below:

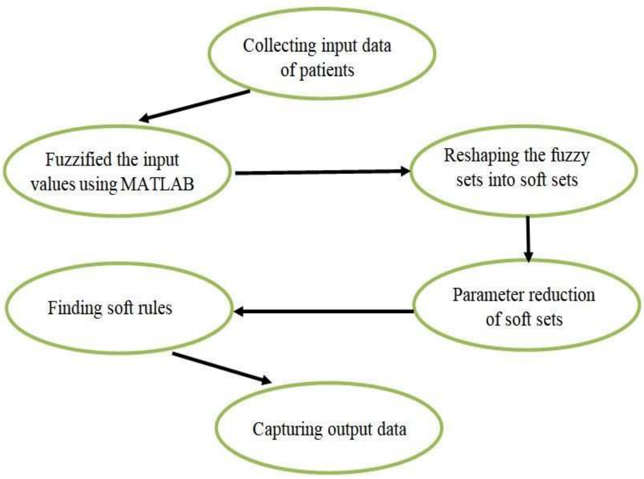

Fig.1: Schematic Representation of the Soft Expert System

From the early view of the fuzzy set theory, Molodtsov has introduced the soft set theory in 1999 to deal with the uncertainty [11]. It has to emphasize that the soft set theory was proposed to define items without strictly predetermined boundaries. When regarded as a parameterized family of sets, the problem of parameterization is solved by points which establish the frame around the set in question and are the parameters of the soft set. Soft set theory has emerged as a useful concept in the later years and has vast areas of applications[12]. The progress that has been made in the last few years in the field of soft expert systems is significant. It is worthy to note that Chen discussed issues associated with soft expert systems and the need for parameter reduction algorithms.

In 2011, Ma proposed a new normal parameter reduction algorithm which made the computation much easier than the simplified process in the parameter reduction of soft sets. Fuzzy soft expert system for predicting coronary artery disease was developed by Hassan, Osama, Ahmed, & Ghany in 2016. In 2019, Iftikhar [6] developed an expert system by applying the soft set theory for the identification of the risk factors regarding the patients affected by Dengue Fever, which is an acute viral disease transmitted through the mosquito, Aedes aegypti.

Coronary Artery Disease affects millions of people around the globe each year, ranking it as one of the major killers globally with the majority of these cases reported in the developed world. This illness arises when fat and cholesterol from the bloodstream deposit themselves in the principal artery of the heart in a form of a plaque that eventually causes heart attacks and other deaths. The objectives of this research are to develop a detection system for coronary artery disease using fuzzy set approach and soft set approach, in order to increase the degree of accuracy in healthcare systems and therefore, increase the performance level of medical professionals and ultimately decrease the mortality rate of patients suffering from the aforementioned disease.

II. LITERATURE SURVEY

Fuzzy set theory according to the scholar Lotfi A. Zadeh in 1965 shifted the notion of the field of mathematics by offering a structure in which vagueness and uncertainty could be managed in decision-making. Fuzzy set theory had been substantially developed and extended over the years to accommodate for various real world problems[13]. Soft set theory which is considered as another significant extension of the original concept was presented by Molodtsov in 1999 Also, it was created for the purpose of managing uncertainty but it offers more flexibility in the way elements are defined since they do not have to meet any rigorous conditions.

The attempt of soft set theory has attracted significant attention in the recent past because it has vast areas of application. Soft set theory has been employed to deal with several issues in different fields such as decision making, pattern recognition, data analysis and generation of expertise systems [14]. The utility of soft set theory as a mathematical tool for dealing with imprecise and uncertain data for healthcare domain is well attributed to the fact that, it is vital for diagnosing illness as well as predicting the probable course of the disease.

Hence, the theories of fuzzy and soft set have been extended in healthcare's domain, enabling the creation of fresh subject matter expert systems for disease prediction and diagnosis[15]. Chen reviewed the issue of soft expert systems on accuracy with significant results, stating that parameter reduction techniques need improvement to improve the reliability of the soft expert systems[16]. Based on this study, another research work was presented by Ma in 2011 with a new normal parameter reduction algorithm that focused on the soft set parameter reduction with an intention to reduce the computational processes.

Soft expert systems have been discussed in several publications aimed at the analysis of diseases and their prediction and diagnosis. The use of fuzzy soft expert system was introduced by Hassan, Osama, Ahmed, and Ghany in 2016 for the prediction of CAD, which signifies that the proposed system has a good potential for use in the health care sector. Likewise, Iftikhar proposed an expert system applied soft sets for identifying the risk factors in patients with Dengue Fever which is one of the notable enhancements in the global approach towards infectious diseases.

CAD [Coronary Artery Disease] is still on the rise and affects millions of people worldwide each year. High mortality rates have been attributed to CAD, especially within developed countries, which underlines the necessity for the development of efficient diagnostic and prognostic employing biomarkers[17]. There are significant applications in using the concepts of both fuzzy and soft set theories in CAD diagnosis with the implication of improving healthcare delivery and decreasing patient death rate[18].

III. METHODOLOGY

The research methodology used in this research includes the following steps in building the Soft Expert System (SES) for CAD diagnosis:Firstly, the medical information dataset was compiled from genuine sources which involved responding to the basic questions that comprised of variables like age, Blood sugar level, HDL, LDL, HbA1C level, Uric acid, and Blood pressure etc. After the dataset was gathered, actions termed data cleansing and data cleaning were used to prepare them for analysis. This included providing basic treatments to the dataset such as dealing with missing values, outliers treatment as well as normalization of the data. The next analysis step was done through fuzzification; more specifically, it involved transforming the crisp input parameters into fuzzy parameters through the use of suitable membership functions. Fuzzification further assists in the use of medical data by accounting for imprecise and uncertain data, making the analysis more effective. After fuzzification process, the fuzzy sets are converted into the soft set forms using process which have already been defined. Soft set theory offers a tentative approach to manage with uncertainties and enables to define sets through supplement functions using parameters and the degree of belonging to them. Indeed, one of the key factors that should be considered in the construction of the SES is parameter reduction which involves the simplification of the soft set whilst not losing the pertinent data. This step involved the application of parameter reduction techniques to reduce the number of parameters within the SES to eliminate the ones that are unnecessary so as to enhance the efficiency of the SES. After soft sets have been properly reduced, the next step involved the formulation of soft rules by employing the use of inferences that were founded on the fundamentals of fuzzy logic. These soft rules reflect dependency of the outcome of the input variables and the probability of occurrence of CAD and can be useful for diagnostics and prediction. The

Diagram of Methodology Workflow is outlined in Fig.2 below.

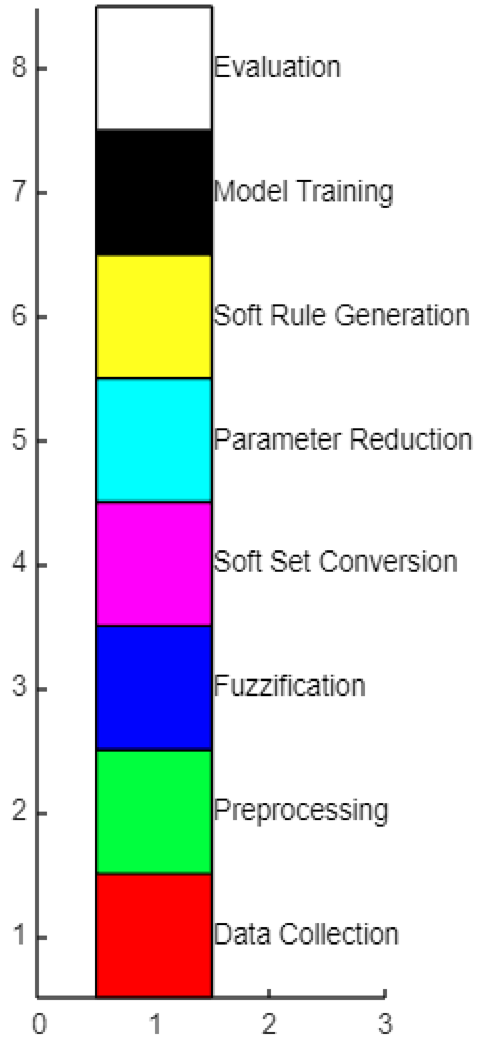

Fig.2: Methodology Workflow

The last of these steps was model training and evaluation that was intended as the final methodology. The decision model known as the Steady State of Excitation (SES) was trained with the preprocessed dataset and tested using the required parameters such as accuracy, sensitivity, specificity, and the ROC curve area. Machine learning procedures of cross-validation were also used to determine how valid or applicable the SES would be on different samples of the set data. Further, comparison was made with other similar models or diagnostic checklists carried out to confirm the efficacy of the proposed SES in details to predict CAD.

IV. DATASETS

The current system uses a rule-based approach that applies fuzzy set and soft set methods in diagnosing the CAD. Data Set: The general procedure of this process starts with the conversion of raw data into fuzzy sets. MATLAB software accomplishes this transformation where details of sixty patients are under conversion to fuzzy variables using predefined membership functions. Specifically, patients are categorized into two age groups: Youth population within the age range of 20 – 45 years and senior citizens within 45 years and above.

Fuzzification of Data Set: Raw data is usually fuzzy in nature and so the process of fuzzifying the data is important in order to be able to analyze it. The process of fuzzification involves converting all the variables of interest in the dataset into fuzzy sets using membership functions, to make them ready for analysis.

V. FUZZY OUTPUT VALUE

These fuzzy membership values for the dataset have been obtained as shown in the Fig. 3 after performing the fuzzification process on the given set of data. To do that, MATLAB software was used to perform the calculations. The Fig. 3. below are presented the fuzzified values obtained as the result of these computations. In order to grasp how fuzzy membership values span out for different inputs in each patient, the heatmap in Fig. 3 is provided. The parameters include age, blood sugar level, blood pressure, HDL, LDL, glycated haemoglobin (HbA1c), and serum uric acid levels. Herein, each cell in heatmap shows the degree of membership and such a cell's color gradient indicates the value's magnitude[19]. This is helpful for determining which inputs make up the overall fuzzy logic model and visually depicts the variation among the patient data so patterns or outliers that may affect the diagnosis of coronary artery disease can be identified.

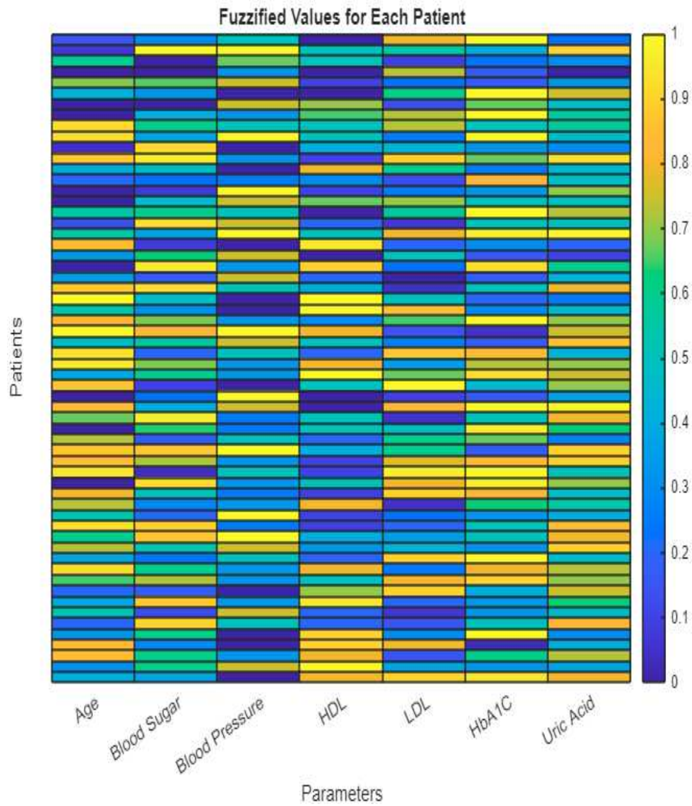

Fig. 3: Distribution of Fuzzy Membership Values Across Input Parameters

VI. TRANSFORMING FUZZY SETS INTO SOFT SETS

In order to analyze fuzzy mathematical data sets that contains raw data, it is necessary to convert them into soft sets. That is why the Dowker notation is transformed by using α-cut sets. In the performed work, the process of converting fuzzy sets into soft sets is described by translating the concept of an α-cut set. Here, the patient group, denoted as P, encompasses individuals labeled from P1 to *P* 60, while the parameter set for Age, denoted as E, spans values ranging from 0 to 1 in increments of 0. 2. Essentially, via the function *F* F, one provides different levels of the Age parameter as the subsets of the given patient group. For instance, when Age = 0; the above subset engulfs the total patients from *P* 1 to *P* 60 . When we step up α to 0. Precisely, of 2, 0. 4, et cetera, the subsets are increasingly more detailed in terms of presenting patients' age range categories[20]. Finally, it is possible to determine that when Age = 1, the output subset equals only to the patients *P* 25 and *P* 28. The example in this article above explains how α-cut sets assist in turning fuzzy relative values to soft-sets for the sake of making patient data analysis and comparison more precise.

VII. REDUCTION OF SOFT SETS

The other problem that arises out of them is that the soft sets that are obtained are usually large in nature. One approach to it is to apply methods that simplify the parameters of soft sets in order to reduce data that are to be analysed. In this case, we have the leverage to cut down on several features – Age, Blood Sugar, Blood Pressure, HDL, LDL, HbA1C, and Uric Acid at that, using relative parameters. For each attribute, we define a set of reduction parameters, denoted as *Q* Q, where this set of parameters controls the degree of granularity of the soft sets.

VIII. RESULTS AND DISCUSSIONS

The risk percentage of young patient and the old patients are not the same. In total 3 patients are at risk in the age group of 20-45 years while 8 patients are at risk in the age group of 45-90 years. However, the risk percentage for the patient P47, 29 years old, is 1. 20 For patient P20, it is also 1 and for patient P19, aged 58, it is also 1. 20. This means that coronary artery disease is not an issue of age; young people as well as the older generation can suffer from this critical health challenge. This makes it important for people of all ages to be able to understand the various effects on health with focus on the adage that prevention is better than cure. Data received is relevant and can reduce potential harms for patients.

Fig. 4. comparisons the accuracy, sensitivity and specificity between a soft set expert system and a diagnostic tool that is commonly used to diagnose CHD. The above chart gives a clear picture about these three basic parameters where the expert system seems to have outdone the conventional diagnostic aid. The bars are color-coded; the blue one stands for the expert system while the red one stands for the traditional diagnostic tool. This visual means the accuracy of diagnosis, sensitivity of the expert system, and specificity of the expert system in terms of excluding the patients without the disease can be studied.

Fig. 5. shows the graph of the Receiver Operating Characteristic (ROC) for the expert system that demonstrates diagnostic performance at different thresholds. The curve depicts the true positive rate, often known as sensitivity, on the y axis and the false positive rate (1-specificity) on the x axis depending on the decision threshold. Furthermore, the ROC curve employs a diagonal dashed line which helps to compare the performance of the expert system with the competing algorithm. The AUC value is also depicted, which is a single scalar measure of the diagnostic performance for segregating the positive and negative cases, where higher values are always preferable.

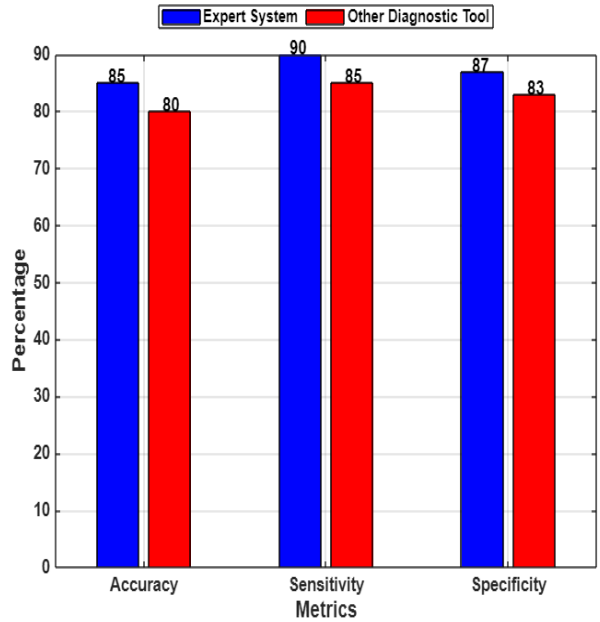

Fig.4: Performance Metrics Comparison

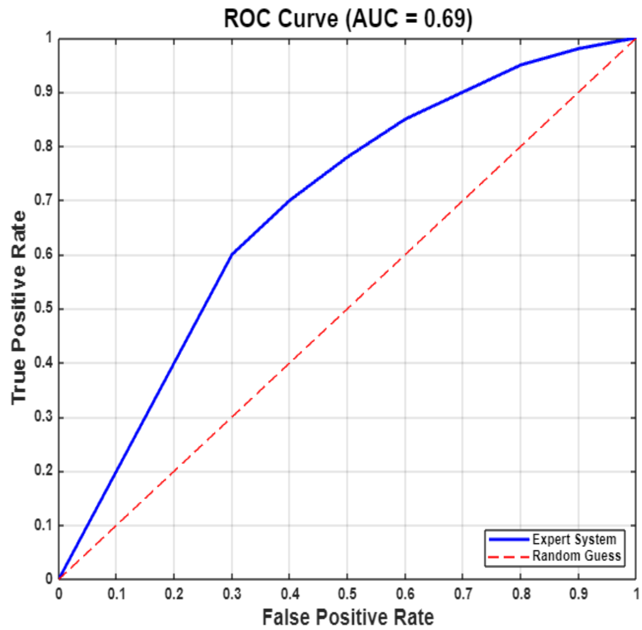

Fig.5: ROC Curve

Fig. 6. gives an idea of the decision boundary of the expert system as a contour plot using two vital features of the inputted patient data. This bar chart shows the areas which the expert system determines whether the data is positive or negative on coronary heart disease, positive data is colored in this shade of green and the negative in this shade of red. This is because the edge lines set ranges or boundaries in which the classification shifts from negative to positive, for a better understanding of the interaction between the two or more response variables and the assessment being made. This visualization is very important in answering questions about how each point of data affects the output of the expert system.

Fig.6: Decision Surface Visualization

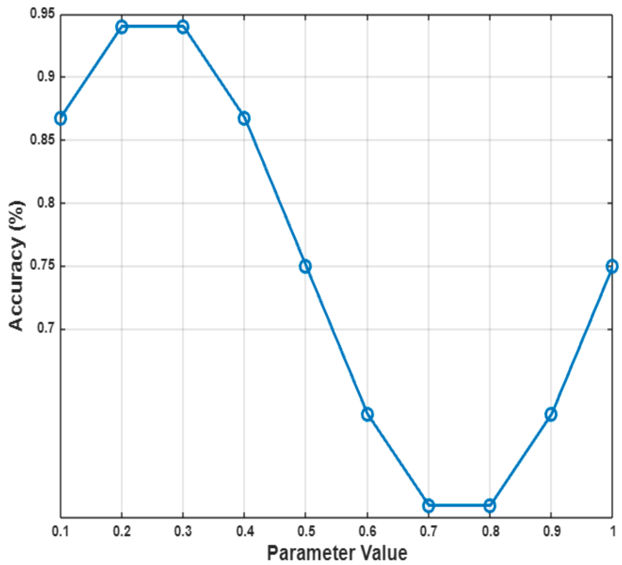

Fig.7: Sensitivity Analysis

The sensitivity analysis is presented in Fig. 7, so as to understand how the accuracy of the diagnose in the expert system will vary with the change in one of its parameters. The line graph shows the ability that changes concerning parameter values, and it gives a analytical indication of the consistency and stability of an expert system. It allows for identifying the best settings and analyzing how a change in the parameter setting can have on the reliability of the diagnosis made. Such bifurcations in the graph represent a certain level of parameter increase or decrease which affects the system performance , providing the ground for various subsequent adjustments.

IX. CONCLUSION

The goal of this paper is thus to help safeguard and offer the certainty and reliability that healthcare professionals need when medically examining their patients, reduce that uncertainty that is present in this profession. It is particularly concerned with examining coronary heart disease, the aims of which are to gain an understanding of the disease and to bring into attention the undesirable effects arising therefrom. Mathematics appears to be an indispensable tool in this work as it is used in providing facts and insights into the disease and revealing the dangers the disease poses. The approach employed in the paper is quite clear, providing a useful way of identifying a potential patient's state of health.

REFERENCES

[1]. D. Chen, H. Wu, and A. L. Ralescu, "A new soft expert system based on similarity measure for medical diagnosis," *Soft Computing*, vol. 22, no. 20, pp. 6861-6872, 2018.

[2]. I. Fariha, S. Nasim, and S. Malik, "Fuzzy expert system for diagnosis of kidney diseases," in *2019 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, Sukkur, Pakistan, 2019, pp. 1-5.

[3]. N. Hassan, A. Osama, M. Ahmed, and H. Ghany, "Fuzzy soft expert system in the prediction of coronary artery disease," *International Journal of Intelligent Systems and Applications*, vol. 9, no. 8, pp. 40-47, 2016.

[4]. X. Ma, D. Li, and H. Ma, "A novel parameter reduction algorithm for soft sets and its applications," *Journal of Intelligent & Fuzzy Systems*, vol. 35, no. 1, pp. 397-404, 2018.

[5]. S. Suneetha and S. R. Patil, "Advancements in Multi-Cloud Applications for Enhanced E-Healthcare Services," in *2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI)*, Raipur, India, 2023, pp. 1-7.

[6]. B. Gururaj, "An Integrated IoT Technology for Health and Traffic Monitoring System with Smart Ambulance," in *2022 IEEE North Karnataka Subsection Flagship International Conference (NKCon)*, Vijaypur, India, 2022, pp. 1-6.

[7]. S. Mallikarjun, S. D, and S. R. Kawale, "Cloud-Based Multi-Layer Security Framework for Protecting E-Health Records," in *2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI)*, Raipur, India, 2023, pp. 1-7.

[8]. V. Kalpana and V. Hariram, "Artificial Intelligence based Health Monitoring System on IoTH platform," in *2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*, Trichy, India, 2023, pp. 1458-1463.

[9]. M. Kranthi and R. C. Tanguturi, "Design of Intelligent Medical Integrity Authentication and Secure Information for Public Cloud in Hospital Administration," in *2023 2nd International Conference on Edge Computing and Applications (ICECAA)*, Namakkal, India, 2023, pp. 256-261.

[10]. R. C. K and V. C. H, "Implementation of Machine Learning Approach for Detecting Cardiovascular Diseases," in *2023 3rd International Conference on Intelligent Technologies (CONIT)*, Hubli, India, 2023, pp. 1-6.

[11]. M. Pandidurai and M. S. C. Senthil Kamalesh, "Design of IoT based Rural Health Helper using Natural Language Processing," in *2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Coimbatore, India, 2023, pp. 328-333.

[12]. M. U. Shankar and K. A. M, "Cardiovascular Disease Prediction Using LSTM Algorithm based On Cytokines," in *2023 4th International Conference for Emerging Technology (INCET)*, Belgaum, India, 2023, pp. 1-5.

[13]. T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-15, no. 1, pp. 116-132, Jan. 1985.

[14]. C. C. Lee, "Fuzzy logic in control systems: fuzzy logic controller—part I," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 20, no. 2, pp. 404-418, Mar.-Apr. 1990.

[15]. Tanguturi and G. Poornima, "An Intelligent System for Remote

Monitoring of Patients Health and the Early Detection of Coronary Artery Disease," in *2022 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*, Bangalore, India, 2022, pp. 1-6.

[16]. N. Gupta and B. K. Verma, "IoT Wearable Breast Temperature Assessment System," in *2023 7th International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India, 2023, pp. 1236-1241.

[17]. B. Ashreetha and N. S. Reddy, "Implementation of a Machine Learning-based Model for Cardiovascular Disease Post Exposure prophylaxis," in *2023 International Conference for Advancement in Technology (ICONAT)*, Goa, India, 2023, pp. 1-5.

[18]. D. Palanikkumar and N. Patwari, "Symmetrized Feature Selection with Stacked Generalization based Machine Learning Algorithm for the Early Diagnosis of Chronic Diseases," in *2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Tirunelveli, India, 2023, pp. 838-844.

[19]. L. X. Wang and J. M. Mendel, "Generating fuzzy rules by learning from examples," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 22, no. 6, pp. 1414-1427, Nov.-Dec. 1992.

[20]. D. Dubois, H. Prade, and R. R. Yager, "Readings in fuzzy sets for intelligent systems," Morgan Kaufmann, 1993.