MARGINALIZED FEATURE ANALYSIS OF CARDIO VASCULAR HEART DISEASE WITH DATA MINING CONCEPTS

Mr. R.VELUSAMY
Part-Time Research Scholar,
Department of Computer Science,
Government Arts and Science
College, Modakkurichi, Erode,
Tamilnadu – 638115,
velusamy.msc75@gmail.com

Dr. P.VIJAYAKUMAR
Assistant Professor of Computer Science,
School of Distance Education,
Bharathiar University, Coimbatore,
Tamilnadu – 641 046,
vijay.hodcs@gmail.com

Abstract---**Heart disease is one of the complex diseases and globally many people suffer from this disease. On time and efficient identification of heart disease plays a key role in healthcare, particularly in the field of cardiology. In this article,we proposed an efficient and accurate system to diagnose heartdisease and the system is based on machine learning techniques. The system is developed based on classification algorithms includes Support vector machine, Logistic regression, Artificial neural network, K-nearest neighbor, Naïve bays, and Decision tree while standard features selection algorithms have been used such as Relief, Minimal redundancy maximal relevance, Least absolute shrinkage selection operator and Local learning for removing irrelevant and redundant features. We also proposed a novel fast conditional mutual information feature selection algorithm to solve the feature selection problem. The Internet ofThings (IoT) is a substantive driving force to ICT technological advancement, leading prospective sectors down the road of automation alongside decentralized intelligence. The IoTevolves incessantly and impacts every facet of our life while resembling a living entity**

Index terms---*Heart Disease Classification, Features Selection, Disease Diagnosis, Intelligent System, Medical Data Analytics.*

I. INTRODUCTION

Information and communication technology (ICT) advancements have laid the foundation for innovative solutions in diverse industry domains, including healthcare, agriculture, transportation, and logistics, among others. The Internet of Things (IoT) is a substantive driving force to ICT technological advancement, leading prospective sectors down the road of automation alongside decentralized intelligence [1]. The IoT evolves incessantly and impacts every facet of our life while resembling a living entity. From household appliances to robots in factories, the IoT connects data, people, things/objects, and processes. Meanwhile, cloud computing (CC) delivers on-demand elastic services with virtually unlimited computationand storage capability [2]. Despite being unique and independent in their respective evolution, cloud computing and IoT aspects complement one another. Eventually, the two technologies converged in recent years, and the confluence became known as a Cloud–IoT paradigm [3,4], offering tremendous prospects for driving new innovative services and applications.

Analytics ensures the systematic quantitative and qualitative analysis of concerned data for efficient decision making, while predictive analytics stems from advanced analytics aiming to elicit the prognosis of future occurrences using the available data [6]. The analytics in healthcare is harnessed for clinical decision support, predictive risk assessment, and remote health monitoring, among other crucial tasks. Predicting and lowering risk based on current and past patient data are a big part of medicine. The integration of humongous data from disparate sources comprising electronic health records, medical imaging, screening results, and administrative information warranting swift decisions is effectively tackled by healthcare analytics [7]. Clinicians must often make decisions with a high degree of uncertainty; however, with the headway of predictive analytics in healthcare, those decisions will be more informed than ever. These cutting-edge predictive analytics approaches help identify trouble early on, avoid complication risks, improve chronic illness management, evade hospital readmission, receive medical research aid, and minimize overhead expenses.

A. MOTIVATION

Predictive analytics is proving its worth, not just in the hospital environment, but also at home by remote monitoring and keeping patients from relapsing into the need for acute treatment. Predictive analytics aid in the diagnosis, prognosis, and therapy at every stage of a patient's treatment [8]. It also helps in designing the treatment course, providing

clinical decision support, decreasing adverse occurrences, and enhancing the overall care quality while lowering healthcare costs. Moreover, the personalized healthcare model shifts fromtreating patients as numbers to treating them as individuals, customizing treatment to their unique medical history, environment, social risk factors, genetics, and biochemistry, among other things [9]—rather than depending on demographic statistics that do not apply to everyone. It tenders real-time clinical decision assistance at the point of treatment, allowing for the most efficient delivery of individualized healthcare [10]. With deadly diseases, spotting them early on and detecting anypossible deterioration in the patients' condition before occurrence can significantly improve the odds of an effective treatment.

The diseases affecting the heart and its related blood vessels are all classified as cardiovascular diseases (CVDs). These include arrhythmia, coronary artery disease, congenital heart disease, valve disease, aortic disease, heart failure, peripheral artery disease, pericardial disease, heart valve disease, cerebrovascular disease, rheumatic heart disease, deep vein thrombosis, cardiomyopathy, myocarditis, atrialfibrillation, ischemic heart disease, and stroke [11,12,13].

The most prominent cause of global mortality is cardiovascular diseases (CVDs), claiming the lives of an estimated 17.9 million individuals and accounting for 32% of all fatalities worldwide [14]. Heart attacks and strokes cause four out of every five CVD fatalities, which are 85% of all CVDmortalities, with one-third occurring before age 70. Identifyingindividuals at risk for CVDs and ensuring that they receive proper therapy can help avert untimely deaths. This is where thepredictive algorithms powered by AI and ML come into play alongside the Internet of Things, as these are adept at managingmassive and diverse data. Pattern classification, as a pattern recognition task, is a crucial supervised learning paradigm for identifying and classifying disease patterns in the medical field The researchers working on classification algorithms concerning heart disease strive to achieve the maximum classification accuracy possible as patients' lives are at stake.

Many individuals are at risk of heart disease due to long- term conditions such as persisting high blood pressure. With theincrease in the aging population across the globe, most of them are diagnosed with chronic heart conditions. This warrants the continuous real-time monitoring of individuals at in-home care and the patients in treatment within hospital premises, entailing timely treatment upon the fluctuation of vital signs. The prolonged tracking of health conditions in the elderly helpsminimize hospitalization cost and enhance the quality of life, but conventional methods are tedious and tiring. This necessitates efficient facilities to mitigate the overwhelming workload of clinicians and hospital staff while minimizing the cost of health monitoring. The pervasive nature of IoT has incited the proliferation of smart, interconnected devices and wearable devices with sensors, thereby facilitating remote patientmonitoring pertaining to heart disease. The IoT for healthcaremonitoring includes smart health watches, wearable blood pressure monitors, and wearable ECG monitors equipped with medical sensors. Thus, the healthcare IoT acquires vital patient data and transmits them to the Cloud for storage and complex deep learning analytics along with prior electronic clinical records for accurate heart risk diagnostics. These IoT devicescan swiftly notify the clinicians and caretakers of the patient's condition. This enables clinicians to better make timely decisions for individuals as well as the population at large by estimating patients' chance of developing a specific heart disease, their prognosis for the given condition, and the corresponding treatment

B. CONTRIBUTION

The pivotal outcomes of this research initiative arelisted as follows:

1. The data collected from IoT sensors pertaining to heart disease risk prediction are subject to the data pre-processingtasks of data cleaning and data filtering at the Cloud layer;
2. The ensuing data are sent to the fuzzy information system(FIS) for the initial classification task;
3. Finally, the proposed Bi-LSTM model is used toaccurately predict the risk of heart disease in patients.

The remaining sections of this article are organized into related work, methodology, experimental setup, performance assessment, experimental results and discussion, comparative analysis, future directions, and conclusions.

C. RELATED WORK

In recent times, diverse systems for heart disease predictions have been propounded. For enhancing heart disease risk prediction accuracy, the deployment ofseveral ensemble classifiers displays an accuracy of 85.4%[16]. A model for diagnosing heart disease disease that combines rough sets-based attribute reduction involving the chaos firefly algorithm with an interval type-2 fuzzy logic system showcases an accuracy of86% [17]. A machine learning hybrid model to predict heart disease [18] by combining random forest (RM) with linear method (LM) approaches exhibits a performance accuracy of88.7%.

D. DATA ACQUISITION/COLLECTION LAYER

The propounded healthcare system acquires data from twoprimary data sources. The physiological data of patients

such as their blood pressure (BP), heart rate, blood sugar/glucose level, respiration rate, blood oxygen, cholesterol level, activity, electrocardiogram (ECG), electromyogram (EMG),and electroencephalogram (EEG) are gathered from the patient's routine health monitoring. These data are transmitted through Bluetooth/Zigbee to related remote gateway devices and then to the cloud data center, where data pre-processing and disease prediction takes place. The other data source is the electronic clinical data (ECD), which comprise the patient's medical history (including their history of smoking and diabetes), observation reports, and comprehensive clinical (lab) reports which offer valuable information on disease prediction and are stored in a cloud database.

II. DATASET

For the experiment, to detect the presence of heart diseasefrom heart patient data, the Cleveland and Hungarian dataset from the UCI machine learning repository are considered.

A. DATA PRE-PROCESSING LAYER

Data pre-processing has become a requisite for ML algorithmdeployment as real-world data are prone to being inconsistent, incomplete, and noisy. Efficient heart disease prediction from the heart disease dataset requires missing data handling, normalization, and feature selection. Data acquired from wearable sensors are impacted due to signal aberrations, such asmissing values and noise, causing havoc in the case of heart disease prediction, compromising the prediction accuracy, or yielding an erroneous result. We utilize a well-known techniqueto filter the data known as Kalman filtering [**19**,**20**], which effectively eliminates duplicate records, noise, anddiscrepancies from the data. Owing to its simple form, it requires low computational power [**21**]. This unsupervised filtering algorithm is specialized to handle vast real-time sensordata and furnish values closer to that of the actual values from the sensor without noise [**22**]. In addition to this, we use two other unsupervised filters in the data filtering stage: removing useless and replace missing values [**23**]. With another 90% of maximum variance, the first filter eliminates irrelevant attributes. The second filter substitutes the mean as well as median values of the existing data for any values missing in thestructured dataset.

B. FUZZY INFERENCE SYSTEM

The term fuzzy refers to something as inexplicit or vague, and the fuzzy system is inspired by the requisite to model inherently vague real-world events [24]. The standard fuzzy system is characterized by four components, namely a fuzzifier,an inference engine, a knowledge base, and a defuzzifier. The inputs to a typical fuzzy system can be crisp data (numeric) andlinguistic values (fuzzy sets). In the case of a crisp input, the fuzzifier assigns to it the applicable fuzzy set and this process is known as fuzzification. Then, the inference engine accomplishes mapping of the input variable values to the linguistic values of the output variable through a suitable approximate reasoning method with expert knowledge indicated by a set of fuzzy conditional rules in the knowledge base. The knowledge base entails the application of domain knowledge which can be divided into a database and a rule base. The database comprises linguistic control rules, and the rule base includes domain expert knowledge. In addition tolinguist values, if numeric data output is needed, then defuzzification assigns crisp data to the resulting fuzzy set.

The classification of heart disease risk based on patients' health data is performed using a fuzzy inference system (FIS), and the algorithm is presented as Algorithm 1.

Algorithm 1: Classification of patients' health data using FIS.

Step 1. The inputs and the respective member functions µ1 determines the fuzzy system

Step 2. Ascertain heart disease risk state using µ1 (ECG1), µ1 (MaxHeartRate1),
µ1 (BloodPressure1) as µ1 (normal) or µ1 (low) or µ1(high)

Step 3. If Health risk state = µ1 (high)

***3.1** Send alert to GD using SPARK as RTA*

***3.2** Store Health risk state of the Puid in CS*

Step 4. Otherwise send Health risk state of the Puid to CS

Step 5. End the process

The inputs for maximum heart rate, ECG, and blood pressure, are created and member functions are fed, which are fuzzified into fuzzy sets using a fuzzy value range. Fig 1 presents the working of FIS for heart disease risk prediction.

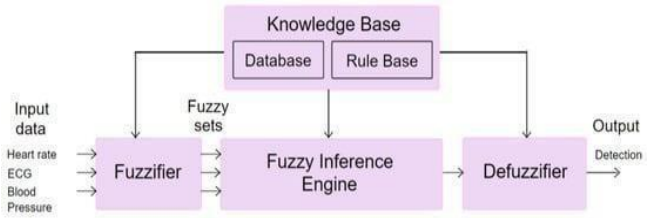

Fig.1. FIS for heart disease risk prediction.

The input variable value is mapped into the output variable's linguistic values through a suitable approximate reasoning method given as fuzzy conditional rules in the knowledge base. The results are classified in function of these

fuzzy rules in the rule base along with corresponding memberfunctions. The notification is sent regarding high-risk patients,and the overall patient risk status is stored in the cloud for future analysis. The data of patients classified as high risk for heart disease are subjected to further analysis in the ensuing prediction layer.

This section delves into the pursuit of the proposed system,and the findings are delineated. After the initial data pre-processing tasks involving data cleaning and data filtering, the ensuing data were examined withthree distinct models, with one model being the generic LSTM for disease prediction. The second model combines the fuzzy information system (FIS) and the LSTM denoted by FLSTM, where FIS is used to initially classify the heart disease risk statusof patients, but for prediction, the LSTM model is utilized. The third model, which the proposed work, combines the FIS with Bi- LSTM for heart disease prediction denoted by FBiLSTM. Thesethree models are assessed in accordance with the performance indices of accuracy, precision, sensitivity, specificity, and function measure concerning the patient heartdisease risk status.

Fig 2, Fig3, depicts the analysis of the accuracyand precision. F1-score displayed by LSTM, FLSTM, and the proposed models. The records are increased from 10% to 100% for the experiment of the threeConsidered models.

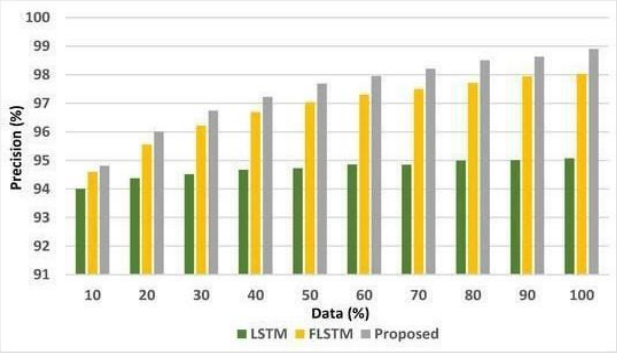

Fig.2. Precision analysis

Table.1. Precision for LSTM-FLSTM

Precision(%) Data(%)	LSTM	FLSTM	Propsed
10	94	94.6	94.8
20	94.4	95.7	96
30	94.6	96.2	96.8
40	94.7	96.8	97.2
50	94.8	97	97.8
60	94.9	97.2	98
70	94.9	97.5	98.2
80	95	97.8	98.6
90	95	98	98.8
100	95	98	98.9

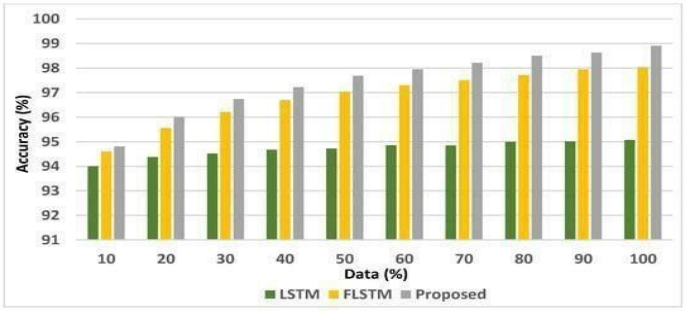

Fig.3. Accuracy analysis.

Table.2. Accuracy for LSTM-FLSTM

Accuracy(%) Data(%)	LSTM	FLSTM	Proposed
10	94	94.6	94.8
20	94.4	95.5	96
30	94.5	96.1	96.7
40	94.7	96.8	97.2
50	94.8	97	97.8
60	94.9	97.2	98
70	94.9	97.5	98.1
80	95	97.8	98.5
90	95	98	98.7
100	95	98	98.9

III.COMPARATIVE ANALYSIS

The proposed work is assessed in terms of prediction accuracy with cutting-edge approaches that harness heart disease datasets. The comparison study analyzing the proposed model's accuracy results with the existing literaturein the order of increasing accuracy. Fig 4 portrays the comparison of the performance results of the proposed model with the existing systems.

The results of the comparison with the related state of the heart disease predictive systems reveal that the proposed system's performance surpasses that of the existing systems. Major IoT-driven tasks for real-time smart systems involving healthcare warrant rapid processing as such applications are delay and context-sensitive. The escalation in the number of IoT devices and the upsurge in the data generated by the smart devices has resulted in immense data traffic resulting in extensive bandwidth utilization and service difficulties.

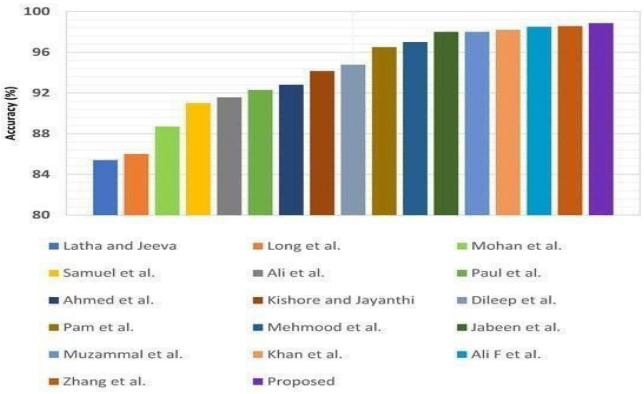

Fig.4. Comparative Analysis

As theCloud–IoT model suffers from limitations such as latency, connectivity, and bandwidth utilization, the cloud computingmodel seems inadequate to manage these challenges solely due to its centralized model. These shortcomings set the stage fordecentralized models of edge computing (EC) and fog computing (FC), wherein computation and storage can behandled at the edge nodes closer to the data source. These newercomputing technologies complement the Cloud and serve as anextension to it while enabling artificial intelligence tasks at theedge nodes.This hierarchical edge–fog–cloud modelconsiderably reduces the delay constraints by efficiently handling the humongous data acquired by the IoT devices whilemitigating latency. Thus, the proposed cloud-based predictionsystem can be deployed at the fog/edge layers in the future toovercome the Cloud's intrinsic constraints, such as increasedlatency and bandwidth use, while managing the surge in IoTdata.

IV. CONCLUSION

In this research initiative, an IoT–Cloud-based smart healthcare system for heart disease cardiovascular risk prediction is proposed, and the fuzzy inference system (FIS) and the recurrent neural network's bidirectional LSTM are harnessed for the predictive task. The proposed system's accuracy, precision, sensitivity, specificity, and F1-score are 98.85%, 98.9%, 98.8%, 98.89%, and 98.85%, respectively, outperforming other state of the art heart disease prediction models. This is just one facet of the healthcare research being done with predictive analytics, with a huge potential of deep learning models yet to uncover. The efficacy of the healthcare domain can be revolutionized with precise and timely disease predictions alongside rapid responses and agile decision-making by clinicians, which will improve the overall quality-of-service when fog/edge computing is involved.

Abbreviations

GD	Gateway device
CS	Cloud server
$\mu 1$	Membership function
RTA	Real-time Analyzer
Puid	Unique identification number of patient

References

[1] Bhatia, M.; Sood, S.K. Game Theoretic Decision Makingin IoT-Assisted Activity Monitoring of Defence Personnel. *Multimed. Tools Appl.* **2017**, *76*, 21911–21935.

[2] .Firouzi, F.; Farahani, B.; Marinšek, A. The Convergenceand Interplay of Edge, Fog, And Cloud in the AI-Driven Internet of Things (IoT). *Inf. Syst.* **2022**, *107*, 101840.

[3] Biswas, A.R.; Giaffreda, R. IoT and Cloud Convergence:Opportunities and Challenges. In *2014 IEEE World Forum on Internet of Things (WF-IoT)*; IEEE: Manhattan, NY, USA,2014

[4] .Botta, A.; de Donato, W.; Persico, V.; Pescapé, A. Integration of Cloud Computing and Internet of Things: A Survey. *Future Gener. Comput. Syst.* **2016**, *56*, 684–700.

[5] Santos, G.L.; Takako Endo, P.; Ferreira da Silva Lisboa Tigre, M.F.; Ferreira da Silva, L.G.; Sadok, D.; Kelner, J.; Lynn, T. Analyzing the Availability and Performance of an E-Health System Integrated with Edge, Fog and Cloud Infrastructures. *J. Cloud Comput. Adv. Syst. Appl.* **2018**, *7*, 16.

[6] Suresh, S. Big Data and Predictive Analytics. *Pediatr. Clin. N. Am.* **2016**, *63*, 357–366.

[7] Simpao, A.F.; Ahumada, L.M.; Gálvez, J.A.; Rehman,M.A. A Review of Analytics and Clinical Informatics in Health Care J. Med. Syst. 2014, 38, 45

[8] Miotto, R.; Wang, F.; Wang, S.; Jiang, X.; Dudley, J.TDeep Learning for Healthcare: Review, Opportunities and Challenges. Brief. Bioinform. 2018, 19, 1236–1246 Pandey, S.; Janghel, R. Recent Deep Learning Techniques, Challenges and Its Applications for Medical Healthcare System: A Review. Neural Process. Lett. 2019, 50, 1907–1935. [Google Scholar] [CrossRef] Pandey, S.; Janghel, R. Recent Deep Learning Techniques, Challenges and Its Applications for Medical Healthcare System: A Review. Neural Process. Lett. 2019, 50, 1907–1935. [Google Scholar] [CrossRef]

[9] Muniasamy, A.; Tabassam, S.; Hussain, M.; Sultana, H.; Muniasamy, V.; Bhatnagar, R. Deep Learning for Predictive Analytics in Healthcare. In Advances in Intelligent Systems and Computing; Springer: Cham, Switzerland, 2019; pp. 32–42. [Google Scholar] [CrossRef]

[10] Smys, S. Survey on accuracy of predictive big data analytics in healthcare. J. Inf. Technol. Digit. World 2019, 01, 77–86. [Google Scholar] [CrossRef]

[11] Amin, P.; Anikireddypally, N.; Khurana, S.; Vadakkemadathil, S.; Wu, W. Personalized Health Monitoring Using Predictive Analytics. In Proceedings of the 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService), Newark, CA, USA, 4–9 April 2019. [Google Scholar]

[12] Joseph, P.; Leong, D.; McKee, M.; Anand, S.S.; Schwalm, J.-D.; Teo, K.; Mente, A.; Yusuf, S. Reducing the Global Burden of Cardiovascular

[13] Disease, Part 1: The Epidemiology and Risk Factors: The Epidemiology and Risk Factors. Circ. Res. 2017, 121, 677–694. [Google Scholar] [CrossRef]

[14] Fuchs, F.D.; Whelton, P.K. High Blood Pressure and Cardiovascular Disease. Hypertension 2020, 75, 285–292. [Google Scholar] [CrossRef] [PubMed