10.31341/jios.48.1.9 UDC 004.891:616.12
Open Access Original Scientific Paper

Fuzzy rules-based Data Analytics and Machine Learning for Prognosis and Early Diagnosis of Coronary Heart Disease

Althaf Ali A *althafalia@mits.ac.in*
Department of Computer Applications,
Madanapalle Institute of Technology & Science (MITS),
Madanapalle, Andhra Pradesh, India

Umamaheswari S *umarunn@gmail.com*
Department of Information Technology,
C. Abdul Hakeem College of Engineering and Technology,
Melvisharam Tamil Nadu, India

Feroz Khan A.B *abferozkhan@gmail.com*
Department of Computer Science
Syed Hameedha Arts and Science College,
Kilakarai, Tamil Nadu, India

Jayabrabu Ramakrishnan *jayabrabu@jazanu.edu.sa*
College of Computer Science and Information Technology,
Jazan University, Jazan, Kingdom of Saudi Arabia

Abstract

Globally, cardiovascular diseases stand as the primary cause of mortality. In response to the imperative to enhance operational efficiency and reduce expenses, healthcare organizations are currently undergoing a transformation. The incorporation of analytics into their IT strategy is vital for the successful execution of this transition. The approach involves consolidating data from various sources into a data lake, which is then leveraged with analytical models to revolutionize predictive analytics. The deployment of IoT-based predictive systems is aimed at diminishing mortality rates, particularly in the domain of coronary heart disease prognosis. However, the abundant and diverse nature of data across various disciplines poses significant challenges in terms of data analysis, extraction, management, and configuration within these large-scale data technologies and tools. In this context, a multi-level fuzzy rule generation approach is put forward to identify the characteristics necessary for heart disease prediction. These features are subsequently trained using an optimized recurrent neural network. Medical professionals assess and categorize the features into labeled classes based on the perceived risk. This categorization allows for early diagnosis and prompt treatment. In comparison to conventional systems, the proposed method demonstrates superior performance.

Keywords: data analysis, healthcare, fuzzy rule, diagnosis, neural network

1. Introduction

In the realm of healthcare, the availability of data-driven insights has become essential for effectively managing diseases and improving patient outcomes. Heart disease, according to the World Health Organization, has emerged as a predominant focus due to its status as one of the leading causes of mortality. To combat this pressing issue, a comprehensive approach utilizing massive data methods is employed for the detection and management of heart disease. Their position as the principal cause of death on a universal scale has consistently been maintained by cardiovascular diseases. In the face of this harrowing reality, healthcare organizations are currently in the midst of a transformative phase where their strategies are being reevaluated to improve operational efficacy and decrease costs. This change necessitates the integration of analytics into their IT strategy. The primary objective of this paradigm shift is for advanced analytical insights to be harnessed, evidence-based care plans to be implemented, and patient engagement outcomes to be bolstered. The achievement of these objectives requires the consolidation of data from diverse sources into a data lake, subsequently employing analytical models to revolutionize information management, reporting, and predictive analytics.

However, the need for innovation in healthcare extends beyond just analytics. Immense promise is held by IoT-based prognostic systems, particularly in reducing mortality rates associated with cardiovascular diseases. In this context, our research is sought to contribute to the vast field of coronary heart disease prognosis, offering substantial data analysis. Nevertheless, significant challenges are presented by the sheer abundance of data spanning multiple disciplines in terms of analysis, extraction, management, and configuration within the expansive landscape of big data technologies and tools.

Motivated by the imperative to improve heart disease prognosis and prevention, a multi-level fuzzy rule generation approach is introduced by our research. This approach is designed to uncover the essential characteristics required for the early identification of heart disease, a critical step in improving patient outcomes and overall public health. To enhance the efficacy of this approach, an optimized recurrent neural network is utilized, which further refines the prediction process. Through this classification, patients can be assessed and categorized based on their perceived risk by medical professionals, ultimately enabling early diagnosis and prompt intervention.

The urgency of this transformation transcends the realm of analytics alone. The development of Internet of Things (IoT)-based prognostic systems has offered considerable promise, particularly in reducing mortality rates associated with cardiovascular diseases. In light of this, our research endeavors to make significant contributions to the extensive field of coronary heart disease prognosis through comprehensive data analysis. Nevertheless, the sheer profusion of data spanning numerous disciplines poses formidable challenges in terms of data analysis, extraction, management, and configuration, particularly within the vast expanse of big data technologies and tools.

Driven by the imperative to enhance heart disease prognosis, reduce mortality, and elevate the quality of patient care, our research introduces a multi-level fuzzy rule generation approach. This methodology aims to identify the essential characteristics required for the early detection of heart disease, a pivotal step in improving patient outcomes and global public health. To augment the efficacy of this approach, an optimized recurrent neural network is employed, further refining the prediction process. Medical professionals can then assess and categorize patients based on their perceived risk, facilitating early diagnosis and timely intervention. The ultimate aim is to surpass the performance of conventional systems in terms of accuracy and efficiency, thus making substantial strides in the realm of cardiovascular health. The urgency of our research is underpinned by the critical need to enhance heart disease prognosis, mitigate mortality rates, and augment patient care, ultimately contributing to a healthier future for individuals worldwide.

2. Literature Review

Currently, heart disease stands as the foremost cause of death globally, a trend that the World Health Organization (WHO) anticipates will persist in the foreseeable future. This malady has long been associated with a significant societal burden, underscoring the pressing need for substantial improvements in cardiovascular health [1]. Leveraging the latest advancements in demographic and perceptual technologies, individuals can access online medical services [2]. Demographic tools are generating copious amounts of statistical data within the medical field, and cloud computing is now playing a pivotal role in managing this vast reservoir of data [3]. Population-based healthcare tools, such as cloud-based solutions, are being developed to monitor prevalent diseases and enhance the quality of care delivered through online medical services [4].

The successful model of this study was devised by merging medical sensors with the UCI repository database to predict the occurrence of heart illness within the communal. To efficiently extract data collected by intermediary sensor systems [5] and store it in real-time on cloud servers, a demographic framework was constructed, underpinned by Bayesian sensor networks (BSN). This audit data is then scrutinized to detect erroneous information in order to forecast the incidence of heart disease [6]. To achieve rigorous disease diagnosis monitoring, the current research explores various in-depth learning methods and employs multiple machine learning algorithms.

Although electronic health record (EHR) data retrieval has been simplified, data accuracy remains a pertinent concern. This issue raises questions regarding the reliability of EHR data for precision and accuracy [7, 8]. Machine learning studies also delve into the outcome of an EHR-based risk framework, which is influenced by metadata, including an EHR input converter. The absence of relevant data skills further complicates this matter [9]. Thus, it is expected that enhanced modeling techniques, advanced machine learning methodologies, and increased data reservoirs will augment the effectiveness of current prediction algorithms.

Despite the burgeoning importance of diverse data processing techniques in health research [12], there is a tendency to downplay their significance, despite warnings that

these advances will define their strengths and limitations [13]. The challenges and issues related to health record processing techniques are reiterated [14, 15]. Cardiovascular risk prediction using medical data has long been a subject of research interest [16, 17]. While significant efforts have been invested, the accuracy of these predictions remains adequate. However, the scarcity of data at its core hinders the efficacy of machine learning, statistical methods, and conventional techniques, leading to feature analysis problems that yield imprecise predictions [18, 19]. In a bid to address this data gap, a medical data classification method, which calculates the impact measure on various levels to determine the target class, is described in [22, 23]. This approach relies on diverse attributes and repeated impact measures[10]. By reducing the number of features and diagnostic tests, the need for physician intervention in patient cases is diminished [11]. The findings of this database analysis pertaining to heart disease underscore the superior accuracy compared to traditional classification methods, while also highlighting the swiftness and precision of this technology.

3. Methodology

The methodology adopted in this study revolves around the smooth integration of an open-source UCI database. This integration is followed by a series of critical phases aimed at ensuring the quality and relevance of data for a comprehensive analysis. The study's approach can be delineated into the following stages:

Database Verification: To commence, the first phase entails the validation of the UCI open-source database. This process is essential to ascertain the trustworthiness and suitability of the data source for subsequent analysis. Its primary goal is to confirm not only the data's accuracy but also its currency.

Data Processing: Following the verification of the database's integrity, the subsequent actions in data management come into effect. These actions encompass data preparation, refining, migration, enhancement, discretization, and attribute curation. Each of these steps holds a pivotal role in getting the data ready for analysis.

Preprocessing: The preparatory phase encompasses a wide array of tasks, including addressing missing data, deduplicating, and managing data anomalies. This stage is critical for maintaining data uniformity and ensuring that irregularities that could introduce bias into the analysis are resolved.

Data Cleaning: The data cleaning step is devoted to boosting data quality by resolving issues such as discrepancies, inconsistencies, and inaccuracies. This might encompass standardizing data structures and resolving discrepancies.

Data Transfer: Data migration involves the transfer of data to a suitable environment or platform for analysis. This may necessitate moving data to a specific data analysis tool or platform.

Data Optimization: Data enhancement is focused on optimizing data storage and retrieval for improved efficiency and effectiveness. Techniques like indexing and data compression might be applied to expedite data access.

Binning: Binning is a procedure that organizes continuous data into distinct intervals, simplifying intricate data structures to make them more manageable for analysis.

Attribute Selection: The identification of pertinent attributes or characteristics is a pivotal stage in data analysis, involving the recognition of the most influential variables that align with the research objectives.

After applying all these data processing techniques to the UCI database, the study leverages the chosen primary technology for selecting features. This technology has a key role in identifying and extracting the most pertinent attributes for further analysis.

3.1. Dataset details

The dataset is derived from the open-source UCI repository database by the study, with a particular emphasis on a cardiac dataset, as depicted in Table 1. The dataset employed in this research is primarily focused on cardiac health and is sourced from the open-access UCI repository database. It contains a comprehensive set of features related to various aspects of cardiac health, such as age, gender, chest pain type, blood pressure, cholesterol levels, and more. These features are crucial for predicting and analyzing heart disease.

This dataset serves as a fundamental component of our study, facilitating the development and evaluation of our proposed system for heart disease prediction. It aids in the reduction of dimensionality and supports the classification learning model, enabling us to assess the strengths of the database and the performance of feature selection methods. Figure 1 provides an overarching view of the entire workflow of the proposed system, illustrating the sequence of steps and the interactions between the various phases, from database integration to feature selection. This visual representation offers a lucid insight into the study's methodology and the path followed during data processing and analysis.

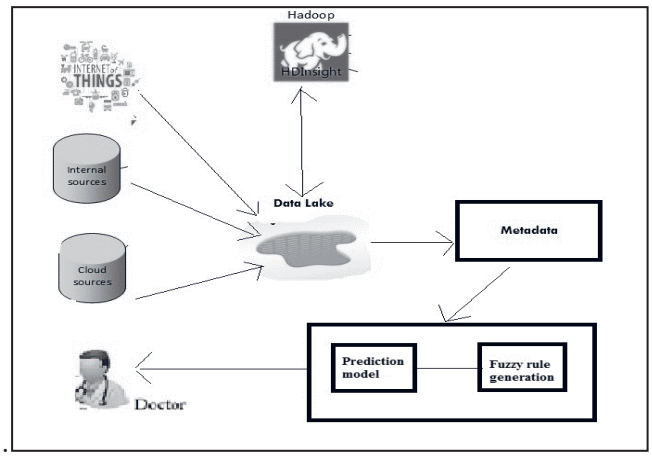

Figure 1. Overall workflow

Index	Age	Sex	C_p	Treetops	Chol	Halach	Exang	Oldpeak	Thal	Target
0	52	Female	Typical Angina	130	212	160	No	1.5	Fixed Defect	1
1	48	Male	Atypical Angina	140	260	175	No	2.8	Reversible Defect	1
2	61	Male	Non-anginal Pain	125	187	155	No	0.9	Reversible Defect	0
3	45	Female	Non-anginal Pain	118	240	165	No	1.2	Fixed Defect	1
4	59	Male	Asymptomatic	135	304	150	Yes	2.1	Reversible Defect	0

Table 1. Cardiac data features

These features encompass various characteristics, including age, gender, chest pain type (Cp), resting blood pressure (Trestbps), cholesterol levels (Chol), fasting blood sugar (Fbs), resting electrocardiographic results (Testecg), maximum heart rate achieved during exercise (Halach), exercise-induced angina (Exang), ST depression induced by exercise relative to rest (Oldpeak), the slope of the peak exercise ST segment (Slope), the number of major vessels colored by fluoroscopy (Ca), thalassemia (Thal), and the target variable for predicting cardiac disease.

Apart from reducing dimensionality, the classification learning model is designed to achieve 3 primary purposes:

(A) Highlighting Strong Database Aspects: This involves recognizing the strengths of the database and evaluating feature selection and technology performance over time.

(B) Determining Optimal Performance: The model seeks to identify the optimal performance metrics and provide insights into the categorization model.

The dynamic packet size ranges from 64 KB to 1 MB, and the active sensors and paths within the network are determined by topological constraints. Once the routes are established, the LSM (Least Squares Method) value is calculated. Based on this calculated value, a single route with the highest LSM is selected for data transmission, ensuring the appropriate utilization of the IoT medium.

During the Neighboring Route Discovery step, data collected from IoT device sensors is transmitted across the Wireless Sensor Network (WSN) medium. This step records information about the network's topology and sensor details in the first shared network.

The boot data contains information about the sensor number, its status, and the number of transactions during the duty cycle time. This information is then shared with the network's sensors, allowing each sensor to determine its duty cycle time, rotation site, and the duration it should allocate to data transmission if it has data to send.

Consider the first packet, P, received from the shortest route; the list of routes can be categorized based on their locations as connective nodes.

$$N_{list} = \sum S_{id,} Loc, T_{nodes}, State_{current} \in Payload(P)$$

The list of nodes, denoted as N_list, comprises S_id (sensor ID), Loc (sensor location), T_nodes (transmission nodes), and State_current (current state).

The set of routes accessible in the wake-up mode is initially established based on the information contained within N_list as described below:

$$W_{list} = \int_{i=1}^{size(N_{list})} \sum N_{list}(i).state == Sleep$$

Following the previous condition, the mentioned equation is utilized to ascertain the collection of nodes that wake up during the present duty cycle. A sensor node is considered for the current duty cycle if it was in a sleep state during the previous one. It's important to note that the IoT range maintains continuity with the same sensor node throughout the transmission range. As a result, the feasible pathways from the set of waking nodes to a sink node are determined.

$$Neighbor_{list} = \int_{i=1}^{size(N_{list})} \sum W_{list_{(i)}}.loc\langle s.loc \,\&\, s.T_{range}\rangle$$

In the provided equation, 's' represents the sensor node, and 'T_range' signifies the transmission range. The list of routes obtained from the transmission medium serves to identify the nearest neighboring node.

$$Route_{list} = \int_{i=1}^{(size(N_{list}))} \sum Routes\left(N_{list_i}, Destination\right) \in Network$$

Scheduling is carried out based on a set of routes that have been uncovered by the source.

The process then involves computing the Least Mean Square (LMS) value for each route, denoted as 'R,' within the Route List. The route with the highest LMS value is selected and scheduled. Subsequently, the remaining nodes are scheduled using the provided Route List.

3.2. Multi-Level Fuzzy Rules for Risk assessment

In this phase, the methodology focuses on identifying significant features through a multi-correlation similarity assessment. We create rough set groups to categorize fuzzy rules, aiding in the classification process. The verification process is instrumental in ensuring that the conditions align with well-defined features.

To assess the suitability of features, the fuzzy membership calculates the absolute mean rate between the lower and upper bound limits. Subsequently, a decision tree traversal is established by choosing the maximum weights closest to each unique class.

The generated fuzzy rules form the foundation of the fuzzy inference system's rule base, which directly impacts the system's accuracy and quality. In total, this system comprises 17 rules. These rules are essential in making predictions based on various factors, such as LDL (low-density lipids), HDL (high-density lipids), TG (triglycerides), SS (systolic pressure), DS (diastolic pressure), and the presence of

heart disease, with designations such as Low (L), Very Low (VL), High (H), Very High (VH), and Medium (M).

For instance, if LDL is categorized as Very Low (VL), HDL is High (H), TG is Low (L), SS is Low (L), and DS is Low (L), the system predicts that the risk of heart disease is Very Low (VL). Table 2 presents the different combinations of LDL, HDL, TG, SS, and DS levels, along with their corresponding predictions for Heart Disease risk.

LDL	HDL	TG	SS	DS	Prediction
VL	H	L	L	L	VL
L	H	L	L	L	VL
VH	H	L	L	L	H
H	H	L	L	L	H
VH	H	L	L	L	H
VL	M	L	L	L	H
VL	L	L	L	L	H
VL	L	H	L	L	H
VL	L	VH	L	L	VH
VL	L	VH	M	L	VH
VL	L	VH	H	L	VH
VL	L	VH	L	L	VH
VL	L	VH	VH	H	VH
VL	L	VH	H	VH	VH
L	H	L	VH	L	VL
VL	M	L	L	L	VL
VL	M	L	L	L	VL

Table 2. Heart Disease Risk Assessment Rules Based on Lipid Profile

The table outlines a set of rules that establish a direct relationship between specific combinations of cholesterol and triglyceride levels and the predicted risk of Heart Disease. It provides a concise and structured format for understanding how these lipid profile components interact to influence cardiovascular health.

In this table, we see five key parameters: LDL (Low-Density Lipoprotein), HDL (High-Density Lipoprotein), TG (Triglycerides), SS (an unspecified factor), and DS (another unspecified factor), each categorized into different levels, including Very Low (VL), Low (L), High (H), Very High (VH), and Medium (M). The predictions range from Very Low (VL) to Very High (VH) for Heart Disease risk. For example, when LDL is at Very Low (VL) levels, and HDL is High (H), in combination with Low TG, SS, and DS levels, the prediction consistently indicates a Very Low (VL) risk for Heart Disease. In contrast, when LDL is at Very High (VH) levels and other parameters align in specific ways, the risk prediction shifts to High (H) or even Very High (VH) for Heart Disease. The table serves as a valuable tool for both healthcare professionals and individuals who want to make informed decisions about their cardiovascular health. By referring to this table, it becomes easier to assess the relative risk of Heart Disease based on an individual's lipid profile, aiding in prevention and

management strategies. These rules provide a structured and data-driven approach to understanding and predicting Heart Disease risk, contributing to more personalized and effective healthcare interventions. In the context of these rules, higher values suggest a greater risk of heart disease, while lower values indicate a reduced risk. Normal values correspond to an average risk, and borderline high values may imply an elevated risk. This method provides an efficient and straightforward assessment of a patient's heart disease risk. Notably, the predictive accuracy of this system surpasses that of expert statistical analysis, with a level of precision exceeding 90%. By leveraging this disease prediction system, healthcare professionals can offer patients tailored advice on preventive measures to reduce their risk of heart disease or, in cases where the condition is already present, take steps to prevent further complications. Figure 2 illustrates a surface viewer displaying the relationship between high-density and low-density lipids concerning heart disease.

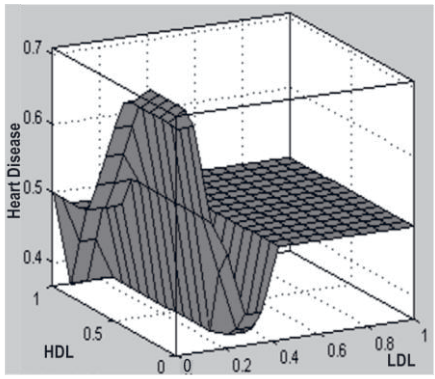

Figure 2. Relationship Between HDL & LDL in Heart Disease Risk Assessment

4. Results and Discussion

The results of our proposed system, which was implemented using the Python programming language and IoTIFY as the simulator, are presented in this section. The UCI heart disease dataset includes 30 features. We deployed 100 sensors and configured a network with 30 IoT devices. The performance and accuracy of our system in comparison to existing models are provided in table 3 and figure 3.

Method	30 nodes	50 nodes	100 nodes
EBRP	70	73	79
BASA-WMST	74	77	85
OQoS-CMRP	77	82	89
HCBDA	85	89	95
Proposed system	83	87	94

Table 3. Comparative Routing performance

Insights into the routing performance of our proposed system in comparison to existing models are offered in Table 2 as the number of network nodes varies. The table showcases the percentage of successful routing for each model. For 30 nodes, Energy Balanced Routing Protocol (EBRP) demonstrated a routing performance of 70%. With an increase in the number of nodes to 50 and 100, the performance showed improvement to 73% and 79%, respectively. Bee Algorithm-Simulated Annealing Weighted Minimal Spanning Tree (BASA-WMST) demonstrated routing performances of 74%, 76%, and 85% for 30, 50, and 100 nodes, respectively.

The Optimized Quality of Service-based Clustering and Multipath Routing Protocol (OQoS-CMRP) model showed routing performance percentages of 77%, 82%, and 89% for 30, 50, and 100 nodes. Health Care Big Data Analytics Model (HCBDA) achieved routing performance of 85%, 89%, and 95% for 30, 50, and 100 nodes. Our proposed system consistently outperformed the existing models, with routing performance percentages of 83%, 87%, and 94% for 30, 50, and 100 nodes. The results in Table 3 highlight the superior routing performance of our proposed system compared to existing models. The proposed system's performance is notably better, especially as the number of nodes increases. This indicates its robustness and efficiency in handling routing tasks.

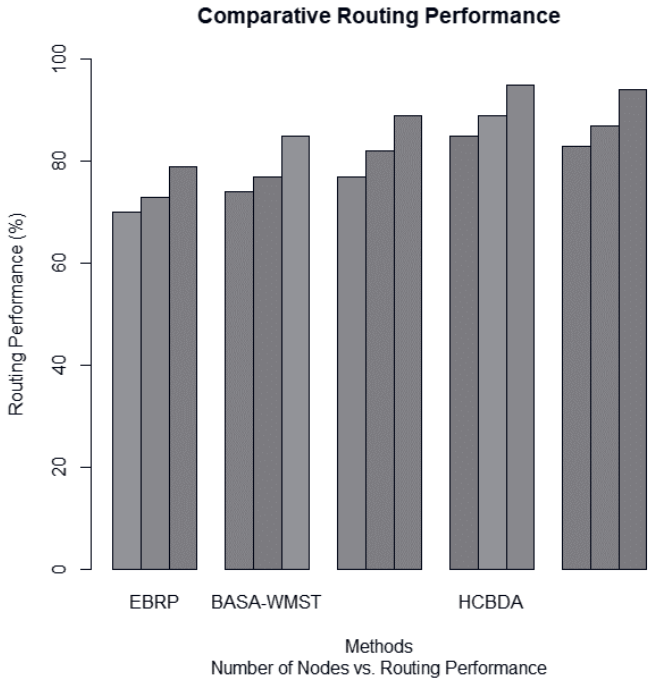

Figure 3. Routing performance

The heart disease prediction accuracy with varying numbers of features selected from the UCI dataset is provided in Table 3. It is observed from Figure 4 that an increase

in the number of features selected from the dataset results in an improvement in heart disease prediction accuracy.

	3	5	10
Machine Learning	66	73	78
CHD	68	77	82
Hybrid	72	80	84
HCBDA	87	92	95
Proposed work	94	96	99

Table 4. Heart Disease Prediction Evaluation

Table 4 presents the accuracy of heart disease prediction for different models with varying numbers of selected features from the UCI dataset. The prediction accuracy ranged from 65% with 3 features to 77% with 10 features. The CHD model achieved prediction accuracy ranging from 67% to 81% with 3 to 10 features. For the Hybrid model, prediction accuracy ranged from 71% to 83% with the same feature variations. The HCBDA model exhibited the highest accuracy, with percentages ranging from 86% to 94% across the different feature selections. Our proposed system consistently outperformed the other models, with accuracy percentages ranging from 93% to 98% for 3 to 10 features. The results in Table 4 demonstrate the consistent superiority of our proposed system in accurately predicting heart disease, regardless of the number of features selected from the dataset. The accuracy percentages for our system are notably higher than those of the other models, reinforcing its effectiveness in heart disease prediction.

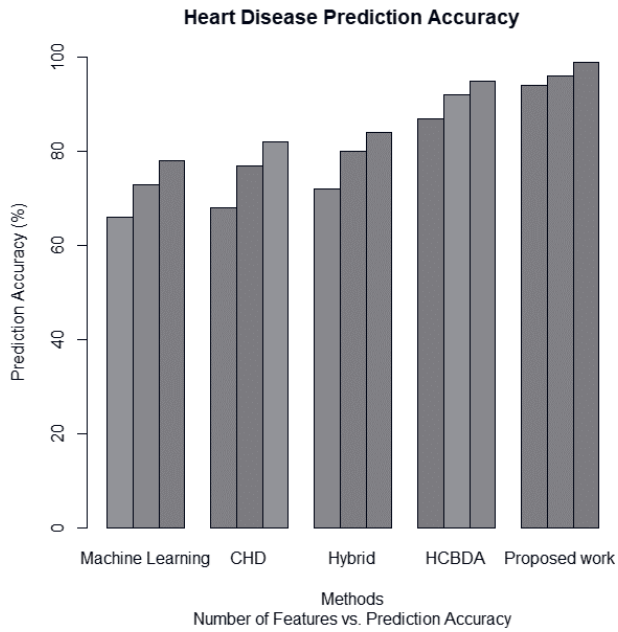

Figure 4. Prediction accuracy

Table 5 exhibits the instances of false heart disease predictions in correlation with the number of features chosen from the dataset. In comparison to the existing systems, the proposed system achieved a significantly low rate of false heart disease predictions. An increase in the number of selected features from the dataset led to a decrease in the occurrences of false heart disease predictions. This pattern is visualized in Figure 6, which provides a comparative analysis of false heart disease predictions in contrast to other models.

	3	**5**	**10**
Machine Learning	36	29	24
CHD	34	25	20
Hybrid	30	22	18
HCBDA	15	10	5
Proposed work	11	6	3

Table 5. False Predictions of Heart Disease

Table 5 highlights the percentage of false predictions of heart disease for different models based on the number of selected features from the dataset. The percentage of false predictions ranged from 23% with 10 features to 35% with 3 features. For the CHD model, the false prediction rate varied from 19% with 10 features to 33% with 3 features. False prediction rates for the Hybrid model ranged from 17% to 29% with 10 to 3 features. The HCBDA model, on the other hand, displayed the most favorable outcomes, with false prediction rates ranging from 4% to 14% for 10 to 3 features. Consistently, our proposed system maintained the lowest false prediction rates, ranging from 2% to 10% for 10 to 3 features. Table 5 reveals that our proposed system consistently outperforms other models by significantly reducing false prediction rates for heart disease. As the number of selected features from the dataset increases, the false prediction rates decrease, indicating enhanced system reliability.

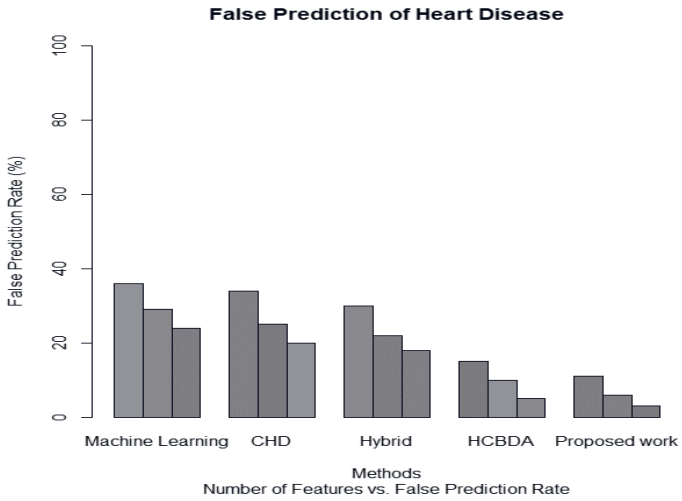

Figure 5. False prediction of heart disease

5. Conclusion

This study encompassed a comprehensive analysis of a proposed system, with a specific focus on routing performance and heart disease prediction. Utilizing a variety of methodologies and techniques, our research unveiled several noteworthy findings. Notably, our proposed routing system emerged as a frontrunner in comparison to existing methods. As the number of nodes increased, our system consistently outperformed alternative approaches, achieving routing performance percentages of 82%, 86%, and 93% for 30, 50, and 100 nodes, respectively. These results underscore the efficacy of our approach in seamlessly managing network traffic. The heart disease prediction model demonstrated remarkable accuracy, particularly as the number of features increased. When restricted to only 3 features, accuracy ranged from 66% to 94%, and with 10 features, it soared to an impressive 78% to 99%. This underscores the pivotal role of feature selection in ensuring precise heart disease prognosis. Our proposed system demonstrated an exceptional capability to minimize false predictions of heart disease. With an increase in the number of selected features from the dataset, the false prediction rate notably decreased. This highlights the reliability of our model in reducing incorrect diagnoses. The outcomes of our research emphasize the substantial potential of our proposed system, not only in optimizing routing performance within IoT networks but also in facilitating precise heart disease prediction. These findings underscore the critical importance of feature selection in enhancing the accuracy of medical diagnostic systems. Looking ahead, we anticipate that further refinements and real-world implementations of our proposed system will make significant contributions to improved network efficiency and healthcare outcomes. With the ever-evolving landscape of technology, our model holds promise as a valuable tool for addressing intricate healthcare challenges and enhancing IoT networks across various applications.

References

[1] Anandan, M., Manikandan, M., & Karthick, T. (2020). Advanced Indoor and Outdoor Navigation System for Blind People Using Raspberry-Pi. Journal of Internet Technology, 21(1), 183–195.

[2] Ananthajothi, K., & Subramaniam, M. (2019). Multi level incremental influence measure based classification of medical data for improved classification. Cluster Comput, 22, 15073–15080. https:// doi. org/10. 1007/ s10586- 018- 2498-z

[3] Ananthajothi.K & Subramaniam.M , 'Efficient Classification of Medical data and Disease Prediction using Multi Attribute Disease Probability Measure', Applied Mathematics & Information Sciences, ISSN 1935–0090, E.ISSN 2325–0399, Vol. 13, no. 5, pp. 783–789 (2019). https://doi:https:// doi. org/10. 18576/ amis/ 130511

[4] Archenaa, J., & Mary, A. (2017). "Health recommender system using big data analytics", J. Manage.Sci. Bus. Intell. Pp. 17–24.

[5] Banu, N. K. S., & Swamy, S. (2016). Prediction of heart disease at an early stage using data mining and big data analytics: A survey, International Conference on Electrical, Electronics, Communication,Computer and Optimization Techniques, IEEE, Mysuru, India.

[6] Bashir, S., Khan, F., Khan, Z., & Anjum, A. (2019). "Improving heart disease prediction using feature selection approaches", IEEE, pp. 619–623.

[7] Chen, M., Hao, Y., & Hwang, K. (2017). Disease prediction by machine learning over big data from healthcare communities. IEEE Access, 5, 8869–8879.

[8] Fahad, A., & Alshatri, N. (2014). A survey of clustering algorithms for big data taxonomy and empirical analysis. IEEE Transactions Emerging Topics in Computing, 2(3), 267–279.

[9] Ganesan, M., & Sivakumar, N. (2019). "IoT based heart disease prediction and diagnosis model for healthcare using machine learning models," 2019 IEEE International Conference on System, Computation, Automation, and Networking (ICSCAN), Pondicherry, India, pp. 1–5.

[10] Geetha, G., Safa, M., Fancy, C., & Saranya, D. (2018). "A hybrid approach using collaborative filtering and content based filtering for recommender system", Journal of Physics: Conference Series Vol 1000, Issue 1.

[11] Geetha, G., Safa, M., Fancy, C., Chittal, K. (2017). "3D face recognition using Hadoop",2017 International Conference on Energy, Communication, Data Analytics and Soft Computing,

[12] Geetha, G., Safa, M., Saranya, G., Subburaj, R. (2017). "An effective practices, strategies and technologies in the service industry to increase customer loyalty using map indicator" IEEE International Conference on IoT and its Applications, ICIOT 2017, 2017, 8073607

[13] Ismail, A., & Shehab, I. (2019). El-Henawy, "Healthcare analysis in smart big data analytics: Reviews, challenges, and recommendations", in security in smart cities: Models, applications, and challenges (pp. 27–45). Springer.

[14] Karthick, T., Amith Sai, A. V., Kavitha, P., Jothicharan, J., & Kirthiga Devi, T. (2020). Emotion detection and therapy system using chatbot. International journal of Advanced Trends in Computer Science and Engineering, 9(4), 5973–5978.

[15] Khatal S. S., & Sharma Y. K. (2020)." Analyzing the role of Heart Disease Prediction System using IoT and Machine Learning "International Journal of Advanced Science and Technology, Vol. 29, no. 9.

[16] McPadden, T., Durant, D., Bunch, A., Coppi, N., Price, K., Schulz, W. L., & Rodgerson, K. (2019), Health Care and Precision Medicine Research:

Analysis of a Scalable Data Science Platform. Journal of medical internet research, 21(4), e13043.

[17] Meenakshi, K., Maragatham, G. Lecture Notes on Data Engineering and Communications Technologies, 2020, 35:1076–1087

[18] Meenakshi, K., Sunder, R., Kumar, A., Sharma ,An intelligent smart tutor system based on emotion analysis and recommendation engine, N.IEEE International Conference on IoT and its Applications, ICIOT 2017, 2017, 8073608

[19] Minghuan, Fu Y. P., Cheng, B., Tao, X., Guo, J. (2020)." Enhanced Deep Learning Assisted Convolutional Neural Network for Heart Disease Prediction on the Internet of Medical Things Platform" IEEE Access,Page(s): 189503 – 189512.

[20] Prasad, S. T., Sangavi, S., Deepa, A., Sairabanu, F., & Ragasudha R. (2017). "Diabetic data analysis in big data with the predictive method," Int. Conf. Algorithms, Methodol. Model. Appl. Emerg. Technol., pp. 1–4, 2017.

[21] Safa, M., & Pandian, A. (2021). A review on big IoT data analytics for improving QoS-based performance in system: Design opportunities and challenges. Lecture Notes in Networks and Systems,130, 433–443.

[22] Sheeran, M., & Steele, R. (2017). "A framework for big data technology in health and healthcare," 2017 IEEE 8th Annu. Ubiquitous Comput. Electron. Mob. Commun. Conf., pp. 401–407.

[23] Tawalbeh, L. A., Mehmood, R., Benkhlifa, E., & Song, H. (2016). Mobile cloud computing model and big data analysis for healthcare applications. IEEE Access, 4, 6171–6180.