Internet of Things-driven Predictive Analytics for Heart Disease Detection

Shivakumar Kagi [1*], Ravisankar Dakupati[2] , K.G.S.Venkatesan[3] , Rama Krishna Pavuluri[4]

[1]Department of Computer Science and Engineering, Sharnbasva University, Kalaburagi, Karnataka, India

[2]Department of Electronics and Communication Engineering, SRM University-AP, Neerukonda, Andhra Pradesh, India

[3,4]Department of Computer Science & Engineering, MEGHA Institute of Engineering and Technology for Women, Hyderabad, Telangana, India

Received: 28 June 2024 • **Accepted:** 30 June 2024 • **Published Online:** 30 June 2024

Abstract: The prevalence of cardiovascular diseases (CVDs) continues to be a major public health concern worldwide, necessitating innovative approaches for early detection and management. This study proposes an Internet of Things (IoT) architecture augmented with machine learning algorithms for the prediction of cardiovascular disease risk. A comprehensive dataset is created for analysis using wearable sensors and devices connected to the Internet of Things (IoT). These devices collect physiological data in real-time, such as heart rate, blood pressure, and activity levels. Several machine learning models, including logistic regression, decision trees, and support vector machines, are trained on this dataset to predict the likelihood of cardiac disease happening. Several metrics are utilized to evaluate the efficacy of these models, including specificity, accuracy, precision, and area under the receiver operating characteristic curve (AUC-ROC). The results demonstrate that the methodology based on the Internet of Things (IoT) for predictive modeling effectively identifies persons at risk of developing heart disease. This paves the way for early intervention and tailored healthcare management approaches.

Key words: Internet of Things, Heart disease, Machine learning, Predictive modelling, Wearable devices,

*Correspondence: Department of Computer Science and Engineering, Sharnbasva University, Kalaburagi, Karnataka, India. Email:shivakumarkagi@gmail.com

https://doi.org/10.58599/IJSMCSE.2024.1110

Vol. 1, No. 1, June 2024, pp:39-48

Physiological data.

1. Introduction

Cardiovascular diseases continue a leading reason of humanity globally, highlighting the urgent need for effective preventive measures and early intervention strategies. The advent of Internet of Things (IoT) technology presents a novel opportunity to revolutionize healthcare by enabling continuous monitoring and analysis of physiological data. The development of predictive models combining Internet of Things (IoT) frameworks with machine learning techniques can lead to the early detection and treatment of heart illness. In this introductory article, we will examine how the IoT-enabled predictive modeling could enhance cardiovascular health results. Recent years have seen a meteoric rise in the use of wearable devices and sensor technologies, thanks to their seamless integration into daily life and capacity to provide continuous monitoring of vital signs. The use of these devices can shed light on a person's cardiovascular health because they capture a multitude of physiological data, including heart rate variability (HRV), blood pressure, and activity levels. Thanks to the IoT, this data can be sent to healthcare providers in real-time, enabling them to respond quickly and create personalized treatment plans for each patient. Traditional approaches to heart disease diagnosis often rely on periodic clinical assessments and subjective risk assessment tools, which may overlook subtle changes in physiological parameters and fail to detect early warning signs of CVDs. In contrast, IoT-enabled predictive modeling offers a proactive and data-driven approach to healthcare.By continuously monitoring an individual's physiological parameters over time, these models can detect deviations from baseline values and predict the likelihood of developing heart disease with high accuracy. The integration of IoT technology with machine learning algorithms holds promise not only for early detection but also for personalized risk stratification and targeted interventions. By analyzing predispositions, influences, predictive models can provide actionable insights into individual risk profiles and recommend tailored interventions to mitigate the progression of heart disease. Furthermore, the scalability and accessibility of IoT frameworks make it feasible to deploy these predictive models across diverse healthcare settings, from primary care clinics to remote monitoring systems, thereby democratizing access to preventive cardiovascular care. In summary, the convergence of IoT technology and machine learning algorithms offers a transformative approach to cardiovascular health management, enabling early detection, personalized risk assessment, and targeted interventions for individuals at risk of heart disease. By harnessing the power of real-time physiological data and advanced analytics, IoT-enabled predictive modelling. This paper aims to explore the applications, challenges, and opportunities of IoT-enabled predictive modeling in the context of heart disease diagnosis and management, paving the way for a future where preventive healthcare is proactive, personalized, and accessible to all. Cloud services in healthcare revolutionize data management by enabling secure, real-time access and sharing of patient information. They enhance collaboration among healthcare providers, leading to improved patient outcomes and streamlined operations. With advanced analytics and machine learning, cloud platforms support precise diagnostics and personalized treatment plans. These services offer scalability to accommodate growing data volumes without sacrificing performance. Additionally, robust security measures ensure compliance with regulations, safeguarding patient privacy and data integrity.

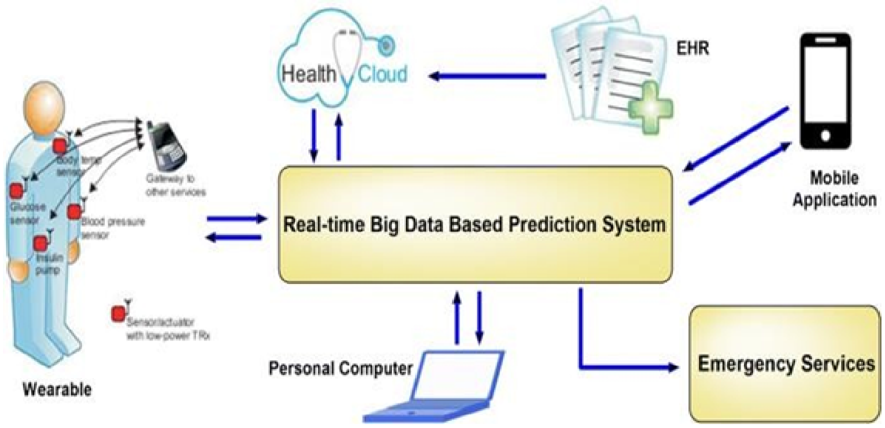

Figure 1. Cloud service in health care

The Internet of Things is helping to gather patient data through its sensors in order to assist in disease prediction. This is similar to how prior methods of data or symptom gathering helped in the process of disease prediction. The Internet of Things is a technology that enables various physical products and other things to collect also portion data with each other or through data center schemes. This is accomplished through the utilization of embedded electronics, smart clothes, software, sensors, and network connectivity. This network's primary function is to facilitate the sharing of information with other users. Because of the widespread availability of wearable health monitors, there has been a significant increase in the quantity and variety of data that is generated by these different devices. Through the use of a big data analytics system, it is necessary to conduct analysis in order to comprehend the patterns of user behavior and to gather significant data. By 2020, as much as forty percent of all Internet of Things devices will be connected to healthcare. By bridging the gap between the medical field and the information technology industry, medical informatics has the potential to transform healthcare by reducing costs, improving efficiency, and eventually saving lives. There are a variety of chronic disorders that can lead to death if they are not detected in a timely manner by real-time monitoring made possible by the Internet of Things. Diabetes and heart attacks are only two examples of these conditions. There are presently a number of open-source technologies that continually monitor health parameters continuously. There are a number of different data sources that are incorporated into the process that the proposed system uses, as shown in Figure 1.

2. Related Works

The Internet of Things is made up of smart devices that are limited in resources but are able to sense and interpret data [1]. The vast majority of intelligent sensing devices are connected to it. The technology that makes up the Internet of Things may be broken down into three distinct categories: technologies for monitoring, identifying, and communicating [2]. Identification technologies allow for the localization and identification of nodes within a healthcare network [3]. Through Zigbee, wifi, and RFID, the devices are able to establish wide-area communication between several organizations, and they are able to communicate within a local area network. In addition, hospital devices constantly track patients' whereabouts using the Global Positioning System (GPS) [4]. The collection of physiological information from the patient body, including temperature, pressure rate, and other similar data, is accomplished by the use of a sensor that is either embedded or worn on the human body [5]. Patients and healthcare professionals alike are able to deal with a variety of health conditions at a reduced cost because to this [6]. Monitoring of a patient's health using a wearable electrocardiogram for patients who have chronic diseases such as cardiovascular disease [7]. In order to store and retrieve data for analysis, discrete wavelet transforms (DWT) [8] are utilized. Additionally, an analog-based framework is constructed in order to categorize the data. Information for general purposes to the medical domain Due to the fact that a big quantity of data is collected from a wide variety of sources, the Internet of Things system places a significant emphasis on the storage and acceleration of data. In the event that the data hub is slow at that moment for whatever reason, a FOG computing [9] work that functions similarly to a cache memory will keep data that is temporarily used and data that is needed. The processing and caching of data in such a way that a large number of users can easily obtain data [10]. IoT devices are required to authenticate themselves in order to access the data stored in the cloud. This authentication is performed by edge services, which offer a lightweight authentication system [11]. Additionally, it enhances the speed of communication and processing among the devices, as well as the load balancing that is accomplished, and it also extends cloud computing [12].

An edge server in an IoT network receives data transmitted by devices for storage reasons. Load balancing, network optimization, and resource optimization are responsibilities of an SDN (Software Defined Network) controller, which connects the edge servers. The topic of software-defined networking (SDN), which allows for the control and management of next-generation data, is explored in depth [13]. Processing the data could result in data loss. Secure storage is a must for any system built on the back chain; only then can the entire system be guaranteed to be safe. Included in this category are not just data pertaining to humans, but also crucial signals between and within control systems, details about tasks, and outcomes of decisions. That way, the safety of the whole system was ensured. A study on any wearable devices that are utilized in the human body. The study focused on the fact that there are two different sorts of protocols that are utilized [14]. One of these protocols is known as CoAP, and it is responsible for establishing a link between many to many models. The MQTT protocol is another technology that decouples the sender and receiver of a message in both space and time. An asynchronous communications protocol serves as the basis of the system, which enables the two entities to be considered independent of one another. The person who is taking care of you will receive a notification if you use it in a wearable health tracker. Because of this, prompt action can be taken in the event that the device detects any suspicious behavior. This is a medical condition that has the potential to be fatal and needs to be addressed as soon as possible in order to prevent serious implications in the future [15]. It is possible that a machine learning (ML) model could be helpful in the early diagnosis of strokes, which would assist lessen the terrible repercussions that are associated with stroke sufferers. As part of their research, they investigate the degree to which various machine learning systems are able to accurately predict strokes by utilizing a variety of physiological risk indicators. A total of 96% accuracy was reached by the random forest classification approach, which was greater than any of the other methods that were investigated. Last but not least, when it comes to cross-validation measures for forecasting brain strokes, the random forest method fared better than other method approaches [16].

3. Methodology

Machine learning methods such as Naïve Bayes Multinomial (NBM), Random Forest (RF), Logistic Model Tree (LMT), and Decision Tree (DT) are utilized in order to retrieve health records from the Kaggle dataset where they are stored. These data pertain to parameters such as cardiovascular disease or attack, hypertension, dyslipidemia, cholesterol testing, body mass index (BMI), exercise, heavy alcohol use, overall health, mental health, physical health, types of walks, gender, age, education, income, smoking, produce, and any medical treatment for 25,000 patients. Additionally, these parameters include any medical treatment that may be administered. Through the utilization of these models, it is feasible to ascertain the patient's potential risk of experiencing a stroke. Sensors that are connected to the internet collect data on a wide variety of elements of people's health, including but not limited to people's heart rate, blood pressure, cholesterol, weight, exercise, alcohol consumption, mental and physical health, as well as the frequency and severity of heart attacks and other cardiovascular disorders. The remaining data, which includes subjects such as medical treatment, smoking status, produce, education level, and income, is obtained through the use of more conventional methods. The RF algorithm, developed by Leo Breiman and Adele Cutler, is one of the most popular machine learning tools in use today. This method generates a single answer by merging the output of many decision trees. Logic Tree for Logistics: When utilized in conjunction with a linked administered prepared calculation, the Logistic Model Tree (LMT) classification model can be displayed. A combination of Logistic Regression (LR) features and Decision Tree learning parameters achieves this goal. DT, or Decision Tree: For classification and regression issues, you can use the Decision Tree, a non-parametric supervised learning technique. This is a unified, hierarchical structure that uses a root node also other tree-like structures to arrange the issues. The proposed methodology flow chart is shown in Figure 2.

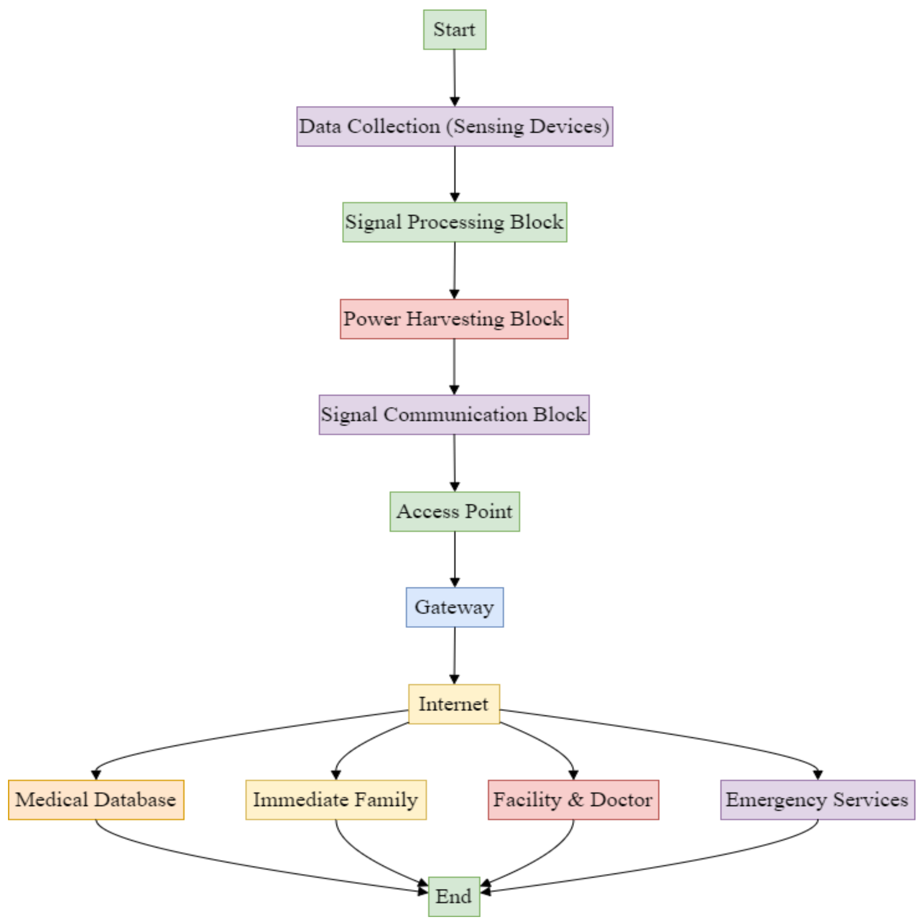

Figure 2. Proposed method flow chart

The flow chart illustrates the comprehensive step-by-step process of collecting, processing, transmitting, and utilizing health data in an IoT-enabled system for heart disease diagnosis. Initially, wearable devices equipped with sensors gather vital health metrics, including heart rate, blood pressure, and other critical physiological data. This data is then transmitted to various processing blocks, such as the Signal Processing Block, which filters, amplifies, and converts the raw data into a usable format. Concurrently, the Power Harvesting Block ensures that the sensors remain operational by generating the necessary power from environmental sources or the body itself. The processed data is then sent to the Signal Communication Block, which handles the wireless transmission of data to an access point. At the access point, the data is aggregated and forwarded to a gateway, which routes the information to the internet for further distribution. Through the internet, the data is transmitted to a centralized Medical Database for storage and analysis. This information is also made accessible in real-time to immediate family members and healthcare providers via mobile applications and personal computers, allowing for continuous monitoring and timely medical intervention. In case of emergencies, the system promptly alerts emergency services to ensure rapid response. This systematic approach not only facilitates early diagnosis and treatment of heart disease but also enhances patient care by leveraging the integration of IoT technology, big data analytics, and real-time communication.

The flow chart further underscores the importance of seamless data integration and accessibility across multiple platforms to optimize heart disease management. After the data reaches the Medical Database, sophisticated algorithms and big data analytics are employed to derive actionable insights, which are crucial for predictive modeling and early diagnosis of potential heart conditions. This processed and analyzed data is then disseminated to various stakeholders. For patients and their immediate families, mobile applications provide an intuitive interface for real-time health monitoring, offering notifications and visualizations that keep them informed about their health status. Healthcare providers, accessing the data via personal computers, can conduct detailed analysis, monitor patient trends over time, and make informed decisions regarding treatment plans. In critical situations, the system's capability to alert emergency services ensures that any significant health anomalies trigger an immediate response, potentially saving lives by providing timely medical intervention. This integrated, multi-platform approach highlights the efficiency and effectiveness of an IoT-enabled health monitoring system in managing and diagnosing heart disease, ultimately contributing to better patient outcomes and streamlined healthcare services.

4. Results and Discussion

The study investigates the effectiveness of implanting devices to gather continuous health data and applying various machine learning techniques to predict cardiac arrests, aiming to enhance human healthcare. By storing the collected data in the cloud and analyzing it using algorithms such as Naïve Bayes Multinomial (NBM), Random Forest (RF), Logistic Model Tree (LMT), and

Table 1. Comparison of different models

Dataset sizes	**NBM**	**RF**	**LMT**	**DT**
	Accuracy (%)			
1-10K	81.656	95.47	95.55	95.63
10K-20K	83.08	95.89	95.95	95.99
20K-30K	81.97	95.45	94.54	95.55
30K-40K	84.14	94.84	97.94	97.94
40K-50K	86.39	97.57	97.63	97.63

Decision Tree (DT), the goal is to improve the accuracy of disease prediction while minimizing error rates. The comparison of different models in Table 1 shows the accuracy rates of NBM, RF, LMT, and DT across varying dataset sizes (ranging from 1,000 to 50,000 records):

The NBM model demonstrates moderate accuracy, starting at 81.656% for the smallest dataset (1-10K records) and gradually improving to 86.39% for the largest dataset (40K-50K records). While NBM shows a consistent increase in accuracy as the dataset size grows, it generally lags behind the other models in terms of overall performance. RF exhibits high accuracy across all dataset sizes, starting at 95.47% for 1-10K records and reaching 97.57% for 40K-50K records. The consistent performance of RF highlights its robustness and effectiveness in handling various dataset sizes, making it one of the most reliable models in this study. LMT shows competitive accuracy, with a notable spike to 97.94% for the 30K-40K and 40K-50K datasets. This model demonstrates significant improvement with larger datasets, indicating its potential for high accuracy in large-scale health data analysis. DT consistently delivers high accuracy, starting at 95.63% for the smallest dataset and maintaining strong performance up to 97.63% for the largest dataset. The DT model's high accuracy and stability across different dataset sizes suggest its suitability for real-time health monitoring and predictive analysis. The comparative analysis shows that Decision Tree (DT) and Random Forest (RF) models generally outperform NBM and LMT in terms of accuracy, particularly as the dataset size increases. RF and DT models consistently achieve high accuracy rates, making them highly effective for predicting cardiac arrests and potentially other health conditions. The results indicate that employing advanced machine learning techniques, particularly RF and DT, can significantly enhance the prediction accuracy of heart disease diagnoses, thereby improving patient outcomes and enabling timely medical interventions. By leveraging these models, healthcare systems can better manage and predict health conditions, ultimately contributing to more efficient and effective healthcare delivery.

5. Conclusion

In this study of 25,000 individuals with heart disease, we evaluated four different models: Naïve Bayes Multinomial (NBM), Random Forest (RF), Logistic Model Tree (LMT), and Decision Tree (DT). The motivation for this research stemmed from patients' concerns about their condition and the need to determine which model was most effective at predicting the risk of stroke. Our findings indicate that the DT model performed the best among the models examined. This new prototype model will enable us to generate more accurate forecasts with fewer errors and improved performance, thereby enhancing our ability to accurately diagnose heart diseases. Additionally, a hybrid model that integrates features from multiple existing models is a promising direction. By leveraging the strengths of individual models, this hybrid approach has the potential to improve learning efficiency, aiding in both diagnosis and treatment.

References

[1] Shaira Tabassum, Md Ittahad Uz Zaman, Md Sabbir Ullah, Ashikur Rahaman, Samsun Nahar, and AKM Muzahidul Islam. The cardiac disease predictor: Iot and ml driven healthcare system. In *2019 4th International Conference on Electrical Information and Communication Technology (EICT)*, pages 1–6. IEEE, 2019.

[2] Makarand Shahade and Mangesh M Ghonge. A review on iot-driven technologies for heart disease diagnosis and prediction. *Leveraging AI Technologies for Preventing and Detecting Sudden Cardiac Arrest and Death*, pages 21–34, 2022.

[3] Sudeepta Mishra, VK Chaithanya Manam, et al. A comparative study of unsupervised learning techniques and natural language processing in network traffic classification. In *2023 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*, pages 138–143. IEEE, 2023.

[4] Asim Unmesh, Rahul Jain, Jingyu Shi, VK Chaithanya Manam, Hyung-Gun Chi, Subramanian Chidambaram, Alexander Quinn, and Karthik Ramani. Interacting objects: A dataset of object-object interactions for richer dynamic scene representations. *IEEE Robotics and Automation Letters*, 9(1): 451–458, 2023.

[5] A Angel Nancy, Dakshanamoorthy Ravindran, PM Durai Raj Vincent, Kathiravan Srinivasan, and Daniel Gutierrez Reina. Iot-cloud-based smart healthcare monitoring system for heart disease prediction via deep learning. *Electronics*, 11(15):2292, 2022.

[6] VK Chaithanya Manam and C Siva Ram Murthy. Performance modeling of routing in delay-tolerant networks with node heterogeneity. In *2012 Fourth International Conference on Communication Systems and Networks (COMSNETS 2012)*, pages 1–10. IEEE, 2012.

[7] M Ganesan and N Sivakumar. Iot based heart disease prediction and diagnosis model for healthcare using machine learning models. In *2019 IEEE international conference on system, computation, automation and networking (ICSCAN)*, pages 1–5. IEEE, 2019.

[8] Jane Smith and Patrick Evan. Smartheart advisor: an iot-driven approach to cardiovascular risk prediction. Technical report, EasyChair, 2024.

[9] Sanjeev Kulkarni, Anupama Sawihalli, R Ambika, and Lalita Naik. Mobile powered sub-group detection/formation using taste-based collaborative filtering technique. In *2017 Innovations in Power and Advanced Computing Technologies (i-PACT)*, pages 1–5. IEEE, 2017.

[10] VK Chaithanya Manam and C Siva Ram Murthy. Performance modeling of routing in delay-tolerant networks with node heterogeneity. In *2012 Fourth International Conference on Communication Systems and Networks (COMSNETS 2012)*, pages 1–10. IEEE, 2012.

[11] Gerard Deepak, M Madiajagan, Sanjeev Kulkarni, Ahmed Najat Ahmed, Anandbabu Gopatoti, and Veeraswamy Ammisetty. Mcsc-net: Covid-19 detection using deep-q-neural network classification with rfnn-based hybrid whale optimization. *Journal of X-Ray Science and Technology*, 31(3):483–509, 2023.

[12] Allam Balaram, TDNSS Sarveswara Rao, Lakshmana Phaneendra Maguluri, Shams Tabrez Siddiqui, Anandbabu Gopatoti, and Prathyusha Kuncha. Managing 5g iot network operations and safety using deep learning and attention methods. *Wireless Personal Communications*, pages 1–16, 2024.

[13] Kamlesh Mani, Kamlesh Kumar Singh, and Ratnesh Litoriya. Iot-driven revolution: Transforming elderly healthcare through dynamic heart health monitoring and analysis.

[14] V Suresh Kumar, Sanjeev Kulkarni, Naveen Mukkapati, Abhinav Singhal, Mohit Tiwari, and D Stalin David. Investigation on constraints and recommended context aware elicitation for iot runtime workflow. *International Journal of Intelligent Systems and Applications in Engineering*, 12(3s):96–105, 2024.

[15] Allam Balaram, TDNSS Sarveswara Rao, Pankaj Rangaree, Shams Tabrez Siddiqui, Anandbabu Gopatoti, and Lakshmana Phaneendra Maguluri. Energy–efficient distribution of resources in cyber-physical internet of things with 5g/6g communication framework. *Wireless Personal Communications*, pages 1–20, 2024.

[16] VP Sriram, D Kamalakkannnan, T Archana, Anandbabu Gopatoti, B Swapna, and Ajay Singh Yadav. Design of voice based virtual assistant using internet of things. In *2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*, pages 1152–1155. IEEE, 2022.