Early Prediction of Cardio Vascular Disease using Machine Learning and IoT Techniques : A Review

Mr.Michael Raj S,
Research Scholar, Department of Computer Science,Sri Krishna Adithya College of arts and Science,Coimbatore
Dr.K.BManikandan,
Associate Professor, Department of Computer Science,Sri Krishna Adithya College of arts and Science,Coimbatore

ABSTRACT

One application where data mining methods are showing promising results is in the detection of diseases. Heart disease and disorders (HD) are among the most complicated and deadly human diseases in the world. This condition causes the heart to operate improperly, resulting in blood channel blockages and angina, heart attacks, and stroke. In order to extract usable knowledge from the vast volumes of medical data that are readily available, strong data analysis techniques are required. The use of data-mining and statistical technologies to enhance data analysis on massive data sets has long been of interest to researchers. A thorough investigation of the use of a specific method of data mining in the detection of heart disease has revealed adequate levels of accuracy. Researchers have recently begun examining the impact of combining many techniques that have improved heart disease detection outcomes. Less attention has been paid to the use of techniques for data mining to find a viable treatment for people with heart disease. In this paper, various data mining techniques proposed by various researchers for identifying the heart failure at its early stage is analyzed and the research gap available in the existing systems are discussed.

Keywords : Heart failure, Machine Learning, IoT,Review,Data Mining

I. Introduction

A human heart is impacted daily by a number of variables. Most issues develop rapidly, and new heart disorders are discovered quickly. In today's pressure-filled environment, the health of the heart, an essential organ that pumps blood for circulating all through the body, is essential for a long and healthy life. Heart problems can also lead to heart failure, which can result in shortness of breath if the heart is unable to adequately pump blood.

Cardiovascular diseases are currently among the most prevalent diseases in the medical world. Heart disease is a leading cause of death worldwide, accounting for more than 17.7 million deaths per year. Out of them, around 7.4 million and 6.7 million persons worldwide per year died from coronary heart disease and stroke-related heart disease, respectively. The World Health Organization (WHO) reported that over twelve million fatalities globally occur each year as a result of heart disorders.That means one person dies from heart disease every 34 seconds. Congestive Heart Failure (CHF), Coronary Vascular Disease (CVD), Coronary Artery Disease (CAD) and Abnormal Heart Rhythms are the four most prevalent forms of heart conditions.

Medical diagnosis is crucial, and it must be carried out precisely and efficiently in cases of complexity. Clinical testing should be less expensive with the aid of suitable computer-based data and decision support.

Software that helps clinicians identifies early-stage cardiac disease based on pathological and clinical data has been created using machine learning algorithms and computer technology. For a patient has many diseases from the same category, their information may include duplicate and related symptoms and indications when making a biological diagnosis. It becomes exceedingly challenging for the doctors to make the proper diagnosis. The aforementioned prediction issue in a medical dataset with many inputs can be resolved using data mining using intelligent algorithms. Recently, complicated and challenging tasks have been performed using artificial neural networks.In order to uncover hidden dependencies and be able to exploit them for prediction, the neural network is trained using historical data. A widely used method for classification and prediction tasks is feed forward neural networks trained by back propagation. A flawless model to assess the risk to categorize people with heart disease can only be achieved with correct data, which is why clinicians and patients need trustworthy information about an individual's risk of getting heart disease. In fact, the ideal model would be able to forecast when the sickness will start to manifest itself.

II. Machine Learning Algorithm for predicting Heart Disease

To assist medical professionals in assessing the severity of heart disease, Pandey et al. [1] created a model for forecasting the disease. Heart disease is classified using the Cleveland Heart Disease dataset using the J38 decision tree, that is based on a number of clinical criteria. The results of the model show that fasting blood glucose levels represent the most significant risk factor for heart disease. A successful two-stage method for predicting heart disease was found by

Mienye et al. [2]. The enhanced sparse auto encoder (SAE), an unsupervised neural network that seeks out the best description of the training data, was first created by the researchers. They employed an artificial neural network (ANN) to forecast patient health based on the learning records.

The use of AI-assisted electrocardiography (ECG) in the detection of heart disease in high-risk areas, as well as the consequences for treatment choices in patients with coronary heart disease, were explored by Siontis et al. [3]. Anitha et al. [4] investigated the possibility of using learning vector quantization approaches to forecast heart disease. Her approach has an accuracy rate of 85.55%. She used 303 entries and 76 attributes from the University of California, Irvine (UCI) machine learning collection as the datasets for her investigation. Due to missing values, the data had to be pre-processed, yielding a sample of 302 records with just 14 heart disease features. The information is divided into two groups: model training costs 70% of the budget, while model testing costs 30% of the budget. Machine learning techniques were utilized by

Kumar et al. [5] to predict whether cardiovascular disease had already manifested itself. In comparison to other machine learning techniques, RFC provides the highest accuracy, at 85.71%, according to the proposed model.

A unique health information system, according to Linda et al. [6], may be beneficial for heart disease patients who are recommended to exercise. Clinicians must use mobile information systems to create exercise prescriptions for patients with a wide range of CVD risk factors based on historical data. The system made available to the patients is a simple, time-saving, and evidence-based manual. The best method for diagnosing human heart disease was compared by G. N. Ahmad et al. [7] using machine learning algorithms with and without sequential feature selection.

In order to identify mortality events in heart disease patients while they are receiving therapy, Ritu et al. [8] proposed a sequential feature selection strategy. LDA, K - nearest neighbors, RF, DT,SVM and GBC are just a few of the machine learning techniques used. To validate the outcomes of the SFS algorithm, other metrics such as the confusion matrix, precision,receiver operating characteristic curve, F1-score and recall rate are also created. The experimental results showed that the random forest classifier's accuracy was 86.67% while using the sequential feature selection technique.

In order to predict cardiac disease, Gao et al. [9] proposed a model combining ensemble techniques (boosting and bagging) with algorithms for feature extraction (LDA and PCA). On a subset of characteristics from the Cleveland heart disease dataset, the authors contrasted ensemble approaches (bagging and boosting) with classifiers (SVM, KNN, RF, NB, and DT). The bagging ensemble learning approach using DT and PCA feature extraction achieved the most impressive performance, according to the experiment's results.

In order to predict cardiac failure, KarenG'arate-Escamila et al. [10] created a hybrid dimensionality reduction technique called CHI-PCA, which combines Chi-square and principal component analysis. Three distinct datasets from the UCI Machine Learning Repository were used in their study: the Hungarian, Cleveland, and Hungarian-Cleveland datasets. Five distinct classifiers—random forests, gradient-boosted trees, decision trees, multilayer perceptrons, and logistic regression—were used to assess the performance of the proposed method. The Cleveland dataset, the Hungarian dataset, and the Cleveland-Hungarian (CH) dataset each had accuracy values of 98.7%, 99.0%, and 99.4% for chi-square and principal component analysis (CHI-PCA) using random forests (RF).

Spence et al. [11] performed experiments on four commonly used heart disease datasets using four different feature selection techniques: principal component analysis, chi-square test, ReliefF, and symmetric uncertainty. As noted by the authors, the benefits of feature selection differ depending on the machine learning approach used for the cardiac datasets. For example, one of the most accurate models discovered had 85.0% accuracy, 84.73% precision, and 85.56% recall when chi-square feature selection was combined with the BayesNet classifier.

A diagnostic model for the early diagnosis of chronic renal disease was built by Semen et al. [12] using a dataset of 400 patients and 24 characteristics. The greatest characteristics were chosen using recursive feature elimination (RFE). The categorization techniques employed in this paper were decision trees, random forests, support vector machines, and k-nearest neighbors (KNN). Excellent performance was attained by all categorization techniques. All parameters for accuracy, precision, recall and F1-score were 100% for the random forest approach, which outperformed all other methods.

Almansour et al. [13] compared two classifiers, ANN and SVM and utilized a random exhaustive search strategy to improve their parameters in order to aid in the early diagnosis of chronic kidney disease. The UCI repository's 400-instance dataset is preprocessed, and features are selected based on correlation coefficients. The effectiveness of the classifiers is evaluated in comparison to the model's training duration and the best features (2, 3, 6, and all). Finally, since ANN outperforms SVM with high accuracy, the 12 best features are employed to predict renal disease using SVM and ANN.

KNN and Random Forest machine learning algorithms were used by Apurv Garg et al. [14] to forecast heart disorders. The data's balance was evaluated after collection and analysis, and a correlation between different features and their impact on the goal value was discovered. The UCI dataset from Kaggle was the one that was retrieved. For training and testing, it was split 80/20, respectively. The desired attribute was found to be positively correlated with Maximum heart rate reached and Chest Pain . Using KNN and Random Forest, this model had an accuracy of 86.89% and 81.967%, respectively.

A web application-based predictive model with a 75–25 split between the training and testing portions of the UCI dataset was proposed by Rishabh Magar et al. [15]. Predictive models based on Logistic Regression were found to be the best accurate, with an accuracy rate of 82.89%, followed by SVM at 81.57%, Naive Bayes, and Decision Tree at 80.43% each. The end user can utilize the online application as a preliminary test to assess their cardiac condition and, if necessary, seek medical guidance.

Four classification algorithms, including Random Forest, Decision Tree, Logistic Regression, and Naive Bayes, are employed in the system described by Apurb Rajdhan et al. [16] to forecast the patient's health. Data are divided into training data and testing data, respectively. It was decided to build a confusion matrix that would show true and false negatives in addition to true and false positives. The highest accuracy achieved with Random Forest classification was 90.16%.

A framework of models using supervised learning techniques through the WEKA tool was proposed by Devansh Shah et al. [17]. The likelihood of getting a cardiac disease was predicted using four different classification techniques: NB, KNN, RF, and DT. The dataset was first cleaned, then smoothed, normalized, and aggregated before being integrated and reduced. KNN approach produced the highest level of accuracy.

A system with an 87.5% accuracy was created by Harshit Jindal et al. [18] using three different classification algorithms: KNN, RF, and LR. In this EHDPS, or effective heart disease prediction system, KNN and Logistic Regression perform better than RF, with KNN offering the greatest accuracy of the three algorithms utilized (88.52%).

On the UCI dataset, AravindAkella et al. [19] applied 6 predictive models, and the Neural Networks model had the highest accuracy (93.03%) and recall (93.8%), both of which indicate low probability of false negatives and, consequently, extremely precise results. The other five models had accuracy of almost 80% and higher.

Five techniques were applied by Ravindhar NV et al. [20] back propagation neural network, fuzzy kNN, naive bayes, logistic regression, and k-means clustering. The experimental investigation of cardiac diseases employs a 10-fold cross validation procedure. Back propagation neural networks were used to gather data with 98.2% accuracy, 87.64% recall, and 89.65% precision.

III. IoT and Machine Learning in Heart Failure Prediction

A wireless gadget created by Kovuru Chandu Chowdary [21] is used to compute blood pressure, temperature, oxygen saturation, and pulse rate. Only the concerned doctor will be granted access thanks to the fingerprint sensor attached to the microcontroller. Data is obtained from a Raspberry Pi, which serves as a sensor node of the given device, and is then sent to a web server for visualization. The real-time physiological parameters are adjusted every 20 seconds. The Raspberry Pi also has GSM connections and Zigbee modules. An alert SMS will be sent to a mobile doctor once the health data hits certain threshold values.

Sood and Mahajan[22] created the healthcare system for the diagnosis and treatment of the chikungunya virus. The information was gathered by sensors in wearable IoT monitors such fitness sensors, opioid sensors, location sensors, ambient and meteorological sensors. The collected data is sent to the fog in real-time for analysis and diagnosis of CHV users who may be compromised, which results in alerts being sent to the patient's cell phones. The cloud-based processing of each user's results and collected medical records determines each user's ORI, which represents their likelihood of contracting and spreading the virus.

A. Alani [23] has created an IoT-based smart healthcare system that employs the Intel Galileo board to track temperature, heart rate, and blood pressure. Its pins are compatible with the Arduino Uno and contains a 32-bit CPU in addition to an Ethernet shield. The Intel Galileo computer is used to create code for the Arduino IDE. The board processes and uploads the sensor data via the Xampp-based storage server. Using a login ID and password, a doctor visualizes the data from a database server.

For older persons and chronic patients, AshwiniGutta [24] has created an IoT-based health management system. A device can be an IoT server, a PC, or a Raspberry Pi. The Raspberry Pi receives the data from the temperature, ECG, and pulse rate sensors, which are subsequently entered into the database server. Correspondence is conducted using the Message Queuing Telemetry Transport (MQTT) protocol. Once three emails or SMS notifications have been sent, the patient's condition is considered critical, and the doctor is notified via email or SMS. A doctor can remotely access clinical records to evaluate the patient's current state of health.

Table 1 gives the comparison of some of the efficient algorithms proposed by authors

Sl.No	Author	Method	Merits	Demerits
1	Domor I Mienye (2020) [2]	Sparse auto encoder Artificial Neural Networks	Improved performance	High Sensitivity Time Constraint
2	Konstantinos C Siontis (2021) [3]	Deep-learning AI-ECG	powerful tool for phenotyping of cardiac health	Insufficient parameters Need optimization Complex to implement
3	Anitha et al.(2019)[4]	Vector Quantization approach	Attains good accuracy	Less training data
4	N. Komal Kumar et al. (2020) [5]	Random woodland Machine learning classifier	Improved accuracy Reduced execution time	Insufficient training data
5	Linda S. Pescatell (2021) [6]	prioritize personalize prescribe exercise [P3-EX]	easy-to-use, time-efficient and Guided evidence-based approach	Need evaluation for its feasibility and acceptability
6	Vineel A. Chakradhar (2018) [25]	Identifying similarity in physiological waveform time-series data.	Uses optimal choice of parameters	Doesn't support multivariate functions
7	Qiang Huang et al (2018) [26]	Asymmetric LSH scheme based on Homocentric Hypersphere partition (H2-ALSH)	Asymmetric Locality-Sensitive Hashing Scheme	Not applied to hear disease dataset
8	YongliangQiao et al (2019) [27]	ConvNet features and localized image sequence matching	Good performance in the case of visualization	Not applied to heart disease prediction.
9	SahilGarg et al (2019)[28]	Dialog Modeling via KernelizedHashcode Representations	Computational efficiency and enhanced response time	Not applied to hear disease dataset
10	Anna Karen Gárate Escamilla et.al (2020) [11]	CHI - PCA with Random Forest	Improved accuracy	Same type of features are analyzed. Less features are considered.

IV. Conclusion and Future work

This overview of the literature shows how heart failure can be detected earlier using various sensors and techniques including machine learning, deep learning, and artificial intelligence on various datasets. Each piece of writing possesses strength and a weakness. Despite the researcher's efforts, there is still some uncertainty regarding prediction model standardization. Other numerous heart disease datasets from various sources with more attributes should be taken into consideration in order to obtain a more generalized classification and prediction accuracy. The main objective of our future research is to develop an effective predictive framework model that completely eliminates the majority of the flaws mentioned in this paper. In order to standardize the working learning approach and assure its dependability with the data, real-time data should indeed be examined on clinical validation and correlation

References

1. Atul Kumar Pandey & Prabhat Pandey & K L Jaiswal, 2013. "**A Heart Disease Prediction Model Using Decision Tree**," The IUP Journal of Computer Sciences, IUP Publications, vol. 0(3), pages 43-48, July.
2. Mienye, I.D.; Sun, Y.;Wang, Z. Improved sparse auto encoder based artificial neural network approach for prediction of heartdisease. Inform. Med. Unlocked **2020**, 18, 100307.
3. Siontis, K.C.; Noseworthy, P.A.; Attia, Z.I.; Paul, A. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. Nat. Rev. Cardiol. **2021**, 18, 465–478.
4. Anitha, S.; Sridevi, N. Heart disease prediction using data mining techniques. J. Anal. Comput. **2019**, 8, 48–55.
5. Kumar, N.K.; Sindhu, G.S.; Prashanthi, D.K.; Sulthana, A.S. Analysis and prediction of cardio vascular disease using MLclassifiers. In Proceedings of the 2020 6th International Conference on Advanced Computing and Communication Systems(ICACCS), Coimbatore, India, 6–7 March 2020; pp. 15–21.
6. Linda, P.S.; Yin, W.; Gregory, P.A.; Amanda, Z.; Margaux, G. Development of a novel clinical decision support system for exercise prescription among patients with multiple cardiovascular disease risk factors. MayoClin. Proc. Innov. Qual. Outcomes **2021**, 5, 193–203.
7. Ahmad, G.N.; Ullah, S.; Algethami, A.; Fatima, H.; Akhter, S.M.H. Comparative Study of Optimum Medical Diagnosis of HumanHeart Disease Using ML Technique with and without Sequential Feature Selection. IEEE Access **2022**, 10, 23808–23828.
8. R. Aggrawal and S. Pal, "Sequential feature selection and machine learning algorithm-based patient's death events prediction and diagnosis in heart disease," *SN Computer Science*, vol. 1, no. 6, 2020.
9. X.-Y. Gao, A. A. Ali, H. S. Hassan, and E. M. Anwar, "Improving the accuracy for analyzing heart diseases prediction based on the ensemble method," *Complexity*, vol. 2021, Article ID 6663455, 10 pages, 2021.
10. A. KarenG´arate-Escamila, A. E. Hassani, and E. Andr`es, "Classification models for heart disease prediction using feature selection and PCA," *Informatics in Medicine Unlocked*, vol. 19, Article ID 100330, 2020
11. R. Spencer, F. (abtah, N. Abdelhamid, and M. (ompson, "Exploring feature selection and classification methods forpredicting heart disease," *Digital Health*, vol. 6, Article ID 2055207620914777, 2020.
12. E. M. Senan, M. H. Al-Adhaileh, F. W. Alsaade et al., "Diagnosis of chronic kidney disease using effective classification algorithms and recursive feature elimination techniques," *Journal of Healthcare Engineering*, vol. 2021, Article ID 1004767, 10 pages, 2021.
13. N. A. Almansour, H. F. Syed, N. R. Khayat et al., "Neural network and support vector machine for the prediction of chronic kidney disease: a comparative study," *Computers in Biology and Medicine*, vol. 109, pp. 101–111, 2019.

14. Garg, Apurv& Sharma, Bhartendu& Khan, Rizwan. (2021). Heart disease prediction using machine learning techniques. IOP Conference Series: Materials Science and Engineering. 1022. 012046. 10.1088/1757- 899X/1022/1/012046.
15. Suraj Raut, Rishabh Magar, Rohan Memane. Prof. V. S. Rupnar "HEART DISEASE PREDICTION USING MACHINE LEARNING", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.7, Issue 6, page no.2081-2085, June-2020, Available :http://www.jetir.org/papers/JETIR2006301.pdf
16. ApurbRajdhan , Avi Agarwal , Milan Sai , Dundigalla Ravi, Dr. PoonamGhuli, 2020, Heart Disease Prediction using Machine Learning, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 04 (April 2020)
17. Shah, D., Patel, S. &Bharti, S.K. Heart Disease Prediction using Machine Learning Techniques. SN COMPUT. SCI. 1, 345 (2020). https://doi.org/10.1007/s42979-020-00365-
18. Harshit Jindal et al 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1022 012072
19. Akella, Aravind and Akella, Sudheer. Machine learning algorithms for predicting coronary artery disease: efforts toward an open source solution. Future Science OA Volume 7, Number 6, Pages FSO698, 2021, https://doi.org/10.2144/fsoa-2020-0206
20. Ravindhar NV, Anand, HariharanShanmugasundaram, Ragavendran, Godfrey Winster. Intelligent Diagnosis of Cardiac Disease Prediction using Machine Learning. Volume-8 Issue-11, September 2019, ISSN: 2278-3075 (Online). Page No: 1417-1421. DOI: 10.35940/ijitee.J9765.0981119
21. Chowdary, K. C., Lokesh Krishna, K., Prasad, K. L. &Thejesh, K. An efficient wireless health monitoring system. *Proc. Int. Conf. I-SMAC (IoT Soc. Mobile, Anal. Cloud), I-SMAC 2018* 373–377 (2019) doi:10.1109/I-SMAC.2018.8653716.
22. Sood, S. K. &Mahajan, I. Wearable IoT sensor based healthcare system for identifying and controlling chikungunya virus. *Comput. Ind.* **91**, 33–44 (2017).
23. A. Alani, IoT Based Smart Healthcare System. Int. Journal of Engg Research and Development, Vol 14, Iss 1 (2018).
24. Ashwini Gutte, RamkrishnaVadali, IoT based Health Monitoring System using Raspberry Pi, IEEE Int. Conf on Computing Communication Control and Automation (ICCUBEA) (2018).
25. Vineel A. Chakradhar, Evaluating Parameter Optimization in Locality-Sensitive Hashing for High-Dimensional Physiological Waveforms, Thesis - MIT,UK 2018
26. Qiang Huang, Guihong Ma, JianlinFeng, Qiong Fang, Anthony K. H. Tung, Accurate and Fast Asymmetric Locality-Sensitive Hashing Scheme for Maximum Inner Product Search**,** KDD '18 Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 1561-1570, 2018
27. YongliangQiao, Cindy Cappelle , YassineRuichek and Tao Yang, ConvNet and LSH-Based Visual Localization Using Localized Sequence Matching, Sensors 2019, 19, 2439; doi:10.3390/s19112439
28. Sahil Garg1 , Guillermo Cecchi2 , Irina Rish2 , Shuyang Gao1 , Greg Ver Steeg1 , Sarik Ghazarian1 , Palash Goyal1 , Aram Galstyan, http://ceur-ws.org/Vol-2202/paper4.pdf , 2019