Analytical review on Heart Disease Detection using Machine Learning Algorithms

M. SIREESHA[#],M.JAMES STEPHEN[*],P.V.G.D.PRASAD REDDY[+]

[#]*Department of IT&CA, Andhra University, Visakhapatnam, AP, INDIA*

[*]*Dr.B.R. Ambedkhar Chair Professor, Andhra University, Visakhapatnam, AP, INDIA*

[+]*Sr.Professor, Department of CS&SE, Andhra University, Visakhapatnam, AP, INDIA*

Email address of corresponding author: mallasireesha.rs@andhrauniversity.edu.in

Abstract: Heart disease is a term used to describe any disorder that affects the cardiovascular system. Heart disease has long been considered one of the leading causes of death. Most patients don't notice their signs until it's too late, and some people work hard to lessen the effects of risk factors that causes the heart disease. Cardiovascular disease highlights the need of getting a clear and early diagnosis. Because they can be fatal, researchers are focusing on creating smart systems to accurately identify cardiac aliments using electrocardiograms (ECGs), medical imaging, patient demographics and machine learning algorithms. In the field of healthcare, machine learning has become a powerful tool that offers potential solutions for early cardiac disease prediction. The main goal of this review paper is to provide a comprehensive analysis and current trends for investigation of various machine learning techniques employed for the early detection of heart disease. These algorithms should also address the class imbalance issue prevalent in heart disease datasets, provide interpretable insights to aid clinical decision-making, and ensure generalizability and scalability for real-world application. Developing innovative machine learning techniques that can tackle these challenges holds immense potential for enhancing early detection and prevention strategies for heart diseases, ultimately leading to improved patient care and reduced healthcare burden.

Keywords: Machine Learning (ML) Algorithms, Prediction Model, Cardiovascular Diseases (CVDs), Electrocardiograms(ECGs), Medical imaging.

Introduction

One of the major components of the human body is the Heart. The Heart's main function is to pump blood throughout the entire body. Heart Disease refers to any condition that may impair the heart's functioning. Cardiovascular disease is another name for it, which refers to both heart and blood vessel disease. In recent years, Heart Disease has become a major issue due to unhealthy diet, inactive lifestyle, tobacco usage and extreme consumption of alcohol. Despite numerous healthcare industries working to identify cardiac disease at an early stage, many people are unaware of the early detection of heart disease. As a result, it becomes nearly hard to pay for the cost of treatment. The early detection of cardiac disease is currently of great importance to researchers. Heart disorders are mostly caused by two categories of risk factors. Family history, age, and gender are examples of uncontrollable elements that fall under this category. The causes of heart disease that are under your control fall under a different group.

Risk factors like smoking patterns are under our control.

Fig: Identification of Heart Disease Risk Factors

Every day, the healthcare facility generates and collects a sizable amount of data. Data innovation makes it possible to extract data automatically, leading to novel findings. The diagnosis of cardiac disease should be made as soon as feasible. To accomplish this, a variety of machine learning and visual representation techniques have been created. The primary cause of heart disease may also be examined using machine learning. Early detection and diagnosis of cardiac disease are essential for bettering patient outcomes, lowering mortality rates and managing healthcare resources efficiently. Machine learning algorithms have showed tremendous promise in transforming the area of cardiac disease prediction, especially with the introduction of highly developed computational tools and the spread of electronic health information. Early interventions and individualized treatment strategies can be developed to slow the spread of the illness by identifying high-risk people before the appearance of clinical signs. Researchers and medical professionals are very interested in machine learning algorithms because of their capacity to evaluate enormous volumes of patient data, recognize complex patterns, and make precise predictions.

The following is a summary of this review's major contributions:

- To provide in-depth analysis of the most recent machine learning techniques used for prediction of Heart disease which focuses particularly on early-stage detection.
- Analyze experimental findings using techniques from published research literature.
- Compare the performance of machine learning algorithms with traditional diagnostic methods used for heart disease detection.
- Describe current research concerns around various elements of healthcare and provide some prospective future research possibilities.

Literature Review

Chaimaa Boukhatem et al. [1] implemented four classification algorithms, Multilayer Perceptron (MLP), Support Vector Machine (SVM), Random Forest (RF), and Nave Bayes (NB) to build the prediction models. The SVM algorithm with a linear kernel produced the best results, with an accuracy of 91.67%. The algorithms employed were able to extract the complex relationships between the disease's symptoms and its manifestations.

Dara DVVNS Saikumar et al. [2] implemented two algorithms KNN and SVM classifier to predict heart disease. It can be shown that the Support Vector Machine Algorithm only has a precision of 81.6%, whereas the K-Closest Neighbour Algorithm has a preferred exactness of 87.1%.

Seema Gulati et al. [3] applied six models on the Cleveland Heart Disease data set and achieved a maximum accuracy of 86.81% with Naïve Bayes over Multi-Layer Perceptron with 81.32% accuracy, K-Star with 84.62%accuracy, Logistic Regression with 82.42% accuracy, Decision Tree with 82.42% accuracy and J48 with 79.12% accuracy.

Priyanka Gupta et al. [4] utilized five machine learning classifiers to predict Heart disease or CVD. They also proposed a web application based predictive model which can take data like age, gender, blood pressure heart rate and other required attributes from the user and predicts the possibility of presence of heart disease. When they compared different classifiers, they found

that Logistic Regression performed the best, with a prediction accuracy of 88.5%.

Harshit Jindal et al. [5] developed a cardiovascular disease detection model using three ML classification modelling techniques KNN, Random Forest, Logistic Regression. In this EHDPS, or effective heart disease prediction system, Logistic Regression and KNN perform better than Random Forest with KNN offering the greatest accuracy of the three algorithms utilized 88.52%.

Sanjay Dhanka et al. [6] proposed Random Forest model trained on Cleveland Heart Disease dataset. The dataset is splitted in 80:20 ratio, 80% for training and 20% for testing. The 10-Fold cross validation was used to validate the Random Forest model, which achieved 90.16% accuracy.

Md. Touhidul Islam et al. [7] utilized PCA and Hybrid genetic algorithm with K-means for final clustering. For improved combinatorial optimization, they merged the unsupervised heuristic k-means method with metaheuristic genetic algorithms. The results show that these data mining algorithms have a 94.06% accuracy rate for early heart disease prediction.

Ravindhar NV et al. [8] applied five algorithms: Logistic Regression, Naive Bayes, Fuzzy KNN, K-Means Clustering and back propagation Neural-Network. 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.

Awais Nimat et al. [9] proposed an expert system based on two support vector machines(SVM) to predict Heart Diseases efficiently. The first one of these two SVMs is used to remove the extraneous features, and the second one is utilized for prediction. Additionally, they optimized the two techniques using the HGSA (hybrid grid search algorithm).they have a 92.22% accuracy rate.

Md. Razu Ahmed et al. [10] implemented a real-time based 4-tier cloud architecture for Heart Disease prediction and tracking record. For the early diagnosis of cardiac disease, they used five well-known supervised based machine learning techniques: ANN, SVM, Naive Bayes, Random forest, and Decision tree. Using a confusion matrix and the ten-fold cross-validation approach, they compared the performance of five classifiers that are used to predict heart disease. ANN attained the best accuracy rate of 84%.

C. B. Erdas et al. [11] utilized three algorithms 1Rule, Random Forest and SVM to detect survival of heart failure patients. The dataset's accuracy for the 1Rule algorithm was 86%, whereas it was 84% for the RF and SVM algorithms.

S. M Mehedi Zaman et al. [12] concentrated on the survival prediction of patients with heart failure using their follow-up information. To improve the performance , they created and put out a proposal for an ensemble machine learning method that makes use of three basic learners (DT, RF, & XGB) and one meta learner (RF). The performance improved to 99.98% accuracy, 100% precision, 99.98% recall, and 99.98% F1 score with an AUC score of 0.99 after merging numerous algorithms.

A. Ishaq et al. [13] implemented Extra Trees Classifier(ETC) algorithm and they significantly improve the prediction of heart failure patients survival with an accuracy of 92.62%.

Shalet K.S et al. [14] proposed the Heart Disease Prediction System (HDPS), which might predict a person's heart condition. Decision Tree (DT) and Support Vector Machine (SVM) Classifier are two examples of the methods employed. The disadvantage is that revealing the resting electrocardiographic output and sex which is not necessary to improve overall outcomes performance. The advantage of this approach is that it can classify various parameters using the classifier, including chest pain, age, type, thalassemia, colored fluoroscopy, and exercise-induced chest discomfort. The accuracy of this research is 82.35%, and the SVM classifier is

used for classification in addition to the DT method for feature selection.

Anshika Trisal et al. [15] employed K-Nearest Neighbor, Decision Tree, and Support Vector Machine as three machine learning methods, and some results have been analyzed in terms of accuracy, confusion matrix, precision, and recall. The Decision Tree achieved the best accuracy score of 98% . The Decision Tree yields the best results, with precision values of 97% for target value "0" and 100% for target value "1". In terms of recall, Support Vector Machine performs best for target values of "0" (61% recall value) and "1" (76% recall value). The decision tree produces the best results in the confusion matrix, with true positive values of 100, false positive values of 3, false positive values of 0, and true negative values of 102.

Rohit Bharti et al. [16] On the UCI heart disease dataset, the author of this research applied machine learning and deep learning techniques. Techniques used in this paper include logistic regression, K neighbors, SVM, Radom Forest, decision tree, and DL. The DL approach yields the best result with accuracy of 94.2% after applying all the techniques to the dataset.

M. Snehith Raja et al. [17] used 13 health parameters in this research. They gathered these factors from a dataset in order to estimate the likelihood of heart failure. In order to anticipate the outcome, they divided the data set into two sections: the training dataset and the testing dataset. For that purpose, they trained the model appropriately. Random Forest, Decision Trees, and ID3 were three distinct algorithms that were put to the test on the data sets. They compared the level of accuracy of these three algorithms. With accuracy rates of 99.67%, 99.81%, and 100%, the results showed that Random Forest fared the best across all three datasets.

Akanksha Kumari et al. [18] considered an individual analysis of the three algorithms LR, LDA, and KNN provide a superior mean value when compared to the seven algorithms (i.e., LR, LDA, CART, KNN, NB, SVM, and RF). The AdaBoost Algorithm can increase the minimum value of the Decision Tree classifier. The Naive Bayes and decision tree classifier combination's mean value is enhanced by the Voting Ensemble technique. The LR, CART, GNB, and RF algorithms are proven to have 100% accuracy.

Pranav Motarwar et al. [19] predicted possibilities of heart diseases more precisely. The accuracy was increased and boosted by the usage of techniques like Gaussian NB, SVM, Random Forest, Hoeffding Tree, and LMT for effective prediction. On the Cleveland dataset, each algorithm's performance was examined, and the results were compared in terms of accuracy. Furthermore, it is discovered that Random forest is more appropriate due to its capacity to take into account unique interests 95.08%.

Kuldeep Vayadande et al. [20] utilized Logistic Regression, NB, K-NN, SVM, Multi-Layer Perceptrons , Artificial Neural Networks, Decision Tree, Random Forest, XG Boost, and Cat Boost in the model. According to the research and the algorithms they utilized, the Logistic Regression, Random Forest, and XGBoost algorithms have produced results with higher accuracy compared to other techniques.. The accuracy obtained by Logistic Regression, Random Forest, and XGBoost are 88.52%, 88.52%, and 88.52 %, respectively.

Dakun Lai et al. [21] gathered information from the MIT-BIH Sudden Cardiac Death Holter (SDDB) database of SCD risk factors, the Normal Sinus Rhythm (NSRDB) database of healthy persons, and the AHA database of heart attack risk variables for this investigation. For this research (RF), five distinct machine learning techniques and algorithms were used: 1. K-Nearest Neighbour (KNN), 2. Decision Tree (DT), 3. Support Vector Machine (SVM), 4. Naive Bayes (NB), and 5. Random Forest. With a 99.49% accuracy percentage, Random Forest performed better than all of these algorithms.

K. Rajalakshmi et al. [22] utilized a tool Hadoop open-source framework's MapReduce , which

makes use of the Hive database. The algorithms employed are Prediction Tree C5.0, WAC (Weighted Associative Classifier), and K-Means. This method's disadvantage is that cardiologist advise on first aid is lacking. The advantage of this approach is that WAC provides a powerful medium for ingraining the classification process through Association Rule Discovery, which helps to increase classification prediction accuracy. The combined DT, WAC, and K-Means approach provides innovative prediction results for Heart Disease symptoms. The proposed methods provide a precise and reliable method of heart disease prediction.

Pabitra K Bhunia et al. [23] utilized heart.csv dataset. The methods that were used on the dataset are decision trees, support vector machines, Radom Forest classifiers, K closest neighbors, and logistic regression. The implementation's results indicate that the RFC and SVM algorithms produce better outcomes, with accuracy rates of 90.32% for RFC and 90.32% for SVM, respectively.

Mihir J. Gaikwad et al. [24] applied five machine learning algorithms Support Vector Machine
(SVM), Random Forest(RF), Gradient Boosting, Logistic Regression(LR), and Decision Tree Classifier. As a result, using the statlog(heart) dataset's characteristics along with the Support Vector Machine technique, heart disease may be predicted with almost perfect accuracy. They also developed a web application that uses the Support Vector Machine algorithm to determine whether a person has a cardiac disease or not.

Arun. R et al. [25] implemented Naïve Bayes and AES algorithms. The drawbacks of this approach include permitting restricted information exchange with the patient's family through the use of social networking tools. The AES encryption technique tends to be advantageous for this strategy since it encrypts sensitive yet unclassified information. By utilizing the recently developed technology of key private proxy re-encryption, CAM's cloud-assisted privacy-preserving mobile health monitoring system helps to shift the computational burden to the cloud.

Analytical Review

According to the research, Technologies based on machine learning are essential for the development of the medical field. The next generation of the predictive healthcare systems, which might serve thousands of patients, could be enabled with the help of machine learning. The organization of hypothesis-driven and clinical exploratory research from multiple sources of information might be guided by machine learning education.

With certain data sets [6], [4], [10], [16], [17], [18], [19], [25] machine learning techniques and algorithms are effective. On some datasets, these approaches showed reasonable precision and accuracy. The research's findings are not tested for new datasets if they only apply to a small number of difficult qualities and datasets.

The methods [4], [6], [10], [17], [19], [20], [22] are regarded as traditional learning-based methods for classification and prediction. These models and methods provide a significant level of accuracy when they are applied to the data sets. When a large dataset is analyzed, these models perform slightly worse.

The techniques [17], [20] performed well in terms of precision success metrics and accuracy resilience. When individual attributes are taken into account, the research findings are optimistic, but the findings are not encouraging when all of the attributes are taken into account at once.

A full analysis of the application of strengthening methods together is provided in [13], [15]. The approaches are based on ensemble machine learning algorithm using base learners Decision Tree and Random Forest. The proposed method outperformed a typical conventional model in terms of performance, and it is best suited for boosting the efficiency of weak algorithms.

Research Challenges

The methodologies outlined in the academic literature are still unable to effectively address the issues of the real world, despite the quick and significant improvements that have been made in machine learning-based healthcare systems.

It can be difficult to find huge, high-quality datasets that are representative of many demographics. Other crucial factors to take into account include addressing data imbalance and protecting data privacy.

For practical use and wide applicability, it is crucial that machine learning models generalize effectively to unknown data from different hospitals or patient groups.

Due to a variety of factors, medical databases frequently have missing values. It is essential to provide reliable ways for dealing with missing data and assign strategies that don't degrade model performance.

Healthcare data may contain noise and artifacts. For accurate predictions, machine learning models must be robust to noisy inputs.

Real-time detection is essential for prompt diagnosis and action in healthcare settings. Developing models that can provide fast and low-latency predictions is a challenge.

Healthcare AI is still working to develop models that can take into account the unique traits and medical histories of each patient.

Due to differences in datasets, preprocessing, and assessment measures, it can be challenging to replicate and validate the outcomes of machine learning models from various research papers.

By addressing these research issues, we may successfully use machine learning to the identification of cardiac disease and enhance patient outcomes in clinical practice.

Conclusion and Future Scope

This review study emphasizes the following observations:

- Most of the researchers have used ANN, SVM, Naïve Bayes, Random Forest, Logical regression, and KNN among other machine learning algorithms to predict Heart disease accurately.

Before any clinical signs appear, it is possible to identify individuals who are at risk. This offers up new opportunities for preventative care and individualized treatment approaches, improving patient outcomes and lowering heart disease-related morbidity and death.

This review study is intended to lay the groundwork for future investigation and innovation in the field of cardiac disease prediction as research and technology continue to develop.

In-depth and precise prediction models might be created through research into the successful combination of these multi-modal datasets with clinical information, allowing for personalized risk assessments and focused interventions.

In the future, we would integrate other datasets with additional observations in order to do more experiments. To increase the scalability and accuracy of this prediction system, a variety of potential changes might be investigated. In order to help doctors identify and diagnose such diseases as early as possible and potentially save millions of lives, this research offers a methodology that uses hybrid machine learning algorithms.

References

[1] Chaimaa Boukhatem, Heba Yahia Youssef, Ali Bou Nassif. (2022). "Heart Disease Prediction Using Machine Learning". Advances in Science and Engineering Technology International Conferences (ASET) IEEE | DOI: 10.1109/ASET53988.2022.9734880.

[2] Dara DVVNS Saikumar, Uma Priyadarsini P.S, I. Meignana Arumugam. (2022). "Prediction of Heart Disease Using K-Nearest Neighbour Algorithm in Comparison with Support Vector Machine Algorithm", 14th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS) IEEE | DOI: 10.1109/MACS56771.2022.10023034.

[3] Seema Gulati, Kalpna Guleria, Nitin Goyal. (2022). "Classification and Detection of Coronary Heart Disease using Machine Learning", 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) IEEE | DOI: 10.1109/ICACITE53722.2022.9823547.

[4] Priyanka Gupta, D.D. Seth. (2022). "Comparative Analysis of Machine Learning Classifiers for Accurate and Early Detection of Heart Disease". 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO).

[5] Harshit Jindal , Sarthak Agrawal ,Preeti Nagrath, Rachna Jain .(2021) “Heart disease prediction using machine learning algorithms” .IOP Conf. Ser.: Mater. Sci. Eng. 1022 012072.

[6] Sanjay Dhanka, surita maini. (2021). “Random Forest for Heart Disease Detection: A Classification Approach”, IEEE 2nd International Conference on Electrical Power and Energy Systems (ICEPES) .

[7] Md. Touhidul Islam, Sanjida Reza Rafa, Md. Golam Kibria . (2020). “Early Prediction of Heart Disease Using PCA and Hybrid Genetic Algorithm with k-Means”. 23rd International Conference on Computer and Information Technology (ICCIT) |IEEE | DOI: 10.1109/ICCIT51783.2020.9392655.

[8] Ravindhar NV, Anand, Hariharan Shanmugasundaram, Ragavendran, Godfrey Winster.
(2019) “Intelligent Diagnosis of Cardiac Disease Prediction using Machine Learning”, Volume-8 Issue-11, ISSN: 2278-3075 (Online). Page No: 1417-1421.
DOI: 10.35940/ijitee.J9765.0981119.

[9] Awais Nimat et al. (2019). “An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of Heart Failure”, IEEE Access, vol. 7, pp. 54007–54014, DOI:10.1109/ACCESS.2019.2909969.

[10] Md. Razu Ahmed, S M Hasan Mahmud, Md Altab Hossin, Hosney Jahan, Sheak Rashed Haider Noori. (2018). “A Cloud Based Four-Tier Architecture for Early Detection of Heart Disease with Machine Learning Algorithms”. IEEE 4th International Conference on Computer and Communications.

[11] C. B. Erdas and D. Olcer . (2020). “A Machine Learning-Based Approach to Detect Survival of Heart Failure Patients” Medical Technologies Congress (TIPTEKNO) pp. 1–4. DOI: 10.1109/TIPTEKNO50054.2020.9299320.

[12] S. M Mehedi Zaman, Wasay Mahmood Qureshi, Md. Mohsin Sarker Raihan, Abdullah Bin Shams and Sharmin Sultana. (2021). “Survival Prediction of Heart Failure Patients using Stacked Ensemble Machine Learning Algorithm”. International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE). IEEE, DOI: 10.1109/WIECON-ECE54711.2021.9829577.

[13] A. Ishaq et al., (2021). “Improving the Prediction of Heart Failure Patients’ Survival Using SMOTE and Effective Data Mining Techniques,” IEEE Access, vol. 9, pp. 39707–39716, DOI: 10.1109/ACCESS.2021.3064084.

[14] Shalet K.S, V. Sabarinathan, V. Sugumaran, V. J. Sarath Kumar (2015), “Diagnosis of Heart Disease Using Decision Tree and SVM Classifier”, International Journal of Applied Engineering Research, Vol. 10, No.68, pp. 598-602.

[15] Anshika Trisal , Vikas Sagar, Roshan Jameel. (2022). “Cardiac Disease Prediction using Machine Learning Algorithms”. International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES). DOI: 10.1109/CISES54857.2022.9844370.

[16] Rohit Bharti, Aditya Khamparia, Mohammad Shabaz, Gaurav Dhiman, Sagar Pande, and Parneet Singh. (2021). “Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning”. Hindawi Computational Intelligence and Neuroscience.

[17] Raja, M. S., Anurag, M., Reddy, C. P., & Sirisala, N. R. (2021). “Machine Learning Based Heart Disease Prediction System”. International Conference on Computer Communication and Informatics (ICCCI). IEEE Digital Object Identifier 10.1109/ACCESS.2020.2968900.

[18] Akanksha Kumari, Ashok Kumar Mehta. (2021). “A Novel Approach for Prediction of Heart Disease using Machine Learning Algorithms”. Asian Conference on Innovation in Technology (ASIANCON). DOI: 10.1109/ASIANCON51346.2021.9544544.

[19] Pranav Motarwar, Ankita Duraphe, G Suganya, M Premalatha. (2020). “Cognitive Approach for Heart Disease Prediction using Machine Learning”. International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 0.1109/ic-ETITE47903.2020.242.

[20] Kuldeep Vayadande, Rohan Golawar, Sarwesh Khairnar, Arnav Dhiwar, Sarthak Wakchoure, Sumit Bhoite , Darpan Khadke.(2022) . “Heart Disease Prediction using Machine Learning and Deep Learning Algorithms”. 1 st International Conference on Computational Intelligence and Sustainable Engineering Solution(CISES) DOI: 10.1109/CISES54857.2022.9844406.

[21] Lai, D., Zhang, Y., Zhang, X., Su, Y., & Bin Heyat, M. B. (2019). “ An Automated Strategy for Early Risk Identification of Sudden Cardiac Death by Using Machine Learning Approach on Measurable Arrhythmic Risk Markers”. Digital Object Identifier 10.1109/ACCESS.2019.2925847.

[22] K. Rajalakshmi and K. Nirmala (2016), “Heart Disease Prediction with MapReduce by using Weighted Association Classifier and K-Means”. Indian Journal of Science and Technology. Vol. 9, No. 19, pp.1-7.

[23] Pabitra Kumar Bhunia, Arijit Debnath, Poulami Mondal, Monalisa D E, Kankana Ganguly, Pranati Rakshit. (2021). “Heart Disease Prediction using Machine Learning”.

[24] Mihir J. Gaikwad, Prathmesh S. Asole, Prof. Leela S. Bitla. (2022). “Effective Study of Machine Learning Algorithms for Heart Disease Prediction”. 2nd International Conference on Power Electronics & IoT Applications in Renewable Energy and Its Control (PARC).
DOI: 10.1109/PARC52418.2022.9726613.

[25] Arun. R and N. Deepa (2018), “Heart Disease Prediction System Using Naive Bayes”. International Journal of Pure and Applied Mathematics. Vol. 119, No. 16, pp. 3053-3065.