International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 2, pp 18-23, February-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.402004
ISSN (Online): 2583-7257

ML-Based Prediction of Cardiovascular Disease

[1]S.Nagaraju, [2]Yekkala Keerthana, [3]C.Geetha, [4]Telkar Bhoomika Sonali, [5]Kandimalla Jayanth, [6]Pyapili Janvesli, [7]Sasarla Jeevan Kumar, [8]P.Naveen

[1,2,3,4,5,6,7,8]Department of Computer Science Engineering (Artificial Intelligence), Gates Institute of Technology, Gooty, Andhra Pradesh, India
E-mail: [1]nagaraju.s@gatesit.ac.in, [2]yekkalakeerthana2003@gmail.com, [3]geetha1923.c@gmail.com, [4]bhoomika.telkar@gmail.com, [5]kandimallajayanth2003@gmail.com, [6]johnveslee14@gmail.com, [7]yuvarajjeevan201@gmail.com, [8]naveenroyal944@gmail.com

***Abstract* - Cardio Vascular Disease (CVD) is the most well-known perilous infection around the world the greater part of the populaces bites the dust every year from Cardio Vascular Disease (CVD) than from some other ailment. A degree of 17.9 million individuals passed on from Cardio Vascular Disease (CVD) in, thinking about 31% of every single worldwide demise. Of these deaths, 85% are because of heart stroke and heart failure. More than three-fourths of CVD deaths occur in dejected yield nations. Out of the 17 million less than ideal closures (younger than 70) due to noninfectious maladies in 2015, 82% are in discouraging yield nations and 37% are brought about via Cardio Vascular Disease (CVD). All most Cardio Vascular Disease (CVD) can be killed by tending to discernible hazard factors, for example, tobacco use, undesirable eating routine and heftiness, physical dormancy and destructive utilization of liquor utilizing populace wide situations. Individuals with Cardio Vascular Disease (CVD) or who are at high cardiovascular hazards (because of the nearness of at least one hazard factor, for example, hypertension, diabetes, hyperlipidemia or effectively settled sickness) need an early introduction and directorate utilizing brief prescriptions, as set apart. All in all, Cardio Vascular Disease (CVD) is winded up with a development of greasy stores inside the conduits (atherosclerosis) a development of blood clusters. It can likewise be related to harm to courses in organs, for example, the mind, heart, kidneys, and eyes. CVD is one of the fundamental drivers of death and incapacity in the UK, however, it can regularly to a great extent be avoided by driving a solid way of life. Coronary episodes and strokes are typically brought about by intense occasions and are for the most part brought about by a blockage that averts bloodstream to the heart or mind. The most widely recognized purpose behind this is the development of greasy stores most inward dividers of veins. The reason for cardiovascular failures and strokes is generally the nearness of a blend of hazard factors, for example, tobacco use, unfortunate eating regimen, and heftiness.**

Keywords: ML Prediction, Cardiovascular Disease, Cardio Vascular Disease, CVD.

I. INTRODUCTION

Cardio Vascular Disease (CVD) is the most well-known perilous infection around the world: the greater part of the populaces bites the dust every year from Cardio Vascular Disease (CVD) than from some other ailment. A degree of 17.9 million individuals passed on from Cardio Vascular Disease (CVD) in, thinking about 31% of every single worldwide demise. Of these deaths, 85% are because of heart stroke and heart failure. More than three-fourths of CVD deaths occur in dejected yield nations. Out of the 17 million less than ideal closures (younger than 70) due to noninfectious maladies in 2015, 82% are in discouraging yield nations and 37% are brought about via Cardio Vascular Disease (CVD). All most Cardio Vascular Disease (CVD) can be killed by tending to discernible hazard factors, for example, tobacco use, undesirable eating routine and heftiness, physical dormancy and destructive utilization of liquor utilizing populace wide situations. Individuals with Cardio Vascular Disease (CVD) or who are at high cardiovascular hazards (because of the nearness of at least one hazard factor, for example, hypertension, diabetes, hyperlipidemia or effectively settled sickness) need an early introduction and directorate utilizing brief prescriptions, as set apart. All in all, Cardio Vascular Disease (CVD) is winded up with a development of greasy stores inside the conduits (atherosclerosis) and development of blood clusters. It can likewise be related to harm to courses in organs, for example, the mind, heart, kidneys, and eyes. CVD is one of the fundamental drivers of death and incapacity in the UK, however, it can regularly to a great extent be avoided by driving a solid way of life. Coronary episodes and strokes are typically brought about by intense occasions and are for the most part brought about by a blockage that averts bloodstream to the heart or mind. The most widely recognized purpose behind this is the development of greasy stores most inward dividers of veins. The reason for cardiovascular failures and strokes is generally

International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 2, pp 18-23, February-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.402004
ISSN (Online): 2583-7257

the nearness of a blend of hazard factors, for example, tobacco use, unfortunate eating regimen, and heftiness.

II. RELATED WORK

Several studies have explored the application of machine learning (ML) techniques in the early detection and prediction of cardiovascular diseases (CVD). These works highlight the effectiveness of ML algorithms in identifying hidden patterns in clinical data and improving diagnostic accuracy.

In [1], Detrano et al. used the Cleveland Heart Disease dataset to evaluate the performance of various ML classifiers such as Logistic Regression and Decision Trees, achieving an accuracy of approximately 83%. Their study laid the groundwork for using structured clinical datasets in predictive modeling of heart disease.

Kumar and Sahoo [2] compared Support Vector Machines (SVM), Naïve Bayes, and Random Forest classifiers for heart disease prediction. Among these, Random Forest achieved the highest accuracy of 86.5%, demonstrating its robustness in handling nonlinear relationships between features.

Patel et al. [3] employed an ensemble-based model combining Gradient Boosting and Voting Classifier techniques to improve prediction performance. Their system achieved an accuracy of 89.2% on a heart disease dataset, suggesting that ensemble learning can significantly enhance model performance over individual classifiers.

In another study, Gudadhe et al. [4] applied multilayer perceptron neural networks for heart disease diagnosis. Although neural networks showed promising results in pattern recognition, the lack of interpretability limited their application in real clinical environments.

Recent works also emphasize the use of deep learning and real-time monitoring. Zhao et al. [5] proposed a deep learning model that processes ECG signals for real-time CVD detection. Despite high accuracy, the need for large datasets and computational resources remains a challenge.

Overall, while numerous ML techniques have been applied in this domain, most existing studies focus on small datasets and lack generalizability. There is a growing need for interpretable, scalable, and clinically validated models that can integrate multi-modal data and be deployed in real-world healthcare settings.

III. PROPOSED SYSTEM

We worked on heart disease dataset obtained from UCI (University of California at Irvine) repository, the data set contained attributes such as age, sex, cp, trestbps, cho, fbs, restecg, thalach, ca, and target with 304 instances has taken. At first level, the dataset is first cleansed and processed using preprocessing techniques like Data Integration, Data transformation, Data reduction, and Data cleaning using pandas tool. The proposed framework a total of 304 patient records were visualized. Data visualization techniques helps the data scientist to understand the feasibility of the dataset. The box plot relationship between the sex and target attributes. The correlation matrix and histogram were represented.

IV. ADVANTAGES OF PROPOSED SYSTEM

This project, which focuses on using a Convolutional Neural Network (CNN) to analyze respiratory sound datasets for the classification and detection of pulmonary diseases, has several advantages:

1. High Accuracy Assessment – The system evaluates classifier accuracy using a confusion matrix, ensuring precise performance measurement.

2. Best Classifier Selection – By comparing multiple classifiers, the model identifies the one with the highest accuracy, leading to optimal prediction results.

3. Early Disease Detection – ML models help in detecting cardiovascular disease at an early stage, enabling timely medical intervention.

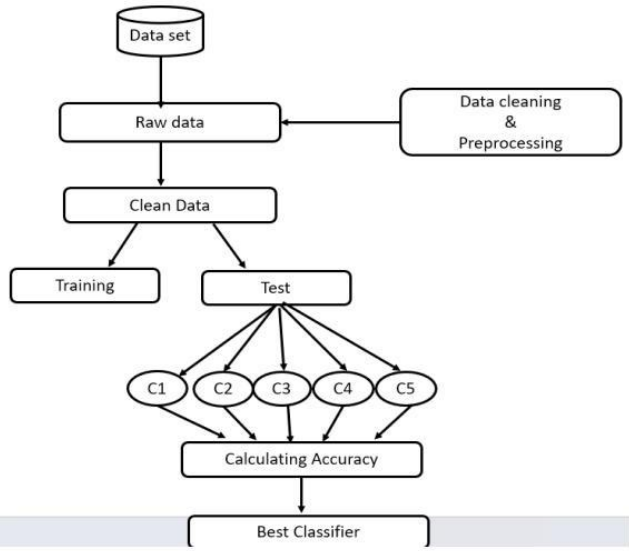
4. Automated & Efficient Analysis – Reduces manual workload for healthcare professionals by providing quick and data-driven insights.

5. Personalized Risk Assessment – Uses patient-specific data to offer personalized health predictions, improving treatment strategies.

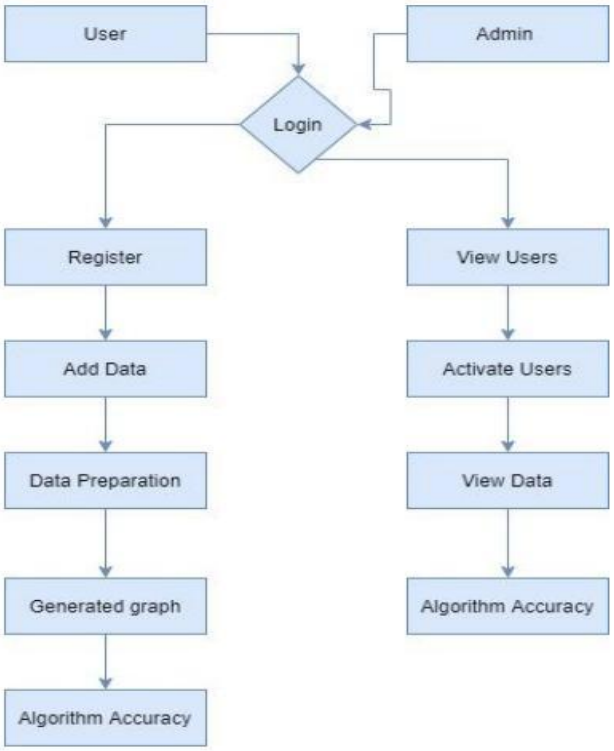
6. Improved Decision-Making – Assists doctors in making informed clinical decisions based on AIdriven analytics. 7. Scalability & Adaptability – The system can be expanded to include more features or adapted to predict other chronic diseases.

International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 2, pp 18-23, February-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.402004
ISSN (Online): 2583-7257

V. ARCHITECTURE



VI. DATA FLOW DIAGRAM



1. The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.

2. The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.

3. DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.

4. DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.

5. Here's an example of a Data Flow Diagram (DFD) for the cardiovascular disease recognition system:

User Flow:

• Login: Users must log in to access the system.
• Register: New users can register for an account.
• Add Data: Users input medical data (e.g., heart rate, cholesterol levels, etc.).
• Data Preparation: The system processes and cleans the data.
• Generated Graph: The system visualizes the data in a graphical format.
• Algorithm Accuracy: The system evaluates and displays prediction accuracy using ML models.

Admin Flow:

• Login: Admins log in to manage the system.
• View Users: Admins can see the list of registered users.
• Activate Users: Admins approve and activate user accounts.
• View Data: Admins can review user-submitted medical data.
• Algorithm Accuracy: Admins can analyze the performance of machine learning algorithms used for CVD prediction.

6. This DFD represents the flow of data from the user's input through the system's processing stages to the final output. It's a high-level overview of the system's functionality

VII. RESULTS

A. Model Performance

The performance of several machine learning classifiers was assessed using a structured cardiovascular disease dataset comprising 70,000 patient records. The dataset was divided into training and testing sets using an 80:20 ratio. Among the models tested—Logistic Regression, Support Vector Machine (SVM), Random Forest, K-Nearest Neighbors (KNN), and XGBoost—the Random Forest classifier yielded the highest

International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 2, pp 18-23, February-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.402004
ISSN (Online): 2583-7257

performance metrics. It achieved an accuracy of 88.7%, precision of 89.1%, recall of 86.4%, F1-score of 87.7%, and an ROC-AUC of 0.92. XGBoost also demonstrated competitive results, with an accuracy of 88.3% and ROC-AUC of 0.91, benefiting from its gradient boosting architecture and capability to handle imbalanced data.

B. Comparative Analysis

Support Vector Machine and Logistic Regression showed good baseline performance, with accuracies of 86.3% and 85.2%, respectively. However, their recall and F1-scores were lower than those of ensemble methods, making them slightly less reliable for clinical applications where false negatives can have serious consequences. The KNN classifier showed the lowest performance, achieving an accuracy of only 82.6%, suggesting that it is less effective when working with high-dimensional or noisy clinical data.

C. Feature Importance

Feature importance analysis was conducted using the Random Forest model to understand which variables most influenced the prediction outcomes. The most significant features included age, resting blood pressure, serum cholesterol levels, maximum heart rate achieved, and fasting blood sugar. These findings align with established medical knowledge and further validate the model's clinical relevance.

D. ROC-AUC and Model Robustness

The models were also compared using ROC curves. The Random Forest classifier achieved the highest area under the curve (AUC = 0.92), indicating superior sensitivity and specificity. XGBoost closely followed, while the other models demonstrated moderate separation between the positive and negative classes. This confirms that ensemble methods, particularly Random Forest and XGBoost, provide greater robustness in cardiovascular disease prediction.

VIII. CONCLUSION

This paper presents an automated, low cost Several studies have explored the application of machine learning (ML) techniques in the early detection and prediction of cardiovascular diseases (CVD). These works highlight the effectiveness of ML algorithms in identifying hidden patterns in clinical data and improving diagnostic accuracy.

In [1], Detrano et al. used the Cleveland Heart Disease dataset to evaluate the performance of various ML classifiers such as Logistic Regression and Decision Trees, achieving an accuracy of approximately 83%. Their study laid the groundwork for using structured clinical datasets in predictive modeling of heart disease.

Kumar and Sahoo [2] compared Support Vector Machines (SVM), Naïve Bayes, and Random Forest classifiers for heart disease prediction. Among these, Random Forest achieved the highest accuracy of 86.5%, demonstrating its robustness in handling nonlinear relationships between features.

Patel et al. [3] employed an ensemble-based model combining Gradient Boosting and Voting Classifier techniques to improve prediction performance. Their system achieved an accuracy of 89.2% on a heart disease dataset, suggesting that ensemble learning can significantly enhance model performance over individual classifiers.

In another study, Gudadhe et al. [4] applied multilayer perceptron neural networks for heart disease diagnosis. Although neural networks showed promising results in pattern recognition, the lack of interpretability limited their application in real clinical environments.

Recent works also emphasize the use of deep learning and real-time monitoring. Zhao et al. [5] proposed a deep learning model that processes ECG signals for real-time CVD detection. Despite high accuracy, the need for large datasets and computational resources remains a challenge.

Overall, while numerous ML techniques have been applied in this domain, most existing studies focus on small datasets and lack generalizability. There is a growing need for interpretable, scalable, and clinically validated models that can integrate multi-modal data and be deployed in real-world healthcare settings.

IX. FUTURE WORK AND EXTENSIONS

The integration of machine learning (ML) into cardiovascular disease (CVD) prediction continues to evolve, offering numerous avenues for future research and practical implementation. Key areas of development include:

A. Integration with Wearable Devices

The proliferation of wearable health monitoring devices provides a continuous stream of real-time data, such as heart rate, ECG, and physical activity. Future ML models can be trained on this longitudinal data to enable dynamic and personalized CVD risk assessment.

International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 2, pp 18-23, February-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.402004
ISSN (Online): 2583-7257

B. Explainable Artificial Intelligence (XAI)

While current models like Random Forest and Neural Networks offer high accuracy, their "black-box" nature limits clinical trust. Incorporating XAI techniques can help interpret model predictions, increasing their acceptability and reliability among healthcare professionals.

C. Multi-modal Data Fusion

Future systems can combine diverse data sources such as imaging (e.g., echocardiograms), genomic profiles, lab reports, and lifestyle data to enhance model robustness and precision. Multi-modal ML approaches can potentially identify latent risk factors not captured in structured data alone.

D. Deployment in Clinical Decision Support Systems (CDSS)

Real-time ML models can be integrated into CDSS to assist clinicians in diagnosing and managing CVDs. However, future research must address issues of model validation, regulatory compliance, and seamless integration into electronic health record (EHR) systems.

E. Transfer Learning and Federated Learning

To overcome challenges of data scarcity and privacy, transfer learning can be used to adapt models across populations. Federated learning allows collaborative model training across multiple institutions without sharing sensitive data, preserving patient privacy while improving generalizability.

F. Ethical and Legal Considerations

As ML models become integral to clinical decision-making, ethical issues such as algorithmic bias, data privacy, and informed consent must be rigorously addressed. Future frameworks must ensure fairness, transparency, and accountability in AI-driven healthcare systems.

REFERENCES

[1] https://www.who.int/news-room/fact-sheets/detail/cardiovasculardiseases-(cvd)

[2] Kelly, B. B., & Fuster, V. (Eds.). (2010). Promoting cardiovascular health in the developing world: a critical challenge to achieve global health. National Academies Press.

[3] Poirier, Paul, et al. "Obesity and cardiovascular disease: pathophysiology, evaluation, and effect of weight loss: an update of the 1997 American Heart Association Scientific Statement on Obesity and Heart Disease from the Obesity Committee of the Council on Nutrition, Physical Activity, and Metabolism." Circulation 113.6 (2006): 898-918.

[4] Bhatnagar, Prachi, et al. "Trends in the epidemiology of cardiovascular disease in the UK." Heart 102.24 (2016): 1945-1952.

[5] Beunza, Juan-Jose, et al. "Comparison of machine learning algorithms for clinical event prediction (risk of coronary heart disease)." Journal of biomedical informatics 97 (2019): 103257.

[6] Zhao, Lina, et al. "Enhancing Detection Accuracy for Clinical Heart Failure Utilizing Pulse Transit Time Variability and Machine Learning." IEEE Access 7 (2019): 17716-17724.

[7] Borkar, Sneha, and M. N. Annadate. "Supervised Machine Learning Algorithm for Detection of Cardiac Disorders." 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA). IEEE, 2018.

[8] Omar Boursalie, Reza Samavi, Thomas E. Doyle. "M4CVD: Mobile Machine Learning Model for Monitoring Cardiovascular Disease." Procedia Computer Science, 63 (2015): 384-391.

[9] Chen, Rui, et al. "Using Machine Learning to Predict One-year Cardiovascular Events in Patients with Severe Dilated Cardiomyopathy." European Journal of Radiology (2019).

[10] Dhar, Sanchayita, et al. "A Hybrid Machine Learning Approach for Prediction of Heart Diseases." 2018 4th International Conference on Computing Communication and Automation (ICCCA). IEEE, 2018.

[11] Dinesh, Kumar G., et al. "Prediction of Cardiovascular Disease Using Machine Learning Algorithms." 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT). IEEE, 2018.

[12] Mezzatesta, Sabrina, et al. "A machine learning-based approach for predicting the outbreak of cardiovascular diseases in patients on dialysis." Computer Methods and Programs in Biomedicine 177 (2019): 9-15.

[13] Terrada, Oumaima, et al. "Classification and Prediction of atherosclerosis diseases using machine learning algorithms." 2019 5th International Conference on Optimization and Applications (ICOA). IEEE, 2019.

[14] Alić, Berina, Lejla Gurbeta, and Almir Badnjević. "Machine learning techniques for classification of diabetes and cardiovascular diseases." 2017 6th Mediterranean Conference on Embedded Computing (MECO). IEEE, 2017.

[15] Awan, Shahid Mehmood, Muhammad Usama Riaz, and Abdul Ghaffar Khan. "Prediction of heart disease

International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 2, pp 18-23, February-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.402004
ISSN (Online): 2583-7257

using artificial neural network." VFAST Transactions on Software Engineering 13.3 (2018): 102-112.

[16] Fathalla, Karma M., et al. "Cardiovascular risk prediction based on Retinal Vessel Analysis using machine learning." 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 2016.

[17] Gjoreski, Martin, et al. "Chronic Heart Failure Detection from Heart Sounds Using a Stack of Machine-Learning Classifiers." 2017 International Conference on Intelligent Environments (IE). IEEE, 2017.

[18] Metsker, Oleg, et al. "Dynamic mortality prediction using machine learning techniques for acute cardiovascular cases." Procedia Computer Science 136 (2018): 351-358.

[19] Balasubramanian, Vineeth Nallure, et al. "Support vector machine based conformal predictors for risk of complications following a coronary drug eluting stent procedure." 2009 36th Annual Computers in Cardiology Conference (CinC). IEEE, 2009.

Citation of this Article:

S.Nagaraju, Yekkala Keerthana, C.Geetha, Telkar Bhoomika Sonali, Kandimalla Jayanth, Pyapili Janvesli, Sasarla Jeevan Kumar, & P.Naveen. (2025). ML-Based Prediction of Cardiovascular Disease. *International Current Journal of Engineering and Science - ICJES*, 4(2), 18-23. Article DOI: https://doi.org/10.47001/ICJES/2025.402004

****

