International Journal of Informatics and Communication Technology (IJ-ICT)
Vol. 15, No. 2, June 2026, pp. 719~728
ISSN: 2252-8776, DOI: 10.11591/ijict.v15i2.pp719-728

Utilizing the machine learning-driven techniques used to ECG dataset for predicting coronary heart disease

Mohd Osama[1,2], Rajesh Kumar[1], Chandrakant Kumar Singh[3]
[1]Department of Electronics and Communication, University of Allahabad, Prayagraj, India
[2]United Institute of Technology, Prayagraj, India
[3]U.P. Rajarshi Tandon Open University, Prayagraj, India

Article Info

Article history:

Received Feb 26, 2025
Revised Jul 2, 2025
Accepted Nov 5, 2025

Keywords:

Coronary heart disease
Decision tree classifier
Logistic regression
Machine learning
Support vector machine

ABSTRACT

The worldwide cause of mortality is cardiovascular heart disease. The automatic prediction of heart disease can be made to possible for accurate detection in initial stage. In recent year, the artificial intelligence approaches giving promising outcomes in predicting various types of cardiovascular conditions. The main focous of this work is to implementation of various machine learning techniques used to predict cardiovascular heart disease (CHD) using electrocardiogram (ECG) datasets. ECG provide the electrical Signal from the heart that identify the presence of disease or not. The preprocessing method are used for improving the quality of ECG signals and extract the features from ECG of patients. There are several well-established machine learning techniques, including support vector machine (SVM) and K-nearest neighbour (KNN)., logistic regression and decision tree classifier used for prediction of the disease. So, our finding of this paper will provide the new understanding regarding CHD prediction using different machine learning techniques. The Decision Tree-based machine learning model demonstrated excellent performance, achieving 98% accuracy, 96% precision, 100% recall, and an F1-score of 97%, which is better than rest of other comparative machine learning models. Finaly expermental results shows that decision tree approach providing better outcome amongs all the algorithms with respect to all above mensioned parameter.

Corresponding Author:

Rajesh Kumar
Department of Electronics and Communication, University of Allahabad
Prayagraj, India
Email: rajeshkumariitbhu@gmail.com

1. INTRODUCTION

Coronary heart disease (CHD) is the very common type of heart disease in the world, and it is a major cause of death. Heart diseases have killed 1.7 million Indians in 2016, increasing health care costs and reducing productivity, costing India up to $237 billion from 2005-2015 [1]. According to World Health Organization (WHO) estimates that 85% people died from heart attack and stroke, that estimated 17.9 million peoples are died from CHD in 2019. Diagnosis of coronary disease is very crucial task. Early diagnosis of disease increases the chances of a better treatment [2]. CHD are a serious and potentially life-threatening condition that affects the cardiovascular system, characterized by the narrowing or blockage of the coronary arteries. The diagnosis of CHD involves a combination of physical examination, medical history assessment and various diagnostic tests [3]. These tests include medical history assessment, physical examination, electrocardiogram (ECG), stress tests, echocardiogram, coronary angiography, blood tests, cardiac CT scans, magnetic resonance imaging (MRI), and nuclear imaging studies [4], [5]. There are mainly two approaches used to diagnosis of CHD. First approach is the invasive process and latter one is the non-invasive processes. Invasive process is a painful diagnostic procedure that provides important information of heart structure and function of the heart [6]. In invasive procedure, a catheter has been penetrated to get X-rays of the heart's arteries (i.e., coronary arteries) called coronary Angiography or Arteriography. In non-invasive process is painless procedure to cardiovascular diagnostic testing includes the complete spectrum of heart [7], [8]. Non-invasive process has been done by highly skilled doctors and nurses, that having the knowledge to latest diagnosis technology [9]. A non-invasive technique that measures the electrical activity of the heart, stress tests evaluate the heart's performance during physical activity or under stress, echocardiogram uses sound waves to create images of the heart's structure and function. Coronary angiography is an invasive techniques that involves injecting a contrast dye into the coronary arteries, and blood tests measure certain substances in the blood [10], [11].

ECG, which is observed by measuring the electrical activity of the heart is one of the most useful set of procedure in detecting CHD [12]. ECG signals are outcome of non-invasive, cheap and readily accessible diagnostic device, which can be very vital in early detection of anomalies in the heart [13]. Raw ECG signals are very susceptible to noise and variability. It is very important to preprocess the Raw signals to enhance the quality of the signal therefore it is able to define some meaningful features [14]. After preprocessing, effectively apply the machine learning (ML) models with data integrity. The used data must maintain and bare minimum errors that result good precision in classification of heart disease [15].

A lot of ML algorithms were implemented in the prediction of CHD and they are logistic regression (LR), support vector machines (SVM), K-nearest neighbors (KNN) and decision trees (DT). These classifiers uses different mathematical concepts with their abilities to learn and interpretability. These models were applied on the same dataset with different evalution parameters [16]. The main objective of this study is to make comparison of performance of these algorithms for predicting occurrence of CHD with use of ECG datasets [17]. Performance metrics are applied to each model, including the accuracy, precision, recall, and F1-score. Although the contributions and successful results achieved in the past studies, having still multiple problems and gaps. The quality and preprocessing of the ECG data are one of major challenge because of noise and artifacts may affects the accuracy of the models without effective processing [18]. Moreover, a number of the available studies have narrow list of ML models or left the models unevaluated through performance variables over sets of datasets. Also, The privious works does not contain comparative studies that apply the same evaluation framework to a number of classifiers on a common dataset to give a clear picture of which algorithm is the most appropriate in CHD prediction [19]. This study fills these gaps by introducing a standardized experimental protocol that process raw ECG data, using several classification methods, and conducts the analysis of the results according to well defined metrics [20]. The experimental results of this work shows that the DT classifier gives better recall and F1-scores, as compared to other ML algoriths [21]. This research also helps to understand better the ways to understand ML in order to increase the predictive potential of the CHD detection systems.

2. LITERATURE REVIEW

This paper will conduct a review of literature on the past studies and will discuss how different machine learning algorithms can be used to predict coronary heart disease (CHD) using ECG data. Another fusion-based model that involves a holistic ultrasound image analysis method through the combination of the merits of several components is also discussed in the study. The model obtains good and reliable classification performance through the use of a pre-trained VGG19 network, discrete wavelet transform (DWT) to pre-process the data, and the use of advanced segmentation methods [22]. Experimental tests carried out on ultrasound images can show the relative performance of various algorithms.

The authors of this paper describe a strategy that should be used to enhance the result tracing, accountability, transparency, and model refinement in the healthcare industry. The analysis also uses the LIME method to determine major factors that can affect the heart disease which includes variables like ST slope and cholesterol levels. The model is based on these features and thus projects the probability of a patient developing cardiovascular disease. Also, SHAP is used to prioritize significant features, order them by their importance, to gain a better understanding of how the model makes decisions. The strategy also proposes the problem in terms of submodular optimization methods to improve interpretability and efficiency [23].

The authors of this study suggested a hybrid decision support system that would be used to identify coronary heart disease in its early stages based on specific clinical parameters of heart disease (HD) patients. In order to enhance the selection of features, hybridization of Genetic Algorithm (GA) and Recursive Feature Elimination (RFE) was used to determine the most applicable features in the dataset. The trained system had an accuracy of 86.6 which is better than some of the current heart disease prediction models mentioned in the literature [24].

In this study, the author introduced a smart healthcare framework for heart disease assessment using deep learning and feature fusion techniques. Initially, the feature fusion approach integrates features extracted from both sensor data and electronic medical records to produce meaningful healthcare information. Furthermore, an additional method is applied to remove irrelevant and redundant features, improving data quality. An efficient feature selection technique is employed to reduce computational complexity and enhance overall system performance. The proposed model was evaluated using heart disease datasets and compared with existing classifiers based on feature fusion, feature selection, and weighting strategies. The results demonstrate that the system achieved an accuracy of 98%, outperforming other existing approaches [25].

Author discusses about the Advancements of technologies have to improved results in field of biomedical science and research processes, particularly use in electrocardiogram. Electrocardiogram generate the electrical signal and indicating a significant increment in the number of people suffering from heart disease. The primary challenge of ECG signals is managing the irregularities and detecting the patient's condition. There are five machine learning model are employed to identify the specific disease and differentiate between the patient that predict the classes of normal and abnormal readings ECG characteristics. The of diseases RBBB and CAS has yielded improved the results used by logistic regression and support vector machine algorithms that achieving the accuracy rate of 95% respectively [26].

3. RESEARCH METHOD

The dataset used in our study, is ECG-Dataset.csv which is accessed the public datasets from Kaggle. This dataset contains information (like age, sex, smoke, ldl, chp, weight, ihd, bpdias, bpsys, and so on) about the group of various person having coronary heart disease [27]. All the implemetation have been done on intell core i5 processor, with 8 GB RAM , and 500 GB SSD on windows 11 plaform in google Co-Lab environment. Then the dataset is splitted in two subparts i.e. the training dataset, testing dataset. The training datasets having the ratio of 80% and testing datasets splits in the ratio of 20% for predict the results. Using some other operation like as remove the null parameter for analysis the better performance for machine learning models. Then we have selected the some of the supervised machine learning approaches use to comparison and evaluation of CHD prediction model. The proposed scheme consist of various steps are data preprocessing, also divide the data in training and testing, results evaluation and building predictive model. Which provides time-saving environment for diagnosis of the coronary heart diseases to the cardiologist/physician All the chosen models are trained on the training dataset, then tested on the testing dataset [28]. These experimenal proposed model shows in Figure 1.

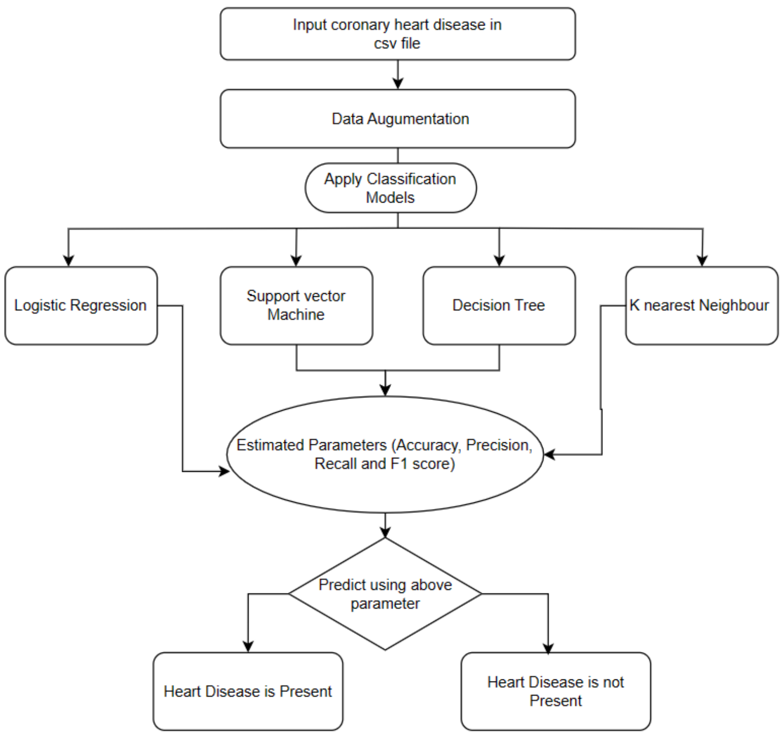

Figure 1. Proposed model of CHD prediction

After receiving the data with the use of the sufficient level of preprocessing, the second step was executing various supervised classification methods to the data. These models used for this research were LR, SVM, DT, and KNN. The choice of these models can be explained by the fact that they are effective in medical and clinical prediction tasks. Final outcome the this framwork for all the classifer is wheather heart disease is present or not. The following algorithms have been used this work is explained as follows:

3.1. Logistic regression (LR)

The best LR is supervised ML algorithms used for most applications. LR categorised on dependent variables and produced discreate variable like as 0 or 1. The cost function works as sigmoid activation function that predicts the probabilistic value deal with 0 or 1. So, the Logistic sigmoid activation function is calculated as (1).

$$P(x) = \frac{1}{(1+e^{-x})} \tag{1}$$

Where the p(x) probabilistic estimation function value that bound by 0 to 1, x is input variable of predicted value of probability functions and e is the Euler's functions. Using above equation (1), to predict the CHD by LR model [29]. This model splits the CHD dataset trained and test with the best accuracy to prediction of CHD.

3.2. Support vector machine (SVM)

In case CHD prediction, SVM play the important role using various parameter like as bp, sugar and others. Svm algorithms mostaly used for classification that find the optimal hyperplane that saparates into various classes with maximum margin [30]. Svm used a binary classifier contain that represent the set of features xi associated with class label yi. so, the decision function of hyperplane that separates with two classes is written as (2).

$$f(x) = \ sign(w.x+b) \tag{2}$$

Where w is define weight vector with b is bias value, and sign function indicate the boundary condition between -1 or +1 and x is sample class that indicate the prediction. The margin distance of hyperplane and denoted as x+ as positive support vector and x- as negative support vector. So, the margin is calculated as (3).

$$margin \ = \frac{2}{||w||} \tag{3}$$

||w|| is defined as Euclidean normal form of weight vectors w. So, the objective function of SVM can be written as with maximum margin with equivalent minimum margin is (4).

$$\begin{gathered} minimize\ of\ \frac{1}{2}\,||w||^2 + C * \boldsymbol{\xi}_I \\ \text{Subject: } y_i\,(w.x_i + b) \geq 1 - \boldsymbol{\xi}_I \text{ and } \boldsymbol{\xi}_i \geq 0 \end{gathered} \tag{4}$$

Where C parameter is balance among the margin and also used for misclassification. When increase the value of C then margin will be small with less misclassification values and vice versa.

3.3. K-nearest neighbour (KNN)

KNN is also a supervise learning simple approach. It is stores whole case and classify the new behaviour on the basis of similarity measurement scale. KNN used for reasoning-based cases and apply knn algorithms. In these techniques, all data is classified into training and testing datasets and evaluated with the point of lowest distance [31]. There are the many ways to find the similarity among the variable with n numbers of attributes values. Let us consider the distance from point A to point B is measure;

$$1. Dist(A,B)\ \geq 0\ and\ Dist(A,B) = 0\ iff\ A = B \tag{5}$$

$$2. Dist(A,B) = Dist\ (B,A) \tag{6}$$

$$3. Dist(A,C)\ \leq\ \ Dist(A,B) + Dist(B,C) \tag{7}$$

All these property of KNN used to measurement of distance between the two points and third property is defined as “Triangle in Equality”. And also increase the accuracy is depend on the value of K.

3.4. Decision tree classifier

Most popular algorithm of ML is DT algorithms. DT algorithms constructed the tree and split the value node of class. the splited leaf node of same class is create the DT [32]. It is defining the gain ration with gain knowledge for every feature a. Then information gain (IG) is calculated.

$$IG(a) = Ent(S) - \sum_{Val\,(a)} \frac{|S_a|}{|a|} * Ent(S_a) \tag{8}$$

Where S_a is the subset of splited feature. Val_a is belong of all possible value of ‘a’ and |a| is defined as total number of ‘a’ value. And also entropy of S is calculated as (9).

$$Ent(S) = \sum_{j=1}^{num\ of\ Class} \frac{freq(L_j,S)}{|S|} * \ log_2(\frac{freq(L_j,S)}{|S|}) \tag{9}$$

Where Lj be the set of classes, and num of class is distinct classes that predict the disease are positive or negative. After that all the selected ML algorithm are evaluated using various parameters such as accuracy score, precision score, recall value and F1-score [33]. Some steps for the prediction of heart disease of model given in Figure 1. Then the results obtained for ML models are represented using tables. And comparisons are made between the selection evaluation metrics of the selected machine algorithms. This will be helpful in identifying the most accurate and efficient ML model among the selected these algorithms that can be used for accurate diagnosis of CHD and advancement in medical science. The accurate diagnose of heart disease using some analytical parameters of confusion metrics like as true negative (TN), true positive (TP), false negative (FN), and false positive (FP) respectively [34]. Using confusion metrics parameters, it helps and understand the performance of classification that analysis and predict the behaviour of CHD. The performance of confusion metrics is identify the types of errors that makes [35]. So, a predicting measurement of machine leaning problems that can be classified two or more classes. There are four different outcomes are given as-

Precision: of precision is the correctly predicted positive parameters between all predicted parameter are positive [36]. The precision value calculated as (10).

$$Precision = \frac{TP}{TP+FP} \tag{10}$$

Recall: the recall value is defined as the correctly predicted positive parameter of CHD from all actual positive parameters [37]. The recall value are calculated by (11).

$$Recall = \frac{TP}{TP+FN} \tag{11}$$

F1-score: the property of F1-score can be defined as the harmonics mean of both precision value and recall values [38]. This is calculated by (12).

$$F1\ Score = 2 * \frac{Precision*Recall}{Precision+Recall} \tag{12}$$

Accuracy: the proportion of all correctly classified parameters is called accuracy [39]. Which is calculated by (13).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{13}$$

It is applied to determine the precision of a classification model that is, the numbers of correct predictions in total.

4. RESULTS

In this experiment, we have to predicted the outcomes and discussed about all related progress ML architecture which calculates the predict method in the modality of CHD that are commonly used to evaluation of CHD prediction. Using successful processing of the dataset, the following results are obtained.

In one iteration, the whole process is performed by considering the selected features of given dataset. To evaluate the result of confusion matrix from Table 1. Using confusion matrix, we have shows the various parameters like as accuracy, recall, precision and F1-score are compare with the existing work of ML approaches in this experiment. Which is shows in Table 2, We have calculated the all-evaluation parameter with the comparison table which is predict the better results from existing model applied on CHD dataset.

From Table 1, we have to show that confusion matrix having value of true positive (TP), true negative (TN), false positive (FP) and false negative (FN). Using these parameters to calculating the various result with the present work of our experiment and visualiesed the result in Figure 2.

Table 1. Confusion matrix of various ML model

Model name	TN	TP	FN	FP
LR	23	39	1	4
SVM	24	39	0	4
KNN	4	37	20	6
DT classifier model	24	42	0	1

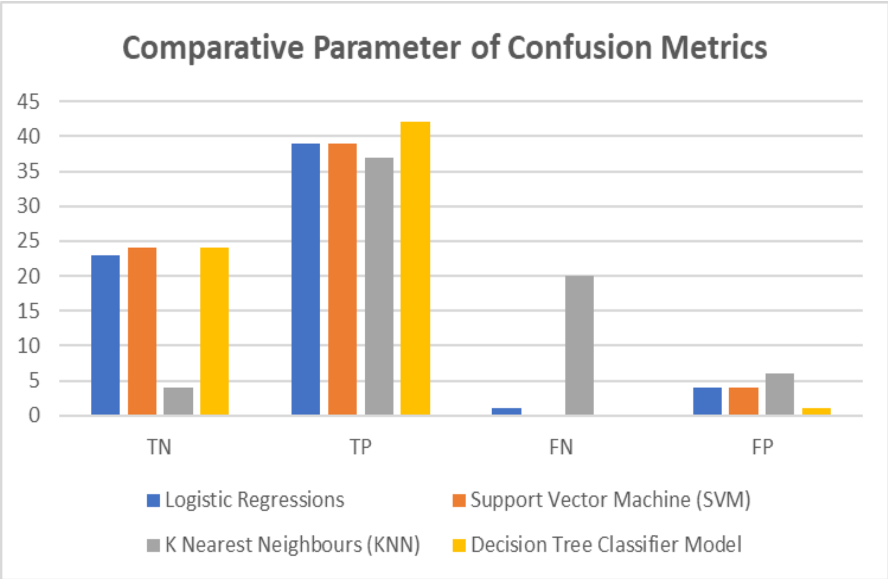

Figure 2. Comparative parameter of confusion metrics

We have calculated the all-evaluation parameter like as precision, recall, F1-score, and accuracy with the comparison of various existing model on CHD dataset. From Table 2, we have calculated the evaluation parameter of precision, recall, F1-score and accuracy. Using all these ML model, we have to compare with better result to predicting the heart disease present or not. In these models, we have evaluated all result of accuracy is present as 92% LR, 94% of SVM, 61% of nearest neighbours, and 98% of DT classifier. All the predicted parameter of various classifier shows in Figure 3.

Table 2. Predicted result of various classifier models

Model	Precision	Recall	F1-score	Accuracy
LR	85	95	90	92
SVM	85	100	92	94
KNN	40	16	23	61
DT classifier	96	100	97	98

The selected ML algorithms like as DT classifier have generated the most accurate results and also having the highest values for the various evaluating parameters. So, the DT classifier will be used for better diagnosis of CHD.

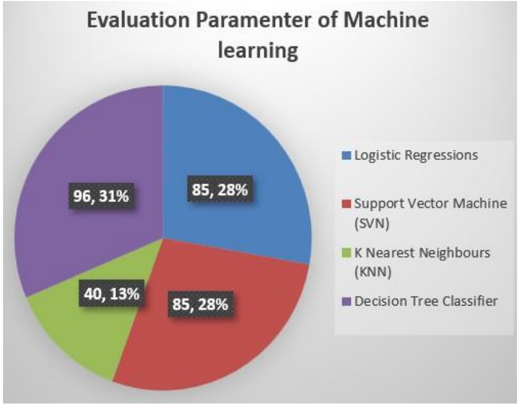

Figure 3. Predicted results of ML

5. DISCUSSIONS

The evaluation of four models, ML-CHDPM, DL, histogram gradient boosting, and our DT classifier, with comparative analysis, shows that there are considerable differences in how they perform when using the following key evaluation parameters, precision, recall, F1-score, and accuracy in given Table 3.

Table 3. Comparision table of various models

Models	Authors	Precision	Recall	F1-score	Accuracy
ML-CHDPM	Pachiyannan *et al*. [40]	87	96	91	94
DL	Eleyan *et al.* [41]	98	99	98	98
Histogram gradient boosting classifier	Sathi *et al.* [42]	88	89	89	90
DT classifier	Our's	96	100	97	98

The ML-CHDPM proposed by Pachiyannan *et al.* [40], has moderate performance as it has an accuracy of 94% and precision of 87%, recall of 96% and F1-score of 91. It has high recall and, therefore, a comparatively high false positive rate [40]. Histogram gradient boosting classifier by Sathi *et al.* [42] is a little bit more balanced though comes with 90% accuracy, 88% precision, 89% recall, and an F1-score of 89% [41]. Eleyan *et al.* [41] DL model presents their results as being top-notch, with almost a perfect 98-percent mark across all metrics with accuracy 98%, F1-score 98%, precision 98%, recall 99% [42]. Such high accuracy, however, is frequently at the expense of greater complexity, computation requirement and decreased interpretation.

Compared to this, our suggested DT classifier becomes an interesting alternative with an accuracy of 98%, precision of 96%, recall of 100%, and F1 accuracy of 97%. It means that not only the model has no false negatives (perfect recall) but the number of false positives is insignificant with high precision ratio. DT model is very simple, transparent and uses less computational overhead and therefore is highly applicable in clinical decision making environment. The DTs, in contrast to black-box DL models, introduce an extremely important layer of trust and usability to medical professionals, since the interpolatory nature of the model is not obscured by black-box functionality. In general, our model is not only highly predictive but also convenient in practice, and thus it becomes a valuable method of detecting cardiovascular diseases at an early stage of development.

6. CONCLUSION

In conclusion, the research of several ML algorithms for ECG data-based CHD prediction has yielded encouraging findings. Numerous algorithms have been investigated and their efficacy in correctly predicting CHD has been established, including LR, decision tress classifier, SVMs, KNN, and ensemble learning techniques. According to ML algorithms can offer accurate analysis and forecasting abilities when processing ECG data for CHD prediction. These algorithms have demonstrated good rates of precision, recall, and accuracy in detecting CHD cases. However, more research is necessary to address a number of problems. Hybrid models, which combine several algorithms or techniques, may increase prediction precision. By including clinical features and domain knowledge in the study, the accuracy and interpretability of the predictions may also be improved.

FUNDING INFORMATION

There is no any funding for this manuscript. The authors did not receive support from any organization for the submitted work.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Mohd Osama	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Rajesh Kumar		✓		✓	✓	✓		✓	✓	✓	✓	✓		
Chandrakant Kumar Singh				✓	✓					✓				

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

There is no Conflict of interest

DATA AVAILABILITY

Dataset used in this experiment are publicly available freely at Kaggle website with named "ECG-Dataset.csv"

REFERENCES

[1] V. V. Ramalingam, A. Dandapath, and M. K. Raja, "Heart disease prediction using machine learning techniques: a survey," *International Journal of Engineering and Technology(UAE)*, vol. 7, no. 2.8 Special Issue 8, pp. 684–687, 2018, doi: 10.14419/ijet.v7i2.8.10557.

[2] A. Lakshmanarao, Y. Swathi, and P. S. S. Sundareswar, "Machine learning techniques for heart disease prediction," *Forest*, vol. 95, no. 99, p. 97, 2019.

[3] H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, "Heart disease prediction using machine learning algorithms," *IOP Conference Series: Materials Science and Engineering*, vol. 1022, no. 1, 2021, doi: 10.1088/1757-899X/1022/1/012072.

[4] A. H. Gonsalves, F. Thabtah, R. M. A. Mohammad, and G. Singh, "Prediction of coronary heart disease using machine learning," in *Proceedings of the 2019 3rd International Conference on Deep Learning Technologies*, Jul. 2019, pp. 51–56, doi: 10.1145/3342999.3343015.

[5] D. Krishnani, A. Kumari, A. Dewangan, A. Singh, and N. S. Naik, "Prediction of coronary heart disease using supervised machine learning algorithms," in *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*, Oct. 2019, vol. 2019-Octob, pp. 367–372, doi: 10.1109/TENCON.2019.8929434.

[6] K. H. Miao, J. H. Miao, G. J. Miao, "Diagnosing coronary heart disease using ensemble machine learning," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 10, 2016, doi: 10.14569/IJACSA.2016.071004.

[7] D. J. I. Z. Chen and P. Hengjinda, "Early prediction of coronary artery disease (CAD) by machine learning method-a comparative study," *Journal of Artificial Intelligence and Capsule Networks*, vol. 3, no. 1, pp. 17–33, 2021.

[8] J. Zheng, J. Zhang, S. Danioko, H. Yao, H. Guo, and C. Rakovski, "A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients," *Scientific Data*, vol. 7, no. 1, p. 48, Feb. 2020, doi: 10.1038/s41597-020-0386-x.

[9] A. H. Chen, S. Y. Huang, P. S. Hong, C. H. Cheng, and E. J. Lin, "HDPS: heart disease prediction system," *Computing in Cardiology*, vol. 38, pp. 557–560, 2011.

[10] Purushottam, K. Saxena, and R. Sharma, "Efficient heart disease prediction system," *Procedia Computer Science*, vol. 85, pp. 962–969, 2016, doi: 10.1016/j.procs.2016.05.288.

[11] H. M. Le, T. D. Tran, and L. van Tran, "Automatic heart disease prediction using feature selection and data mining technique," *Journal of Computer Science and Cybernetics*, vol. 34, no. 1, pp. 33–48, 2018, doi: 10.15625/1813-9663/34/1/12665.

[12] G. Georgieva-Tsaneva, E. Gospodinova, and K. Cheshmedzhiev, "Examination of cardiac activity with ECG monitoring using heart rate variability methods," *Diagnostics*, vol. 14, no. 9, p. 926, Apr. 2024, doi: 10.3390/diagnostics14090926.

[13] P. Ingle, S. D. Kale, M. R. Shirole, and N. Shirsat, "Early heart attack detection using real-time ecg signals: a systematic survey," *Cureus Journal of Computer Science*, vol. 2, no. 1, Apr. 2025, doi: 10.7759/s44389-024-02662-6.

[14] M. Y. Ansari, M. Qaraqe, R. Righetti, E. Serpedin, and K. Qaraqe, "Enhancing ECG-based heart age: impact of acquisition parameters and generalization strategies for varying signal morphologies and corruptions," *Frontiers in Cardiovascular Medicine*, vol. 11, Jul. 2024, doi: 10.3389/fcvm.2024.1424585.

[15] K. Anderson, "The role of data preprocessing in machine learning accuracy for heart disease prediction hybrid models for heart disease prediction: combining neural networks with traditional," 2024.

[16] E. Asgarov, "A comprehensive analysis of machine learning techniques for heart disease prediction," *OALib*, vol. 11, no. 04, pp. 1–16, 2024, doi: 10.4236/oalib.1111490.

[17] S. I. Ayon, M. M. Islam, and M. R. Hossain, "Coronary artery heart disease prediction: a comparative study of computational intelligence techniques," *IETE Journal of Research*, vol. 68, no. 4, pp. 2488–2507, Jul. 2022, doi: 10.1080/03772063.2020.1713916.
[18] S. Saleem, A. H. Khandoker, M. Alkhodari, L. J. Hadjileontiadis, and H. F. Jelinek, "A two-step pre-processing tool to remove Gaussian and ectopic noise for heart rate variability analysis," *Scientific Reports*, vol. 12, no. 1, p. 18396, Nov. 2022, doi: 10.1038/s41598-022-21776-2.
[19] G. Varoquaux and O. Colliot, "Evaluating machine learning models and their diagnostic value," in *Neuromethods*, vol. 197, 2023, pp. 601–630.
[20] K. Nezamabadi, N. Sardaripour, B. Haghi, and M. Forouzanfar, "Unsupervised ECG analysis: a review," *IEEE Reviews in Biomedical Engineering*, vol. 16, pp. 208–224, 2023, doi: 10.1109/RBME.2022.3154893.
[21] J. Miao and W. Zhu, "Precision–recall curve (PRC) classification trees," *Evolutionary Intelligence*, vol. 15, no. 3, pp. 1545–1569, Sep. 2022, doi: 10.1007/s12065-021-00565-2.
[22] M. Edupuganti, V. Rathikarani, and K. Chaduvula, "Classification of heart diseases using fusion based learning approach," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 8s, pp. 570–580, 2024.
[23] A. S. Deokar and M. A. Pradhan, "Prediction of cardiovascular diseases using explainable AI," in *Artificial Intelligence, Blockchain, Computing and Security Volume 1*, vol. 1, London: CRC Press, 2023, pp. 201–206.
[24] P. Rani, R. Kumar, N. M. O. S. Ahmed, and A. Jain, "A decision support system for heart disease prediction based upon machine learning," *Journal of Reliable Intelligent Environments*, vol. 7, no. 3, pp. 263–275, Sep. 2021, doi: 10.1007/s40860-021-00133-6.
[25] F. Ali *et al.*, "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion," *Information Fusion*, vol. 63, pp. 208–222, Nov. 2020, doi: 10.1016/j.inffus.2020.06.008.
[26] R. Aggarwal, P. Podder, and A. Khamparia, "ECG classification and analysis for heart disease prediction using XAI-driven machine learning algorithms," in *Intelligent Systems Reference Library*, vol. 222, 2022, pp. 91–103.
[27] N. N. Anuar *et al.*, "Cardiovascular disease prediction from electrocardiogram by using machine learning," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 16, no. 07, pp. 34–48, Jun. 2020, doi: 10.3991/ijoe.v16i07.13569.
[28] H. Q. Ahmed, S. O. Amen, A. B. Q. Rassol, and I. I. Hamad, "Erbil heart disease dataset," *Kaggle*, 2022. .
[29] A. G, B. Ganesh, A. Ganesh, C. Srinivas, Dhanraj, and K. Mensinkal, "Logistic regression technique for prediction of cardiovascular disease," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 127–130, Jun. 2022, doi: 10.1016/j.gltp.2022.04.008.
[30] B. Duraisamy, R. Sunku, K. Selvaraj, V. V. R. Pilla, and M. Sanikala, "Heart disease prediction using support vector machine," *Multidisciplinary Science Journal*, vol. 6, p. 2024ss0104, Dec. 2023, doi: 10.31893/multiscience.2024ss0104.
[31] B. L. Deekshatulu, P. Chandra, and others, "Classification of heart disease using k-nearest neighbor and genetic algorithm," *Procedia Technology*, vol. 10, pp. 85–94, 2013.
[32] K. Rai, M. S. Devi, and A. Guleria, "Decision tree based algorithm for intrusion detection," *International Journal of Advanced Networking and Applications*, vol. 07, no. 04, pp. 2828–2834, 2016, [Online]. Available: https://www.researchgate.net/publication/298175900.
[33] G. Naidu, T. Zuva, and E. M. Sibanda, "A review of evaluation metrics in machine learning algorithms," in *Lecture Notes in Networks and Systems*, vol. 724 LNNS, 2023, pp. 15–25.
[34] A. Arias-Duart, E. Mariotti, D. Garcia-Gasulla, and J. M. Alonso-Moral, "A confusion matrix for evaluating feature attribution methods," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2023, vol. 2023-June, pp. 3709–3714, doi: 10.1109/CVPRW59228.2023.00380.
[35] O. Caelen, "A Bayesian interpretation of the confusion matrix," *Annals of Mathematics and Artificial Intelligence*, vol. 81, no. 3–4, pp. 429–450, Dec. 2017, doi: 10.1007/s10472-017-9564-8.
[36] K. Krasnodębska, W. Goch, J. H. Uhl, J. A. Verstegen, and M. Pesaresi, "Advancing precision, recall, f-score, and jaccard index: An approach for continuous gridded data," *Recall, F-Score, and Jaccard Index: An Approach for Continuous Gridded Data*, 2024, doi: 10.1016/j.envsoft.2025.106614.
[37] Jude Chukwura Obi, "A comparative study of several classification metrics and their performances on data," *World Journal of Advanced Engineering Technology and Sciences*, vol. 8, no. 1, pp. 308–314, Feb. 2023, doi: 10.30574/wjaets.2023.8.1.0054.
[38] G. M. Foody, "Challenges in the real world use of classification accuracy metrics: from recall and precision to the Matthews correlation coefficient," *PLOS ONE*, vol. 18, no. 10, p. e0291908, Oct. 2023, doi: 10.1371/journal.pone.0291908.
[39] D. Valero-Carreras, J. Alcaraz, and M. Landete, "Comparing two SVM models through different metrics based on the confusion matrix," *Computers & Operations Research*, vol. 152, p. 106131, Apr. 2023, doi: 10.1016/j.cor.2022.106131.
[40] P. Pachiyannan, M. Alsulami, D. Alsadie, A. K. J. Saudagar, M. AlKhathami, and R. C. Poonia, "A novel machine learning-based prediction method for early detection and diagnosis of congenital heart disease using ECG signal processing," *Technologies*, vol. 12, no. 1, p. 4, Jan. 2024, doi: 10.3390/technologies12010004.
[41] A. Eleyan, E. AlBoghbaish, A. AlShatti, A. AlSultan, and D. AlDarbi, "RHYTHMI: a deep learning-based mobile ECG device for heart disease prediction," *Applied System Innovation*, vol. 7, no. 5, p. 77, Aug. 2024, doi: 10.3390/asi7050077.
[42] T. A. Sathi *et al.*, "An interpretable electrocardiogram-based model for predicting arrhythmia and ischemia in cardiovascular disease," *Results in Engineering*, vol. 24, 2024, doi: 10.1016/j.rineng.2024.103381.

BIOGRAPHIES OF AUTHORS

Mr. Mohd Osama was born 12-Augest-1995. He is a resident from Prayagraj. He received his Schooling from Mother Suhag Inter College, after that he received his B.Sc. in Computer science from University of Allahabad in 2016, M.Sc. in Computer Science from J.K. Institute of Electronics and Information Technology, University of Allahabad at 2019 and Qualified UGC-NET in December 2020. He is Currently Persuing Ph.D. at Department of Electronics and Communication, University of Allahabad. His research interests: medical image processing, artificial intelligence, and machine learning. He can be contacted at email: mohdosama@allduniv.ac.in.

Dr. Rajesh Kumar is an assistant professor at the Department of Electronics and ommunication, University of Allahabad Prayagraj, India. He received an M.Tech. in software engineering from Motilal Nehru National Institute of Technology in Allahabad, India; a Ph.D. in computer engineering from the Indian Institute of Technology (BHU) in Varanasi, India. He has taught for around 12 years and conducted research for 8 years. Computer vision, image processing, and medical image analysis are among his areas of interest. He can be contacted at email: rajeshkumariitbhu@gmail.com.

Dr. Chandrakant Kumar Singh has received B.Sc., M.Sc. Physics, M.Sc. Computer sc, M.Tech. Software Engineering and Ph.D. degree from V.K. S University, Patna University, A.P.S. University Rewa, M.N.N.I.T. Allahabad and U.P. Rajarshi Tandon University Prayagraj respectively. He is currently working as an assistant professor in Computer Science Department of U.P. Rajarshi Tandon open university Prayagraj India Electronics and communication, University of Allahabad, India. He has more than 15 years of teaching experience and most challenging administrative experience in same university. He can be contacted at email: cksingh_apsu@hotmail.com.