

**How to Cite:**

Kalaivani, B., & Ranichitra, A. (2022). A comparative study of machine learning approaches for proactive cardiovascular disease prediction. *International Journal of Health Sciences*, 6(S8), 5390–5400. Retrieved from <https://sciencescholar.us/journal/index.php/ijhs/article/view/13462>

# **A comparative study of machine learning approaches for proactive cardiovascular disease prediction**

**B. Kalaivani**

Research Scholar, Department of Computer Science, Sri S.Ramasamy Nadiu Memorial College, Sattur, India  
Corresponding author email: [kalaivanisrnm2008@gmail.com](mailto:kalaivanisrnm2008@gmail.com)

**A. Ranichitra**

Assistant Professor, Department of Computer Science, Sri S.Ramasamy Nadiu Memorial College, Sattur, India  
Email: [ranichitra@srmcollege.ac.in](mailto:ranichitra@srmcollege.ac.in)

**Abstract**--The most common cause of death worldwide is cardiovascular disease (CVD), and its prediction is both difficult and crucial in the medical field. Bad eating habits, lack of exercise, cigarette and alcohol use, high blood pressure, elevated fasting, increased blood lipid levels, and being overweight or obese are the main behavioral causes of disease and stroke. Machine learning has emerged as an important tool in making healthcare decisions and predictions from healthcare information. This paper's goal is to evaluate and examine the risk factors for heart disease prediction as well as the different Machine Learning methods put out by different researchers for heart disease classification and prediction. This paper highlights and compares the algorithms and Methodologies used for early prediction of heart disease in order to discover the best technique to control the mortality rate in case of earlier detection of heart disease and adopting some preventive measures.

**Keywords**--cardiovascular disease (CVD), machine learning (ML), deep learning (DL), feature extraction.

**Introduction**

Cardiovascular disease (CVD) develops when the blood supply that supplies the coronary heart with blood and oxygen is completely stopped. This may also involve artery damage in organs like the kidneys, heart, vision, and central nervous system. In 2019, WHO says that 17.7 million individuals died from

coronary heart disease, accounting for 32% of all deaths worldwide; of these fatalities, 85% were attributable to heart attacks and strokes. Another factor is that middle class and baltic countries account for more than 75% of CVD-related fatalities. These figures suggest that there may be 23.6 million CVD patients worldwide in 2030 [1]. A variety of risk factors have been found, including increased blood pressure, smoking, blood cholesterol, diabetic, obese, genetic predisposition, and age, even if the precise causes of CVD is still unknown. A cornerstone is identifying those who are at risk of CVD. Due to the numerous limitations of manual detection of coronary heart disease, experts have shifted their focus to modern day techniques such as Data Mining, Machine Learning, and Deep Learning approaches for predicting the ailment, which will be an effective way to support in decision-making and predictions. As a result, in order to minimize the incidence of negative invasive coronary angiography tests, healthcare administrators are looking for accurate, reliable, non-invasive, and cost-effective ways for CVD diagnosis.

Complex measurements in medical applications can be effectively handled using machine learning. Modern classification approach provides disease prediction systems can be made more intelligent and effective. [2]. Various Machine Learning (ML) techniques are currently being utilized to predict the severity of cardiac disease. Healthcare professionals are attempting to harness this period of their field by employing constantly growing algorithms and providing information to machines that can assist them in imaging and inspecting human bodies for abnormalities. Machine learning has a variety of additional uses in a variety of sectors that aim to improve human existence. This paper aims to study and analyses the various research techniques and methodologies developed to predict the CVDs at the earlier stage. This publication reviewed the outcomes of state-of-the-art approaches and provides recommendations for further research.

### **Related Works**

Cardiovascular disease (CVD) is a group of conditions that impact the cardiovascular system (the heart and blood vessels). The objectives of the study are to determine the frequency and distribution of CVD risk factors, as well as the technologies employed in CVD prediction and the prevalence of high-risk diseases are discussed below and summarized in Appendix I. Mohan, S. et al. [3] designed a new Hybrid HRFLM approach which combines the traits of Random Forest and Linear method. The proposed model uses four different UCI repository dataset with 13 attributes to predict whether the patients are affected by the heart disease or not. The expectation model is made up of numerous combinations of features and a few well-known arrangement algorithms. The HRFLM technique was found to be quite accurate in predicting heart disease, with 88.7% of accuracy.

Khan, M.A., et al. [4] proposed the Wearable IoT enabled heart disease prediction system using a Modified Deep Convolution Neural Network model. Predicting cardiac disease was done with pre-processing, feature selection, and classification as the three stages. The proposed model uses 303 records with 14 attributes to predict whether the patients are affected by the heart disease or not. The feature selection was diagnosed using the Mapping-based Cuttlefish optimization

Algorithm (MCFA). For evaluate the system performance, the proposed In addition to logistic regression, MDCNN is examined with current deep learning neural networks. Finally, the MDCNN has a higher accuracy than other classifiers, with 98.2 %.

In order to analyse the Cleveland heart disease dataset, Kavitha, M. et al. [5] used a hybrid machine learning model that combines regression and classification techniques. The dataset is examined using the machine learning methods Random Forest, Decision Tree, and Hybrid model. The hybrid model produces more optimized outcomes and has with 88.7% accuracy. Praneetha, M. et al. [6] describes a web-based technique used for the prediction of Cardiovascular Disorder. This paper provides a unique technique for enhancing the accuracy of cardiovascular infirmity estimation by applying AI methods to find basic features. A variety of AI approaches are used to predict cardiovascular disease with high precision. The proposed model shows 79% of accuracy for Decision tree classifiers, 83% for SVM, 84% for Random Forest classifiers and 87% for KNN. Anyone can use wearable IoT devices to assess the accuracy of disease prediction. Finally, the paper concludes that KNN shows acceptable results for the considered dataset.

A DNN and Linear SVC integrated feature selection-based cardiac disease prediction technique was proposed by D. Zhang et al. [7]. The weight of the network is adjusted to avoid gradient vanishing or explosion, hence improving the predictor's performance. The Inter Quartile Range (IQR) approach is the best outlier reduction strategy for the Kaggle dataset. It extracted significant characteristics from the dataset using Linear Discriminant Analysis and Principal Component Analysis (PCA). Experiment findings suggest that a high-accuracy prediction model is viable. The suggested technique achieves 98.56 percent accuracy, 99.35 percent recall, 97.84 percent precision, an F1-Score of 0.983, and an AUC score of 0.983, proving the feasibility and usability of this deep neural network feature selection method.

Rahim, A. et al. [8] proposed a model for the early detection of cardiovascular disease based on several risk variables associated with an unhealthy lifestyle. Effective CVD prediction utilising a MaLCaDD is the subject of the proposed study. The system primarily addresses missing values and data imbalance. Following that, the Feature Significance approach is used to choose features. Finally, prediction is performed using an ensemble of Logistic Regression and KNN classifiers. The prediction accuracy for Framingham, Heart Disease, and Cleveland benchmark datasets is 99.1 percent, 98.0 percent, and 95.5 percent, respectively. Finally, a comparison of MaLCaDD predictions to existing approaches shows that proposed model's predictions are more accurate than existing approaches.

Guo, C. et al. [9] mentioned a Recursion Enhanced Random Forest with an Improved Linear Model to analyse the coronary heart disorder The proposed model of RERF-ILM technique combines a linear model and a random forest approach. In the IoMT platform, heart disease prediction Machine learning classification approaches comparisons of essential aspects are used. The suggested model was compared to the varied outputs of Global Classifier, Deep

Learning, Effective Heart Disease Prediction System, and Fast-Correlation-Based Selection method techniques using the UCI repository dataset with 13 attributes. Finally, RFRF-ILM predicts CAD heart disease with high accuracy and low classic error in the early stages.

Junejo, A.R.et.al.[10] used a dataset from Shaukat Khanum clinic and UCI research centre Molecular Diagnostics Laboratory at the University of California, Cardiovascular infection contains 294 records with 14 attributes. The Robust Healthcare Industry (RHI) assesses the precision of several AI grouping algorithms. The present order execution is the major focus, which is integrated with Deep Learning Techniques such as ANN, Naive Bayer, Decision Tree, and KNN. An ANN technique with intra-organ capacity and a variety of AI strategies are used to evaluate this model. The precision of the ANN is 98.4 percent, the precision of the K-nearest neighbor (KNN) is 98.01 percent, the precision of the Nave Bayes (NB) is 96.99 percent, and the precision of the Decision tree (DT) is 87.81 percent. Finally, ANN has a higher level of expertise than the others.

Pushpavathi.T.P.et al.[11] concentrate on designing and developing feature ranking for predicting the heart disease. This study employs the Kaggle dataset of 303 patients, in which only 6 records are incorrect, considering the 297 datasets for analysis. The proposed model uses three distinctive correlation techniques between the data followed by the applications of different AI techniques like Naïve Bayes, KNN utilizing sklearn packages with deep learning calculation. Random Forest has an accuracy of 81.6 %, K-nearest neighbor (KNN) precision of 55 %, Nave Bayes (NB) precision of 81 %, and Convolution Neural Network precision of 99.6%. Finally, CNN was quite accurate in its prediction of coronary disease.

Using various techniques, Nikam, A., et al. [12] were able to predict CVD based on specific features. BMI is one of the distinguishing characteristics used to predict heart disease. The key focus of this study is that it has two models, one containing BMI features and the other without. Childhood BMI is a risk factor for developing coronary heart disease in adulthood. The addition of the BMI feature enhanced prediction accuracy. As a result of analyzing results, it is obvious that the proposed method provides a more accurate forecast of cardiovascular disorders. The Decision tree algorithm is more efficient and tested with the maximum accuracy. The XGB classifier was used to determine the importance of each component in the prognosis of heart disease.

Samir, A.et al. [13] proposed an Heart disease prediction using an evolutionary method based on a convolutional neural network. CNNs had great success creating health support systems because of their capacity to recognize connections and hidden patterns in healthcare data. The unique method predicts heart disease using CNN –jSO and compares it to existing algorithms utilizing datasets from the kaggle heartbeat sound system and physioNet heart sound dataset. CNN-jSO techniques for cardiac disease prediction, in which the jSO optimization technique is used to optimize certain hyper parameters. Using Python, this model achieves 97.76 percent accuracy for the training dataset and 94.12 percent accuracy for the test dataset. Finally, the article indicates that the CNN-jSO strategy outperforms the other.

Talasila, V. et al.[14] has developed a disease prediction system based on rough set theory and a RNN. The Rough Set Theory is used to identify the most important characteristics, which aids in medical data categorization and illness diagnosis. For illness prediction, the selected characteristics are fed into the Recurrent Neural Network (RNN) approach. This RST-RNN combination approach employs Python to categories five unique datasets. The existing BPA-NB strategy is less accurate on the same heart disease dataset than the recommended RST-RNN method, which uses only relevant information for successful classification. On the same heart disease dataset, the current Big Data Predictive Analytics-Naive Bayes (BPA-NB) technique is less accurate than the suggested RST-RNN method, and feature selection is critical for effective classification. RST-RNN has been demonstrated to be highly accurate when it comes to accurately predicting cardiac disease, with a computed accuracy of 98.57 percent.

Tamal, M.A., Islam, M.S., Ahmmed, M.J., Aziz, M.A., Miah, P., and Rezaul, K.M. et al.[15] investigated Heart Disease Prediction based on a range of peripheral factors was examined. Extrinsic factors that cause heart disease are investigated using machine learning techniques such as Decision Tree, Random Forest, Naive Bayes, SVM, Quadratic Discriminant, and Logistic Regression. Python classification to detect cardiac issues early on. It extracts key information to increase accuracy and controls over-fitting. The suggested technique is a machine learning-based heart disease prediction model that can predict heart disease with 95% accuracy

Nadakinamani, R.G., et al.[16]. Used the Weka tool, different classification REP Tree, M5P Tree, Random Tree, Linear Regression, Naïve Bayes, J48, and JRIP were analyzed using different datasets Hungarian and Statlog (heart).Recent algorithms, on the other hand, analyse health data using the Mean Absolute Error (MAE), Root Mean Squared Error (RSME), Accuracy, and Time to analyse health information. The Cardiovascular Disease Prediction System (CDPS) analyses medical information to provide reliable forecasts for expert advice. The objective is to make increase the predictability of cardiovascular disease. According to the results the Random Tree performed extraordinarily well in predicting cardiovascular disease in sick individuals, with a high accuracy of 99.81 percent. The quickest prediction time was 0.01 seconds, with an Mean Absolute Error of 0.0011, an Root Mean Squared Error of 0.0231, and a fastest MAE of 0.0011.

Mortality is another term for death. The mortality rate is calculated as the number of fatalities brought on by an illness divided by the total population. Asian Indians have a death rate from coronary artery disease that is 20–50% greater than the general population. Compared to the global average of 235, India has a CVD death rate of 272 per 100,000 people, according to WHO estimates from the Global Burden of Disease research. For comparative epidemiological investigations of death by cause, the WHO mortality database is the most often used data source. The visualization website states that ischemic heart disease and stroke, which account for more than 80% of all fatalities, are the main causes of CVD death. The names of the diseases and their overall death rates are summarized in Table 2 and Figure 1. [17].

Table 2: Name of the Disease

S.No	Name of the Disease	Number of Death rate
1	Ischemic heart disease	1520034
2	Chronic obstructive pulmonary disease	879732.4
3	Stroke	697038.1
4	Diarrhoeal disease	670314.5
5	Tuberculosis	435881.7
6	Neonatal conditions	434148.3
7	Lower respiratory infections	392424.8
8	Diabetes mellitus	271495.3
9	Cirrhosis of the liver	268579.9
10	Falls	225755.2
11	Kidney diseases	222902.7
12	Road injury	212595.9
13	Asthma	196050.7
14	Self-harm	173347.3
15	Mouth and oropharynx cancers	134186.8
16	Alzheimer disease and other dementias	126593
17	Rheumatic heart disease	126079
18	Hypertensive Heart Disease	108050.1
19	Trachea, bronchus, lung cancers	88275.78

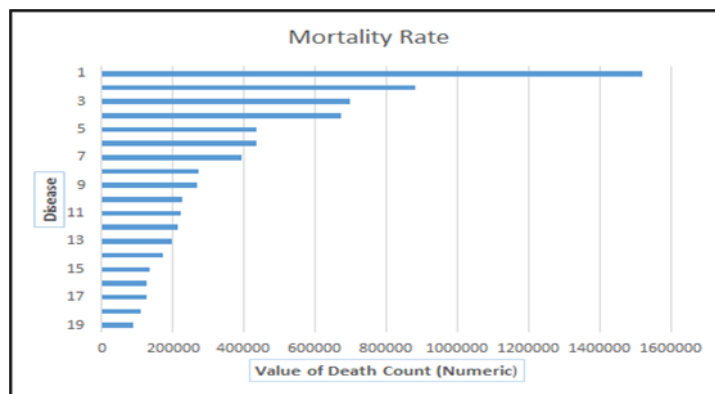


Figure 1: Total Number of death count vs diseases.

The cardiovascular diseases Population Attributable Risk or Fraction for blood pressure were calculated to be 40.6%. Among risk factors, blood pressure was the most critical factor for the cause of cardiovascular disease. The risk factor for high blood pressure had a more population attributable fraction than other physiological factors in the development of cardiovascular disease. Thus, by implementing comprehensive health policies, promoting healthy lifestyles, screening for and identifying relevant cases, and carrying out health promotion activities, these diseases can be prevented [18].

Table 3: Mortality Rate for CVD Components vs Population Attributable Fraction

CVD Components	
High Blood Pressure	40.60%
SMOKING	13.70%
POOR DIET	13.20%
Abnormal Glucose level	11.90%
Pulmonary Artesia	8.80%

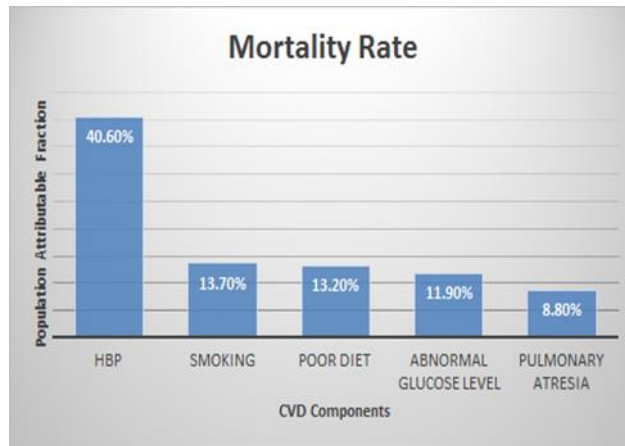


Figure 2: Mortality Rate for CVD Components vs Population Attributable Fraction

Cardiovascular disease is a risk factor for both men and women. Men were found to be roughly twice as likely as women to have a heart attack. Men, on average, have a higher risk of CVD and are affected at an earlier age than women. High cholesterol, high blood pressure, diabetes, BMI, and physical activity are all associated with a higher risk of heart disease in both men and women. Over one crore annual deaths are reported in India 2021 and 20.3% men and 16.9% of women were died due to CVD. According WHO, Gender-wise survey in 2021 shows that CVD disease has a high fatality rate when compared to other diseases. For all diseases, male mortality is at a higher rate than female. The gender-wise mortality rate for the year 2021 in shown in Figure 3.

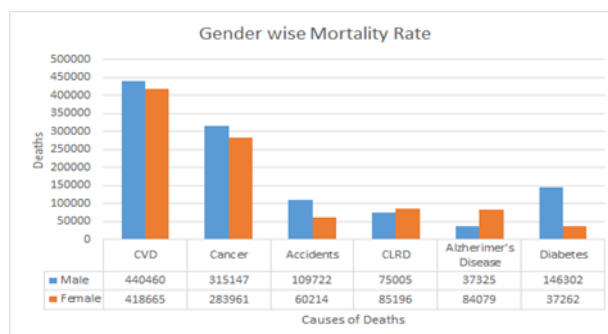


Figure3: Gender Wise Mortality Rate year (2021) [19].

## Conclusion

Cardiovascular disease (CVD) is a group of disorders affecting the circulatory system. Risk assessment is crucial for early diagnosis and treatment. The risk assessment is used to anticipate the total severity of future cardiovascular disease mortality and to suggest further therapy to save a person's life. Deep learning and Artificial intelligence have produced some remarkable outcomes in several fields with high accuracy in medical diagnosis. Hence, the various Machine Learning techniques proposed by the academicians for classification and prediction in diagnosing heart disease were studied and analyzed in this work. This article highlights the algorithms and strategies used for early prediction of heart disease and compares them to discover the best technique. Every suggested model produces different outcomes in different conditions for different datasets. It was also focused on the accuracy attained in detecting heart disease earlier. In future, an approach for predicting heart disease with high accuracy by minimizing expense and complexity will be proposed for earlier prediction of heart diseases.

## References

1. Benjamin, E.J., Muntner, P., Alonso, A., Bittencourt, M.S., Callaway, C.W., Carson, A.P., Chamberlain, A.M., Chang, A.R., Cheng, S., Das, S.R. and Delling, F.N., 2019. Heart disease and stroke statistics—2019 update: a report from the American Heart Association. *Circulation*, 139(10), pp.e56-e528.
2. Guo, C., Zhang, J., Liu, Y., Xie, Y., Han, Z. and Yu, J., 2020. Recursion enhanced random forest with an improved linear model (RERF-ILM) for heart disease detection on the internet of medical things platform. *IEEE Access*, 8, pp.59247-59256.
3. Institute of Health Metrics and Evaluation. GBD Compare 2010. <http://vizhub.healthdata.org/gbd-compare/>. Accessed April 30, 2014
4. Junejo, A.R., Shen, Y., Laghari, A.A., Zhang, X. and Luo, H., 2019. Molecular diagnostic and using deep learning techniques for predict functional recovery of patients treated of cardiovascular disease. *IEEE Access*, 7, pp.120315-120325.
5. Kavitha, M., Gnaneswar, G., Dinesh, R., Sai, Y.R. and Suraj, R.S., 2021, January. Heart Disease Prediction using Hybrid machine Learning Model. In 2021 6th International Conference on Inventive Computation Technologies (ICICT) (pp. 1329-1333). IEEE.
6. Khan, M.A., 2020. An IoT framework for heart disease prediction based on MDCNN classifier. *IEEE Access*, 8, pp.34717-34727.
7. M. Chen, Y. Hao, K. Hwang, L. Wang and L. Wang, "Disease Prediction by Machine Learning Over Big Data From Healthcare Communities," in *IEEE Access*, vol. 5, pp. 8869-8879, 2017, doi: 10.1109/ACCESS.2017.2694446.
8. Mohan, S., Thirumalai, C. and Srivastava, G., 2019. Effective heart disease prediction using hybrid machine learning techniques. *IEEE access*, 7, pp.81542-81554.
9. Nadakinamani, R.G., Reyana, A., Kautish, S., Vibith, A.S., Gupta, Y., Abdelwahab, S.F. and Mohamed, A.W., 2022. Clinical Data Analysis for Prediction of Cardiovascular Disease Using Machine Learning Techniques. *Computational Intelligence and Neuroscience*, 2022.

10. Nikam, A., Bhandari, S., Mhaske, A. and Mantri, S., 2020, December. Cardiovascular Disease Prediction Using Machine Learning Models. In 2020 IEEE Pune Section International Conference (PuneCon) (pp. 22-27). IEEE.
11. Praneetha, M., Jesudoss, A. and Mayan, A., 2021, May. Cardiovascular Disorder Prediction using Machine Learning. In 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 1665-1670). IEEE.
12. Rahim, A., Rasheed, Y., Azam, F., Anwar, M.W., Rahim, M.A. and Muzaffar, A.W., 2021. An Integrated Machine Learning Framework for Effective Prediction of Cardiovascular Diseases. *IEEE Access*, 9, pp.106575-106588.
13. S. Pouriye, S. Vahid, G. Sannino, G. De Pietro, H. Arabnia and J. Gutierrez, "A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease," 2017 IEEE Symposium on Computers and Communications (ISCC), 2017, pp. 204-207, doi: 10.1109/ISCC.2017.8024530.
14. Samir, A.A., Rashwan, A.R., Sallam, K.M., Chakraborty, R.K., Ryan, M.J. and Abohan, A.A., 2021. Evolutionary algorithm-based convolutional neural network for predicting heart diseases. *Computers & Industrial Engineering*, 161, p.107651.
15. T. P. Pushpavathi, S. Kumari and N. K. Kubra, "Heart Failure Prediction by Feature Ranking Analysis in Machine Learning," 2021 6th International Conference on Inventive Computation Technologies (ICICT), 2021, pp. 915-923, doi: 10.1109/ICICT50816.2021.9358733.
16. Talasila, V., Madhubabu, K., Madhubabu, K., Mahadasyam, M., Atchala, N. and Kande, L., 2020. The prediction of diseases using rough set theory with recurrent neural network in big data analytics. *International Journal of Intelligent Engineering and Systems*, 13(5), pp.10-18.
17. Tamal, M.A., Islam, M.S., Ahmmed, M.J., Aziz, M.A. and Pabel Miah, K.M.R., 2019. Heart disease prediction based on external factors: A machine learning approach. *Int J Adv Com Sci Appl (IJACSA)*, 10.
18. WHO-World Health Organization; Geneva: 2016. *Global Health Estimates 2015: Deaths by Cause, Age, Sex, by Country and by Region, 2000–2015*
19. Zhang, D., Chen, Y., Chen, Y., Ye, S., Cai, W., Jiang, J., Xu, Y., Zheng, G. and Chen, M., 2021. Heart Disease Prediction Based on the Embedded Feature Selection Method and Deep Neural Network. *Journal of Healthcare Engineering*, 2021.

## Appendix I

Table 1: Comparative Analysis of the various Techniques and Methodologies applied in early prediction of CVD

Ref.	Pre-processing	Feature selection Selected/ Total Attributes	Machine Learning Algorithms / Proposed Techniques	Methods	Performance Metrics	Percentage	Main objectives and Contribution in the paper
[3]	Yes 1.Feature selection 2.Classification modeling 3.Performance measure	13/13	Hybrid Random Forest Linear Model	Classification -Hybrid Machine Learning Technique	Accuracy Sensitivity Specificity Precision F-Measure	88.7% 92.8% 82.6% 87.5% 90%	<ol style="list-style-type: none"> <li>1. Identified the best ML algorithms based on the less error rate.</li> <li>2. Classified the data based on the high Accuracy using the identified the best algorithm.</li> <li>3. Performance comparison with different model</li> <li>4. It shows low computation time to access and has low accuracy</li> </ol>
[4]	Yes 1. Missing values replacement 2. Redundant values Elimination. 3. Separation	13/14 Mapping-based cuttlefish optimization algorithm (MCFA)	Modified Deep Convolutional Neural Network	Classification	Accuracy Error Precision F1 Recall Specification	98.2% 1.8 95.1 95 97.8 92.6	<ol style="list-style-type: none"> <li>1. Predicting cardiac disease was done with pre-processing, feature selection, and classification as the stages.</li> <li>2. The feature selection was diagnosed using the Mapping-based Cuttlefish Optimization Algorithm. The proposed MDCNN is compared to existing deep learning neural networks in addition to logistic regression to evaluate the system's performance.</li> <li>3. The proposed algorithm, MDCNN shows greater accuracy compared to other algorithms classifiers.</li> </ol>
[5]	No	14	Decision Tree, Random-Forest, and Hybrid model	Classification	Accuracy	88%	<ol style="list-style-type: none"> <li>1. Data mining approach using Cleveland heart disease dataset such as regression and classification procedures.</li> <li>2. The Machine Learning (ML) algorithms Random Forest, Decision Tree and Hybrid model are used to analyze the dataset.</li> <li>3. The hybrid model produces more optimized results and has an accuracy rate of 88.7%.</li> </ol>
[6]	Yes 1.Data Cleaning missing data information	13	Choice Tree, K-Nearest Neighbor, K-means gathering, Adaboost.	Classification	Accuracy Decision Trees Support Vector Random Forest K-Neighbor	79% 83% 84% 87%	<ol style="list-style-type: none"> <li>1. A novel system that aims to find basic features using AI methods to enhance the accuracy of cardiovascular infirmity estimation.</li> <li>2. Wearable IoT devices is used to predict the accuracy of disease prediction.</li> <li>3. Finally, KNN shows acceptable results.</li> </ol>
[7]	Yes 1. Outlier removal using InterQuartile range (IQR) method 2. Data Standardization Process	14 Embedded process (Linear SVC method)	Linear Support Vector Classifier +Deep Neural Network	Classification	Accuracy Recall Precision F1-Score AUC Score	98.56% 99.53% 97.84% 98.3% 98.3%	<ol style="list-style-type: none"> <li>1. Classification and feature selection method to predict heart disease using an ensemble method.</li> <li>2. For the Kaggle dataset, the Inter Quartile Range (IQR) method is the best outlier elimination technique. It uses two method to extract key features from the dataset: Linear Discriminant Analysis and Principal Component Analysis.</li> <li>3. Experiments suggest that employing a deep neural network, a high-accuracy heart disease prediction model can be developed.</li> </ol>
[8]	Yes 1. Mean Replacement techniques 2. SMOTE	Yes 5/14	MalCaDD	Classification Ensemble boosting Ensemble Logistic Regression, K-NN	Accuracy Random Forest SVM Logistic Regression K-Nearest Neighbor Ensemble	82% 79% 94.2% 74.3% 99.1%	<ol style="list-style-type: none"> <li>1. The missing value and data imbalance methods are used to identify missing values.</li> <li>2. The technique of Feature Importance is used for feature selection.</li> <li>3. Evaluation of several machine learning approaches on various datasets.</li> <li>4. For high-accuracy prediction, an ensemble of Logistic Regression and K-NN classifiers are proposed.</li> </ol>
[9]	Yes 1. The clustering of datasets is based on Decision Trees.	14	Recursion Enhance Random Forest with an improved linear model	Classification	Accuracy Support Vector Machine	96.8%	<ol style="list-style-type: none"> <li>1. A model was compared with the various outputs of Global Classifier, Deep Learning, Effective Heart Disease Prediction System, and Fast-Correlation-Based selection method techniques using the UCI repository dataset</li> <li>2. RFRF-ILM predicts CAD heart disease with high accuracy and low classic error in the early stages.</li> </ol>
[10]	Yes 1. Designing ANN	Yes	Decision Tree, Naïve Bayes K-NN, ANN	Classification	Accuracy ANN	98.4%	<ol style="list-style-type: none"> <li>1. A grouping AI strategy measures the exactness.</li> <li>2. The current order execution is the major focus, which is integrated with Deep Learning Techniques</li> </ol>

	2. Back propagation 3. learning techniques 4. Feature selection 5. classification				K-NN Naïve Bayes Decision Tree	98.01% 96.99% 87.81%	such as ANN Nave Bayer, Decision Tree, and K-NN. Finally, ANN is more proficient than the others.
[11]	Yes	Yes 303 records	K-NN, Naïve Bayes, Random Forest, Conventional Neural Network	Classification	Accuracy K-NN Naïve Bayes Random-Forest Conventional Neural Network	55% 81% 81.6% 99.6%	1. Identify the suitable feature ranking method. 2. To identify the important feature by machine learning techniques such as KNN, Naive Baye, CNN and deep CNN 3. CNN shows the best results with an accuracy of 99.6% compare with the other ML techniques.
[12]	Yes 1. Objective 2. Examination 3. Subjective (BMI)	Yes 12 /14	Decision Tree, Light Gradient Boosting Machine, Extreme Gradient Boosting, Neural-Network, Logistic-Regression, K-NN, Naïve Bayes	Classification	Accuracy Decision-Tree Light Gradient Boosting Machine Extreme Gradient Boosting Neural Network Logistic Regression K-Nearest Neighbor Naïve Bayes	72.68% 80.12% 83.52% 71.85% 71.65% 72.13% 70.27%	1. BMI is one of the distinguishing features used to predict heart disease. 2. The decision tree technique has proven to be more efficient and accurate
[13]	No	Yes Audio processing (ISO17 hyper parameters )	Decision Tree, Random Forest, K-Nearest Neighbor	Classification	Accuracy CNN-iSQ Train Test	97.76% 94.12%	1. Trained the model to extract valuable information for decision-making was cleaned up in a data preparation stage. 2. The suggested model's hyper parameter was tuned using the iSQ technique. 3. Reformulation of the loss function to improve the model's accuracy, which now stands at 97.76 percent for training and 94.12 percent for testing. 4. The CNN-iSQ technique produced the highest accurate prediction results across all metrics.
[14]	No	13	RST-RNN Rough Set Theory (RST) Recurrent Neural Network (RNN)	Classification five different dataset	Accuracy AUC Sensitivity Specificity F-Measure	98.57% 89.58% 98.24% 98.64% 95.47%	1. Big Data Predictive Analytics- Naive Bayes (BPANB) approach is less accurate than the proposed RST-RNN method on the same heart disease dataset, and feature selection is crucial for good classification.
[15]	Yes 1. Incomplete & inconsistent information removal.	12/14	Decision Tree, Random Forest, Naive Bayes, Support Vector Machine, Quadratic Discriminant, and Logistic Regression	Classification	Accuracy Naive Bayes Quadratic Discriminant Analysis Logistic-Regression SVM Decision-Tree Random-Forest	92% 94% 94% 95% 91% 92%	1. Using Machine Learning and extrinsic factors to detect cardiac disease at an early stage. 2. Providing services that take less time in exchange for a more accurate diagnosis of cardiac disease.
[16]	Yes 1. The data contains a high number of missing values that have been deleted, as well as noisy values.	14 /14 303 Records	Random Forest, REP Tree (Reduced Error Pruning), JRIP_148, Naïve Bayes	Classification	Accuracy Random Forest JRIP J48	100% 99.9% 99.9%	1. The accuracy of the prediction is tested on two separate datasets. 2. The output of a decision tree is computed using a random subset of features. 3. The Random Tree model performed extremely well in predicting cardiovascular disease in sick individuals, with a maximum accuracy of 100 percent.