# A Comparative Study of Heart Disease Prediction using Machine Learning

Sreekumari S [1], Rajni Bhalla [1] and Geetha Ganesan [2]

[1] *Lovely Professional University, Phagwara, Punjab, India*
[2] *Jain (Deemed-to-be) University, Bengaluru, India*

**Abstract**

The heart, as the primary organ responsible for circulating blood throughout the body, is of great concern in society due to heart disease. Diagnosing heart disease presents significant challenges for doctors, considering its diverse types. Multiple prediction methods have been explored by researchers to address this issue. The accuracy of these predictions remains a key consideration. In this study, we focused on five different machine learning algorithms, including Random Forest, Decision Tree Classifier, Logistic Regression, K-Nearest Neighbor Classifier, and Decision Tree Classifier with Grid search. Additionally, we developed an ensemble model with the primary objective of accurately predicting heart disease. Our analysis utilized a heart disease dataset from Kaggle, and among the five algorithms examined, the Decision Tree Classifier achieved the highest accuracy of 92%. This finding highlights its effectiveness in predicting heart disease.

**Keywords**

Ensemble model, Prediction, machine learning algorithms.

## 1. Introduction

The heart is the primary organ that sends blood to entire parts of our body. This pumped blood carries oxygen and nutrients. The embryology of cardiovascular disease has now become an unsolved problem. Due to increasing stress and change in lifestyle the number of people getting affected by cardiovascular diseases is growing enormously day by day. Earlier heart disease prediction is very strenuous. Because it depends upon several factors. There are a variety of features that cause heart disease such as Hypertension, cholesterol, abnormal pulse rate, etc. The symptoms also may vary for different genders. For example: if females have symptoms such as Chest pain with chest discomfort, men may have only chest pain. One American dies every 34 seconds due to cardiovascular diseases.

There are diverse Types of cardiovascular diseases. Each has distinct types of symptoms and various causes. Diverse types of CVD are cardiac arrhythmia, Aorta disease, coronary Thrombosis, ischemic heart disease, angina pectoris, cardiac infarction, myocardial infarction, stroke, etc. If it is possible to predict heart disease prematurely, it will save several lives.

CAD is a condition that narrows or blocks the arteries that supply blood to our heart due to plaque buildup. This decreased blood flow may lead to chest pain and dyspnea, and a complete impasse of blood flow may cause a heart attack. The most widespread symptom of CAD is chest discomfort.

CVMLH 2023: Workshop on Computer Vision and Machine Learning for Healthcare, May 20 – 21, 2023, Chennai, India.
EMAIL: sreekumari.sreelesh@gmail.com (Sreekumari S)
ORCID: 0000-0002-0207-4738 (Sreekumari S)

Following a heart-healthy lifestyle can stave off one's life from CAD. In this work, we are going to use machine learning techniques to predict heart disease.

## 1.1. Machine Learning

Artificial Intelligence is a sub-part of Machine Learning, in which machines are learning things from past data like human beings learn from experiences. Nowadays, it is used commonly in the medical field, computer - vision, speech recognition, etc. Supervised Learning, Unsupervised Learning, Semi-supervised Learning, and Reinforcement Learning are four kinds of machine learning.

### 1.1.1. Supervised Learning

It is a kind of Machine Learning, where we are providing a labeled dataset to train the model, and based on this it predicts the output. In Supervised learning, the system creates a model by using the labeled data and will test the model by providing sample data to check whether it predicts accurately or not. Classification and Regression are the two classifications of supervised learning.

### 1.1.2. Unsupervised Learning

Here, we are providing an unlabeled dataset. In unsupervised learning, training is provided without any supervision. The two classifications of Unsupervised learning are Clustering and Association.

### 1.1.3. Semi-Supervised Learning

Here, we are providing both labeled as well as unlabeled datasets.

### 1.1.4. Reinforcement Learning

Reinforcement learning (RL) is a branch of machine learning that focuses on an agent's learning process through interactions with an environment. It involves an agent learning optimal actions by receiving rewards or penalties based on its actions. RL has gained significant attention in research due to its ability to tackle sequential decision-making problems. It has been applied in various domains, such as robotics, game playing, and autonomous systems. RL algorithms, such as Q-learning and Deep Q-Networks (DQN), have shown promising results in training agents to make intelligent decisions in dynamic and uncertain environments. It is feedback-based learning. It learns from experience and the feedback it tries to improve.

## 2. Related Work

Heart disease prediction is extremely critical. A plethora of studies are conducted in this area. Different methods are already used. Some are using different machine learning algorithms; some are using deep learning, and some are using data mining techniques. The selection of features is especially important.

In [1] they used CT and CCTA scans as input and then they applied ML algorithms like DNN, SVM, and Random Forest. They have concluded by stating that if it is possible to use AI-based Cardiovascular

images then it will be helpful for the prediction. Authors in [3] stated that this was the first paper that uses a modern technology called Optimally Time- Frequency Concentrated (OTFC) Even-length Bio-orthogonal wavelet Filter Bank (BWFB). They have concluded that they need more CAD ECG data sets for better prediction. Authors in [4], have used the CAD dataset and applied ten different machine learning algorithms such as C-SVC, nu-SVC, Logistic regression, Naïve Bayes, etc. Applied Genetic Algorithm and Particle Swarm Optimization. They eliminated the sparse set, but still, they noticed that there is no effect on the accuracy. In [7], the authors explained very well about invasive and non-invasive methods used in CAD detection. They have used another method called Gaussian Mixture Model (GMM) classifier. Different classifiers were used in this. In this work, they have used six classifiers Decision Tree, Fuzzy Sugeno, GMM, KNN, RBPNN, and Naïve Bayes Classifier. They have introduced a new parameter, the Heart index which is a 16-digit number getting after combining all the feature's values. For normal subjects, the heat index is $2.52 \pm 0.07$, and for CAD subjects $2.86 \pm 0.11$. They concluded by stating that they have 100% accuracy, which can be developed as software and can be implemented on mobile phones. Different methods used for prediction are well explained in [5], [6] and [19] and about ANN in [21] as well.

In [9], they used SVM for Classification and Regression tasks. They used Principal Component Analysis for identifying patterns in such a way that it should highlight all the similarities and differences as well. They have implemented this in Matlab 7.0 and tried the SVM method for 23 features. But they got less accuracy. Then using PCA they reduced the features to 18. Then it shows an accuracy of 79.17%. In [2] they used SVM-RFE to reduce the dimension to find the optimal feature subset. In [10], they compared various CAD prediction methods. In this study, they have used images for prediction and two different training models such as Naïve-Bayes Classifier and SVM. They concluded by stating that SVM is better for prediction and that a database needs to be updated with more descriptions of patients.

In [12] they used the Weighted Associative Rule Mining method to predict heart disease. In this model, weights were assigned to unique features, and calculated the total score obtained for features was. Then they applied WARM using Apriori Algorithm and selected the best features. When compared with the previous models, this model got a 98% Confidence score. Vardhan S et al. [13] in 2021, used the CVD dataset from Kaggle. Along with traditional classifiers, they applied some ensemble techniques such as boosting, bagging, and stacking. The bagged model shows 1.96% more accuracy, the bagged model shows an accuracy of 73.4%, and Boosted model shows an accuracy of 75.1%.

In [14], a new method called Hybrid Random Forest with Linear Model (HRFLM) got introduced. DT, Language model, SVM, RF, Naïve Bayes, neural network, and K nearest neighbor were the six classifiers used for this study. They got an accuracy of 0.8847. In addition to these basic methods AdaBoost M1 and MLP were used in [11]. Sabrina m et al., [15] used two data sets, an Italian data set, and an American dataset. They have applied several scaling methods such as normalization, Min Max scalar, and standardization. LR, KNN, classification decision tree (CART), NB, linear svc, and support vector classifier with radial basis function Kernel were the different Machine learning algorithms. They got more accuracy for SVC with RBF and Grid search algorithm.

Sumit S et al. in [16] created a deep-learning Neural Network for heart disease prediction. Also used a new optimization technique called Talos hyperparameter. The different learning algorithms used were LR, KNN, SVM, NB, and Talos. Talos follows prepare, optimize, and deploy process flow. The last process is the evaluation. Compared to other methods, it shows an accuracy of 90.78 percent. S M M Hasan et al.[17] proposed a study with basic machine learning methods and LR got more accuracy and in [28] KNN got 90.79% accuracy. In [18], they have applied enhancement methods such as AdaBoost, Bagging, and Boosting along with LMT and Hoeffding Tree Techniques. For LMT they selected AdaBoost with 80.32% accuracy and for Hoeffding, got 81.96% accuracy with Bagging Technique.

Jaymin Patel et al., [20], To extract hidden patterns they used data mining techniques. They used Waikato Environment for Knowledge Analysis (WEKA) Tool and the classification Tree Algorithms used were J48 with reduced error pruning, Logistic model Tree Algorithm, and Random Forest Algorithm. The J48 algorithm shows the highest accuracy, and it is 56.76% and LMT shows the lowest accuracy at 55.77%. Their accuracy was less compared to other models. In [22], they used the KEEL tool for feature selection, and for feature extraction, they used PCA under the WEKA tool. Under WEKA Tool, 10-fold cross-validation majority voting got an accuracy of 80.20, and AdaBoostM1 without an ensemble model got an accuracy of 80.01. Under KEEL tool GFS- LogitBoost-c got an accuracy of 80.53. Bagging and Boosting ensemble models were applied in [23].

In [26], they used the Phonocardiogram heart sound dataset. They applied PCG augmentation, then samples were doubled. On all signals, amplitude normalization was performed. CNN was used and this study shows an accuracy of 98.60. In [8], they proposed voting ANN. CNN is used in [25]. In [27], they used a new methodology called Swarm ANN on the UCI repository and they got an accuracy of .9578. Abidishaq et al., [29], applied one oversampling method called SMOTE (Synthetic Minority Oversampling Technique). Among 9 classifiers, they got more accuracy for RF, and it was .8889. With SMOTE, it shows 10% more accuracy. Extra Tree Classifier shows an accuracy of .9262.

Random Forest algorithm on spark framework is applied in [30] and they got an accuracy of 98%. Basic machine learning models were applied in [31] and SVM got more accuracy, and it was 94.60% and boosted SVM got more accuracy in [33]. ANN was created in [32] with a performance of 90%. In [36] presents a detailed examination of the effectiveness and performance of both proposed and existing methodologies for classification task. The study aims to compare and evaluate the advantages and limitations of different methodologies for collecting and analyzing data for structured data. By conducting a comparative analysis, the paper provides valuable insights for researchers and practitioners seeking to optimize their methodologies for classification tasks and for data that is in structured format. Liu J et al in [34], proposed a stacking model. They applied different methods such as LR, RF, ET, MLP and CatBoost. And for stacking model they got an accuracy of 84.62%. In [35] they applied RB-Bayes, NB and SVM and got more accuracy for SVM and it was 85.71%.

In [37] concludes by identifying the current research trends and challenges in sentiment classification using hybrid ensemble-based approaches. It highlights the need for further investigation into the combination of diverse classifiers and the integration of deep learning techniques in ensemble frameworks. The hybrid ensemble approach can be applied to classification dataset like heart disease detection to achieve better accuracy. In [38] created a review of the application of data mining in heart disease prediction. In this, they have compared different data mining algorithms and concluded like heart disease prediction with data mining will become most successful with a smaller number of attributes, and text mining the medical data needs to be extended in predicting the health care data.

## 3. Proposed Methodology

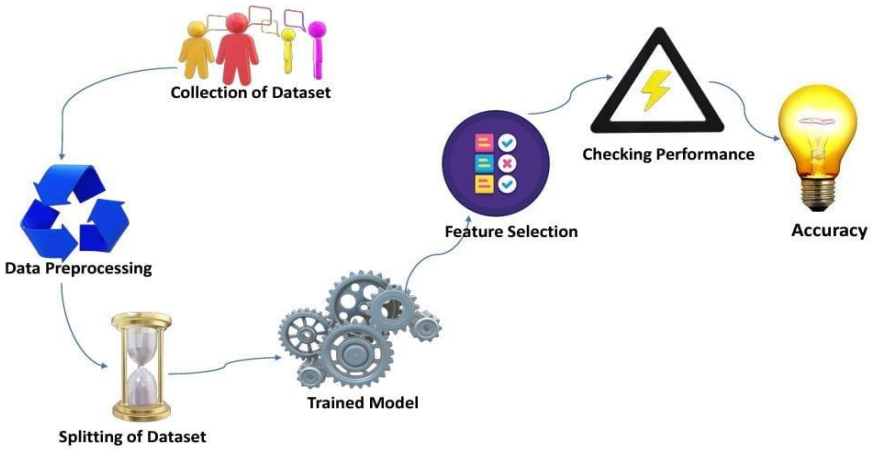



**Figure 1:** Proposed Methodology

This section discusses the structure of the Model. Data collection is the first step. There are different datasets available online or we can collect real-time datasets from hospitals. In this proposed methodology we have used Kaggle dataset which contains 303 records of patients with 14 attributes including the result. Then data preprocessing is needed. Since this was an online dataset, there were no missed values or redundant values. Split the dataset into two as Training set and Testing set. Choose the best correlated feature using correlation matrix. Train the model with the training set. Now the model has been trained with the labels and tested the model with Testing set. Performance of the model will be evaluated. Different evaluation metrics used here are Precision, Recall, F1-score, Accuracy, Confusion matrix and ROC curve. Then compared the accuracy of different models. The framework of the proposed model is in figure 1. These are briefly discussed in the following subsections.

## 3.1. Collection of Dataset

Dataset collection is the initial step in heart disease prediction. Here, we have used the Kaggle dataset. This dataset contains 14 attributes including the output. Table 1 depicts the attributes of the datasets used in this study.

**Table 1:** Attributes of Dataset

| Sl No: | Attribute Name | Description | Type |
|---|---|---|---|
| 1 | age | Patient's age | int64 |
| 2 | sex | Sex of the patient | int64 |
| 3 | exng | Angina induced by Exercise (1 = yes; 0 = no) | int64 |
| 4 | caa | Count of major vessels | int64 |
| 5 | cp | Chest pain type | int64 |
| 6 | trtbps | Resting blood pressure (in mm Hg) | int64 |
| 7 | chol | Cholesterol in mg/dl fetched via BMI sensor | int64 |
| 8 | fbs | FBS | int64 |
| 9 | restecg | Resting electro-cardiographic results | int64 |
| 10 | thalachh | Most heart rate attained | int64 |
| 11 | oldpeak | Exercise induced ST -segment depression relative to restfulness | float64 |

| | | | |
|---|---|---|---|
| 12 | thall | Thalium Stress results | int64 |
| 13 | slp | The slope of the peak exercise-ST segments | int64 |
| 14 | output | 0 = less chance of heart attack, 1= more chance of heart attack | int64 |

## 3.2. Selection of Attributes

The correlation matrix is used for feature selection. Before doing this step, we must remove the missing values to clean the dataset. Figure 2 shows the correlation matrix.

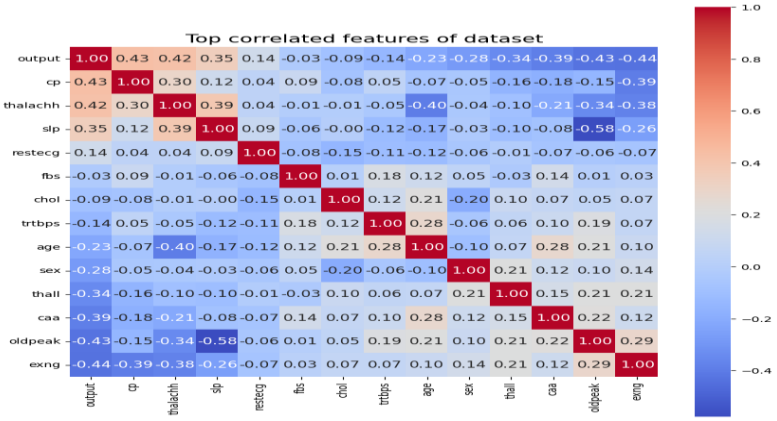



**Figure 2:** The correlation matrix

## 3.3. Splitting of the Dataset

We must split the dataset into Training and Testing. Here, we have split the dataset in 75:25 proportion. The training set is used for training the model and for testing the model test set is used.

## 3.4. Applying Machine Learning Algorithms

Different machine learning algorithms were applied to the dataset to find out the finest prediction model. The 5 types of algorithms compared here are Decision Tree Classifier, Random Forest, Logistic regression, KNN classifier, and Decision Tree Classifier with Gridsearch.

### 3.4.1. Decision Tree Classifier

It is a supervised machine-learning approach used in predictive models. If discrete values are taken by the target variable, then it is called a classification tree and if it takes continuous values, then it is called a regression tree.

### 3.4.2. Random Forest

It is suitable for both classification and regression tasks. Random forest algorithm combines several decision trees. In this, each tree in the forest predicts in which category a new record belongs. Then it is assigned to the category which got the majority vote.

### 3.4.3. Logistic Regression

It is used to solve binary classification problems. This algorithm predicts one of the possible outcomes based on distinctive features pertinent to the problem.

### 3.4.4. KNN Classifier

KNN is used in both classifications as well as regression tasks. Here, we use the Euclidian-distance formula to calculate the distance. 'k' in KNN is the number of neighbors we need to take into consideration while predicting the output.

### 3.4.5. Decision Tree classifier with Grid search

We are using Grid search to get the best parameters. It passes all the possible parameters one by one into the model and finally, it gives out the best parameters. These best parameters are used in Decision Tree Classifier. It is a supervised machine-learning approach used in predictive models. If the target variable takes discrete values, then it is called a classification tree and if it takes continuous values, it is called a regression tree.

### 3.4.6. Voting Method

Here, the voting method combines Naïve Bayes, Deep learning, Generalized Linear Model, and Random Forest.

## 3.5. Evaluation Metrics

Here we are calculating different evaluation metrics such as Precision, Recall, confusion matrix, F1-score, and accuracy.

Precision is calculated as,

$$P= \frac{TP}{TP+FP}$$

Recall is calculated as,

$$R=\frac{TP}{TP+FN}$$

F1 -Score is calculated as,

$$F1=2*(P*R)/P+R$$

$$Accuracy= \frac{True\ Positive+True\ Negative}{Total\ \ Sample}$$

## 3. Results and Discussions

**Table 2:** Results Comparison

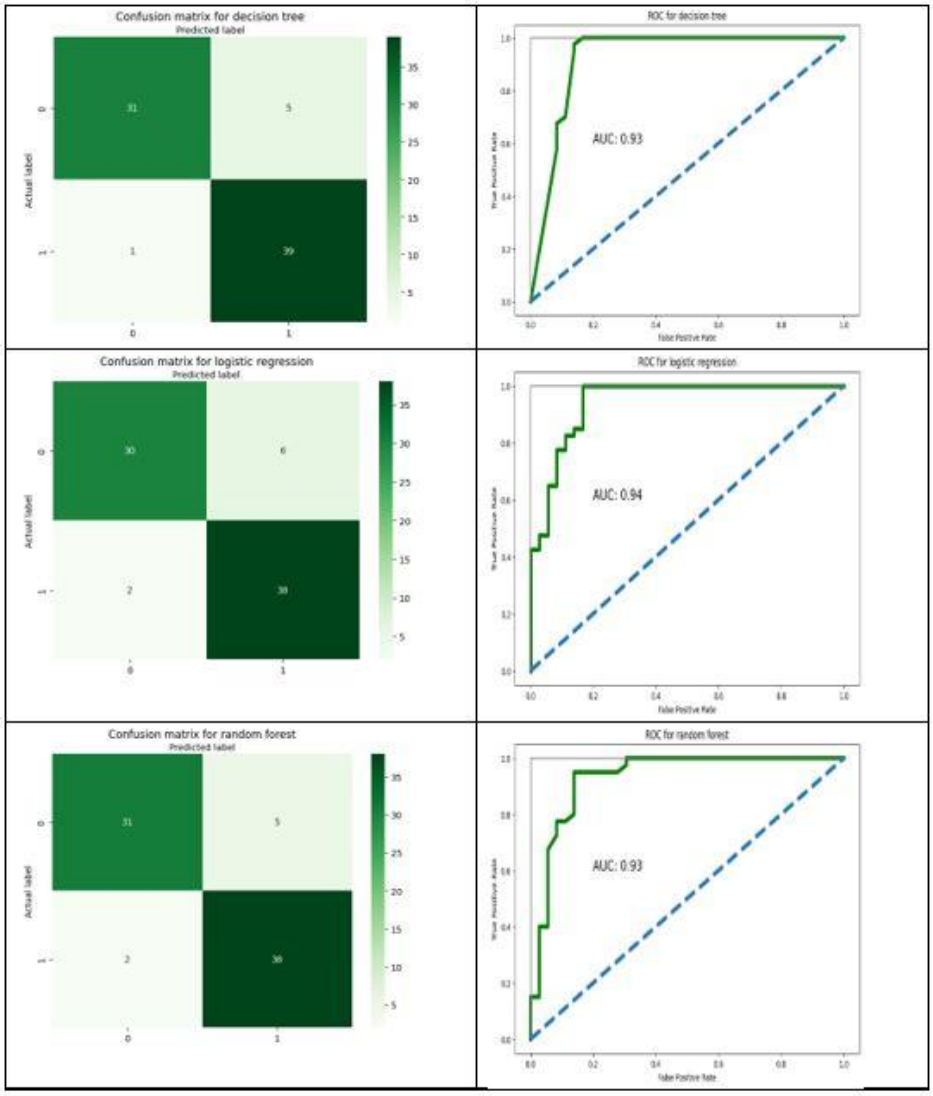
| Method | Precision | Recall | F1-score | Accuracy |
|---|---|---|---|---|
| Decision Tree Classifier | .89 | .97 | .93 | .92 |

| | | | | |
|---|---|---|---|---|
| Random Forest | .88 | .95 | .92 | .91 |
| Logistic regression | .86 | .95 | .90 | .89 |
| Decision Tree with Grid search | .86 | .93 | .89 | .88 |
| KNN | .70 | .82 | .76 | .72 |

In this study, as of now, we have compared five machine learning algorithms such as Decision Tree Classifier with Grid search, DT, RF, LR, and KNN. Precision, Recall, F1- score, confusion matrix, ROC curve, and accuracy are the different evaluation metrics calculated here.

Table 2 and Figure 3 depict the confusion matrix and ROC curves for the five diverse machine-learning algorithms.

| | **Predicted** | | |
|---|---|---|---|
| **Actual** | Total - 303 | Positive (1) | Negative (0) |
| | Positive (1) - 165 | TP | FN |
| | Negative (0) - 138 | FP | TN |



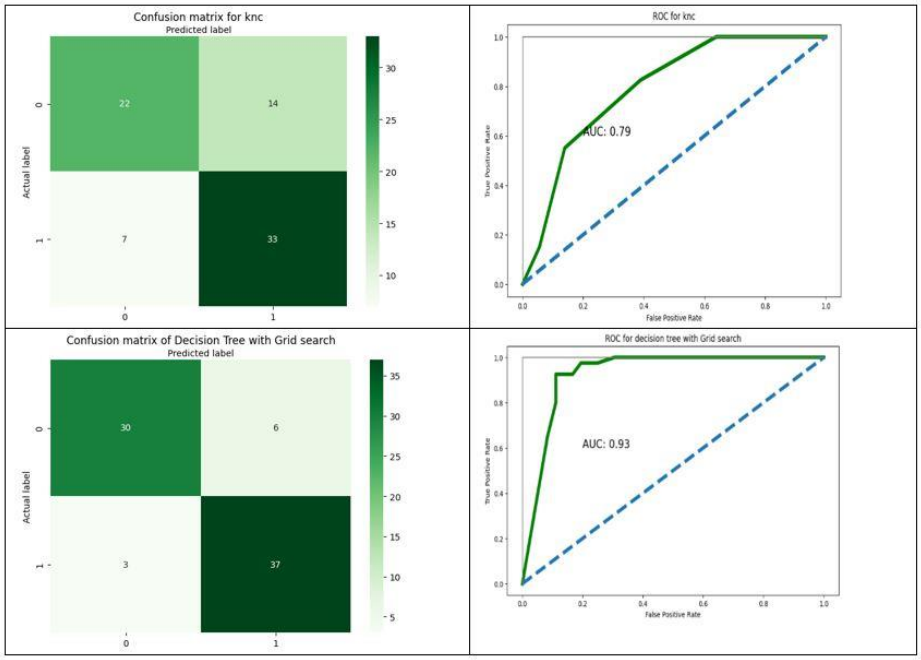



**Figure 3:** Confusion matrix and ROC curve for these four different machine learning algorithms.

| | score |
|---|---|
| Decision Tree Classifier | 0.921053 |
| Random Forest Classifier | 0.907895 |
| Logistic Regression | 0.894737 |
| Decision Tree Classifer with Gridsearch | 0.881579 |
| K-Neighbors Classifier | 0.723684 |

**Figure 4:** Shows the algorithm accuracy comparison.

Here, the Decision Tree classifier got more accuracy, and it is 92%. In this study, we developed a novel ensemble model which is a voting method that combines four different algorithms such as Naïve Bayes, Deep learning, Generalized Linear Model, and Random Forest. For this model, we got an accuracy of 85.44%. Figure 5 depicts the same.

# PerformanceVector

```
PerformanceVector:
accuracy: 85.44% +/- 5.34% (micro average: 85.48%)
ConfusionMatrix:
True:    1       0
1:       150     29
0:       15      109
absolute_error: 0.168 +/- 0.048 (micro average: 0.167 +/- 0.325)
```

**Figure 5:** Performance Vector of the ensemble model

## 4. Conclusion

The field of heart disease prediction continues to witness extensive research, reflecting the critical nature of accurate predictions in saving lives. As demonstrated in this study, numerous methods, including deep learning, machine learning, data mining, Artificial Neural Networks, and IoT, are being explored by researchers to predict heart disease. However, there is a persistent need for improved results in this domain. Given the gravity of this life-threatening disease, the development of more effective ensemble models becomes crucial. The Decision Tree Classifier demonstrated a higher accuracy rate in predicting heart disease in this study. With an accuracy of 92%, it outperformed the other evaluated machine learning algorithms, including Random Forest, Logistic Regression, K-Nearest Neighbour Classifier, and Decision Tree Classifier with Grid search. This finding highlights the effectiveness of the Decision Tree Classifier in accurately predicting heart disease. Future endeavors aim to incorporate real-time data and identify the optimal ensemble model for heart disease prediction. It is imperative to continue this quest for enhanced accuracy, as even a small error can have fatal consequences.

## 5. References


[1] Oikonomou E K, Siddique M & Antoniades C, Artificial Intelligence in Medical Imaging: a radiomic guide to precision phenotyping of Cardiovascular disease, *Cardiovascular Research*, **116(13)**, 2040-2054. https:// doi:10.1093/cvr/cvaa021.

[2] Li, B, Ding S, Song G, Li J, Zhang Q. 2019. Computer-Aided Diagnosis and Clinical Trials of Cardiovascular Diseases Based on Artificial Intelligence Technologies for Risk-Early Warning Model. *J Med Syst* **43,** 228. https://doi.org/10.1007/s10916-019-1346-x.

[3] Manish Sharma, U. Rajendra Acharya, 2019. A New Method to Identify Coronary Artery Disease with ECG Signals and Time-Frequency Concentrated Antisymmetric Biorthogonal Wavelet Filter Bank, *Pattern Recognition Letters*. **S0167-8655(19)**30121-7. https://doi.org/10.1016/j.patrec.2019.04.014.

[4] Abdar M, Ksiazek W, Rajendra Acharya U, Tan R S, Makarenkov V, Plaviak P. 2019.A New Machine Learning Technique for an Accurate Diagnosis of Coronary Artery Disease, *Computer Methods, and Programs in Biomedicine*. **S0169-2607**(18)31458-5. https://doi.org/10.1016/j.cmpb.2019.104992.

[5] Romiti S, Vinciguerra M, Saade W et al.,2020. "Artificial Intelligence (AI) and Cardiovascular Diseases: An Unexpected Alliance," Cardiology *Research and Practice*, vol. **2020**, Article ID 4972346, 8 pages. https://doi.org/10.1155/2020/4972346

[6] Johnson, K. W., Torres Soto, J., Glicksberg, B. S., Shameer, K., Miotto, R., Ali, M., ... & Dudley, J. T. (2018). Artificial intelligence in cardiology. *Journal of the American College of Cardiology*, **71**(23), 2668-2679. https://www.jacc.org/doi/abs/10.1016/j.jacc.2018.03.521.

[7] Acharya U R, Sree S V, Krishnan M M R, Krishnananda N, Ranjan S, Umesh P & Suri J S, (2013). Automated classification of patients with coronary artery disease using grayscale features from left ventricle echocardiographic images, *Computer Methods and Programs in Biomedicine*, Volume **112**, Issue 3,2013, Pages 624-632. https://doi.org/10.1016/j.cmpb.2013.07.012.

[8] Golovenkin S E, Dorrer M G, Nikulina S Y, Orlova Y V, & Pelipeckaya E Y, (2020), Evaluation of the effectiveness of using artificial intelligence to predict the response of the human body to cardiovascular diseases. *Journal of Physics: Conference Series*, Volume **1679**. No. 4, https://iopscience.iop.org/article/10.1088/1742-6596/1679/4/042017/meta.

[9] Babaoglu I, Findik O & Bayrak M, (2010). Effects of principle component analysis on assessment of coronary artery diseases using support vector machine, *Expert Systems with Applications*, Volume **37**, Issue 3, 2010, Pages 2182-2185. https://doi.org/10.1016/j.eswa.2009.07.055.

[10] Chen J I Z, & Hengjinda P. (2021). Early Prediction of Coronary Artery Disease (CAD) by Machine Learning Method- A Comparative Study, *Journal of Artificial Intelligence and Capsule Networks*, Vol.**03**/ No.01, Pages: 17-33. https://doi.org/10.36548/jaicn.2021.1.002.

[11] M M Ali, B K Paul, Kawsar Ahmed, Francis M B, Julian M W, Q M A Moni, (2021), Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison, *Computers in Biology and Medicine,* Volume **136**, https://doi.org/10.1016/j.compbiomed.2021.104672.

[12] Yazdani A, Varathan K D, Chiam Y K, Malik A W, & Wan Ahmad, W A (2021). A novel approach for heart disease prediction using strength scores with significant predictors. *BMC Medical Informatics and Decision Making* **21**(1). https://doi.org/10.1186/s12911-021-01527-5.

[13] Vardhan Shorewala,(2021), Early detection of Coronary Heart Disease using ensemble Techniques, Informatics in Medicine Unlocked, Volume 26, 100655, https://doi.org/10.1016/j.imu.2021.100655.

[14] Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access*, ***7*, 81542-81554**.

[15] Sabrina Mezzatesta, Claudia Torino, Pasquale De Meo, Giacomo Fiumara, Antonio Vilasi, (2019), A MACHINE LEARNING-BASED APPROACH FOR PREDICTING THE OUTBREAK OF CARDIOVASCULAR DISEASES IN PATIENTS ON DIALYSIS, *Computer Methods and Programs in Biomedicine, https://doi.org/10.1016/j.cmpb.2019.05.005*.

[16] Sumit Sharma, Mahesh Parmar (2020), Heart Diseases Prediction using Deep Learning Neural Network Model, *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, Volume-**9**, Issue-3. https://doi.org/10.35940/ijitee.C9009.019320.

[17] S M M Hasan, M A Mamun, M P Uddin, and M A Hossain, (Hasan, S. M. M., Mamun, M. A., Uddin, M. P., & Hossain, M. A. (2018, February). Comparative analysis of classification approaches for heart disease prediction. In 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2) (pp. 1-4). IEEE.

[18] Pranav Motarwar, Ankitha Duraphe, G Suganya and M Premalatha (2020), Cognitive Approach for Heart Disease Prediction Using Machine Learning, IEEE 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 1-5, doi:10.1109/ic-ETITE47903.2020.242.

[19] Ramalingam, V. V., Dandapath, A., & Raja, M. K. (2018). Heart disease prediction using machine learning techniques: a survey. International Journal of Engineering & Technology, 7(2.8), 684-687.

[20] Patel, J., TejalUpadhyay, D., & Patel, S. (2015). Heart disease prediction using machine learning and data mining technique. *Heart Disease*, *7*(1), 129-137.

[21] Diwakar, Manoj; Tripathi, Amrendra; Joshi, Kapil; Memoria, Minakshi; Singh, Prabhishek; Kumar, Neeraj (2020). *Latest trends on heart disease prediction using machine learning and image fusion. Materials Today: Proceedings, (), S2214785320367675–*. doi: 10.1016/j.matpr.2020.09.078

[22] Abdeldjouad, F. Z., Brahami, M., & Matta, N. (2020, June). A hybrid approach for heart disease diagnosis and prediction using machine learning techniques. In *International conference on smart homes and health telematics* (pp. **299-306**). Springer, Cham. https://doi.org/10.1007/978-3-030-51517-1_26.

[23] Pouriyeh, Seyedamin; Vahid, Sara; Sannino, Giovanna; De Pietro, Giuseppe; Arabnia, Hamid; Gutierrez, Juan (2017). *[IEEE 2017 IEEE Symposium on Computers and Communications (ISCC) - Heraklion, Greece (2017.7.3-2017.7.6)] 2017 IEEE Symposium on Computers and Communications (ISCC) - A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease.*, **204–207**. doi:10.1109/ISCC.2017.8024530.

[24] Bharti R, Khamparia A, Shabaz M, Dhiman G, Pande S, Singh P. Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning. Comput Intell Neurosci. 2021 Jul 1; 2021:8387680. doi 10.1155/2021/8387680. PMID: 34306056; PMCID: PMC8266441.
[25] Neeraj Baghel, Malay Kishore Dutta, Radim Burget, Automatic diagnosis of multiple cardiac diseases from PCG signals using convolutional neural network, Computer Methods, and Programs in Biomedicine, Volume 197, 2020, 105750, ISSN 0169-2607, https://doi.org/10.1016/j.cmpb.2020.105750.
[26] Harshit Jindal *et al* 2021 *IOP Conf. Ser.: Mater. Sci. Eng.* **1022** 012072
[27] Nandy, S., Adhikari, M., Balasubramanian, V. et al. An intelligent heart disease prediction system based on a swarm-artificial neural network. Neural Comput & Applic (2021). https://doi.org/10.1007/s00521-021-06124-1
[28] Shah, D., Patel, S. & Bharti, S.K. Heart Disease Prediction using Machine Learning Techniques. *SN COMPUT. SCI.* **1,** 345 (2020). https://doi.org/10.1007/s42979-020-00365-y
[29] ABID ISHAQ, SAIMA SADIQ, MUHAMMAD UMER, SALEEM ULLAH, SEYEDALI MIRJALILI, VAIBHAV RUPAPARA, AND MICHELE NAPPI, Improving the Prediction of Heart Failure Patients' Survival Using SMOTE and Effective Data Mining Techniques, *IEEE Access*, VOLUME 9, 2021, Pages 39707 -39716.
[30] Saboji, R. G. (2017, August). A scalable solution for heart disease prediction using classification mining technique. In *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)* (pp. 1780-1785). IEEE.
[31] Vaidya, N., Kandu, M., Yadav, R. K., & Bharadwaj, N. (2020). The Working of Various Prediction Techniques for Heart Diseases–A Case Study. *IJRAR-International Journal of Research and Analytical Reviews (IJRAR), 7*(1), 447-453.
[32] Nikolaiev, S., & Timoshenko, Y. (2015, October). Reinvention of the cardiovascular disease prevention and prediction due to ubiquitous convergence of mobile apps and machine learning. In *2015 Information Technologies in Innovation Business Conference (ITIB)* (pp. 23-26). IEEE.
[33] Owusu, E., Boakye-Sekyerehene, P., Appati, J. K., & Ludu, J. Y. (2021). Computer-aided diagnostics of heart disease risk prediction using boosting support vector machine. *Computational Intelligence and Neuroscience, 2021*.
[34] Liu J, Dong X, Zhao H, Tian Y. Predictive Classifier for Cardiovascular Disease Based on Stacking Model Fusion. *Processes*. 2022; 10(4):749. https://doi.org/10.3390/pr10040749
[35] Bhalla, R. and Bagga, A., 2019. Opinion mining framework using proposed rb-bayes model for text classification. International Journal of Electrical & Computer Engineering (2088-8708), 9(1).
[36] Bhalla, R., 2020. A comparative analysis of application of proposed and the existing methodologies on a mobile phone survey. In Futuristic Trends in Networks and Computing Technologies: Second International Conference, FTNCT 2019, Chandigarh, India, November 22–23, 2019, Revised Selected Papers 2 (pp. 107-115). Springer Singapore.
[37] Bhalla R, Sharma A, Gupta J. An Ensemble-based Machine Learning Model for Accurate Predictions using Multiple Categorical Datasets. In2022 Second International Conference on Interdisciplinary Cyber Physical Systems (ICPS) 2022 May 9 (pp. 1-4). IEEE.
[38] R. Bhalla, A. Sharma, and G. Ganesan, "A Bibliographic Survey of Sentiment Classification using Hybrid Ensemble-based Machine Learning Approaches," in ceur workshop proceedings, 2022, vol. 3146, pp. 33–38, [Online]. Available: https://ceur-ws.org/Vol-3146/PAPER_04.pdf.