# A Voting Ensemble based Machine Learning Approach to Predict Cardiovascular Disease

**Conference Paper** · December 2023

DOI: 10.1109/ICCIT60459.2023.10441059



**6 authors**, including:

Ahmad Jainul Abidin
Bangladesh University of Professionals


Rownak Borhan
Nara Instiute of Science and Technology


Md. Abdul Based
Dhaka International University


Abdul Kader Zilany
Gono Bishwabidyalay

2023 26th International Conference on Computer and Information Technology (ICCIT)
13-15 December, Cox's Bazar, Bangladesh

# A Voting Ensemble based Machine Learning Approach to Predict Cardiovascular Disease

Ahmad Jainul Abidin
*Information and Communication Technology*
*Bangladesh University of Professionals*
Dhaka, Bangladesh
ahmadadan75@gmail.com

Rownak Borhan
*Information Science*
*Nara Institute of Science and Technology*
Nara, Japan
rownakborhan34@gmail.com

Arifa Akhter
*Computer Science and Engineering*
*Jagannath Univeristy*
Dhaka, Bangladesh
arifarini27@gmail.com

Sanjeda Jahan
*Computer Science and Engineering*
*Dhaka International University*
Dhaka, Bangladesh
sanjeda8850@gmail.com

Md Abdul Based
*Computer Science and Engineering*
*Dhaka International University*
Dhaka, Bangladesh
baset123@gmail.com

Abdul Kader Zilany
*Computer Science and Engineering*
*Gono Bishwabidyalay*
Dhaka, Bangladesh
akzilanygb@gmail.com

2023 26th International Conference on Computer and Information Technology (ICCIT) | 979-8-3503-5901-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCIT60459.2023.10441059

***Abstract*—Heart disease has a huge impact on the worldwide mortality picture and accounts for a large number of annual deaths. This rate is growing due to factors like unhealthy diets, lack of physical activity, and tobacco consumption. As a result, adopting preventative measures and early detection has the potential to lessen the severity of the condition. This research aims to create a machine learning model that uses patients' medical variables as input to predict cardiac disease. To predict cardiovascular disease, this study analyzes various machine learning classifiers. The best outcome is finally obtained when employing the voting classifier ensemble technique. Random Forest and Gradient Boost combined technique and Random Forest and KNN combined approach-based voting classifiers both achieved 90% accuracy.**

***Index Terms*—Machine learning, heart disease prediction, voting classifier, cardiovascular disease (CVD), boosting ensemble**

## I. Introduction

Heart disease, commonly termed Cardiovascular Disease (CVD), is a general terminology used to describe an array of medical conditions that impact both the heart and the blood arteries [1]. Heart failure arises when the heart loses its ability to pump an adequate amount of blood and oxygen to fulfill the body's requirements [2]. Currently, cardiovascular disease stands as a prominent cause of death worldwide. As per the World Health Organization (WHO), in 2019, approximately 17.9 million individuals lost their lives due to heart disease, making up about 32% of all deaths globally [3]. Additionally, non-communicable diseases were responsible for nearly 85% of adult deaths in Bangladesh, with ischemic heart disease, stroke, and chronic respiratory disease accounting for about 60% of the total [4]. Therefore, an automated technique is needed to identify heart disease in its early stages.

Although the healthcare sector has advanced significantly, the current methods used for predicting heart disease still struggle with challenges that undermine their effectiveness [5]. The diagnosis mostly depends on a limited set of conventional risk factors, which might not capture the full complexity of an individual's susceptibility to heart disease. However, predicting heart disease requires taking into account several factors, including high blood pressure, age, elevated cholesterol levels, diabetes, and irregular pulse rate [6]. In order to address this problem, machine learning can be extremely helpful.

Utilizing machine learning, one may forecast CVD by analyzing multiple risk variables and medical data to build a prediction model. Many investigations have been conducted using machine learning algorithms to forecast cardiovascular disease.

The majority of previous research adopted the UCI Cleveland heart disease dataset [7], comprising 303 instances and 14 attributes, for their study. Our study distinguished itself by incorporating a dataset [8] that combines four additional datasets with the Cleveland dataset. In this research, the performance of various machine learning algorithms such as Support Vector Machine (SVM), Naive Bayes (NB), K-nearest neighbor (KNN), Logistic Regression (LR), Random Forest (RF), AdaBoost, Gradient Boosting, and XGBoost are assessed for predicting heart disease. To enhance the accuracy of these models, the Voting Classifier technique is implemented. The voting classifier yielded the best results among the techniques used. Moreover, a comparative analysis is conducted between the methods proposed in this paper and other existing methods.

The subsequent sections of this paper are organized as follows: Section II provides a review of existing related works, Section III covers the techniques used in this study, Section IV presents experiments and results analysis, and Section V summarizes the research findings.

## II. Literature Review

In recent times, the healthcare sector has witnessed substantial progress in the domain of machine learning and data mining. Many researchers adopted various machine learning (ML) techniques to predict cardiac disease by analyzing its

979-8-3503-5901-5/23/$31.00 ©2023 IEEE

risk factors. In this section, some of the existing literature is presented that employed ML algorithms for diagnosing heart disease.

Motarwar et al. [9] developed a framework for forecasting the likelihood of cardiac disease using five machine-learning algorithms. This framework employed the Cleveland dataset to assess the performance of the Hoeffding Decision Tree, Random Forest (RF), Naive Bayes (NB), Logistic Model Tree (LMT), and Support Vector Machine (SVM). Following feature selection, RF outperformed the other four algorithms, obtaining an accuracy of 95.08%. However, before the hyper-parameter adjustment, the initial accuracy was 88.52%.

Shah et al. [10] used Naive Bayes, Decision Tree, Random Forest, and K-nearest neighbor (KNN) algorithms for the prediction of heart disease. Among the four algorithms, the KNN achieved the highest accuracy of 90.78% for the training set, while its testing accuracy was 78.94%.

Bhatt et al. [11] proposed a novel k-mode clustering approach to enhance the classification of heart disease data and improve accuracy. The researchers utilized RF, DT, Multilayer Perceptron (MLP), and XGBoost (XGB) classifiers with GridSearchCV to optimize their parameters. The dataset for this study consists of 70,000 patient records with 12 features, which underwent preprocessing steps, including outlier removal, feature selection, and reduction. The multilayer perceptron with cross-validation stood up as the most accurate method, achieving an accuracy of 87.28% along with an AUC value of 0.95. The findings of this research highlight that transforming continuous data into categorical input through the process of binning can improve the efficiency and comprehensibility of classification algorithms. However, the major drawback of this study is that the model was not evaluated on an independent test dataset, which is essential for testing its generalization to new and unexplored data.

Boukhatem et al. [12] developed a heart disease prediction model employing SVM, MLP, RF, and NB. Prior to model development, the study executed feature selection and data preprocessing steps involving dimensionality reduction and the removal of severe outliers. The evaluation criteria for the models included accuracy, F1-score, precision, and recall. Through the process of feature selection, the SVM model showed a significant accuracy improvement, reaching 91.67%, in contrast to its initial accuracy of 88.33%. Therefore, this observation suggests that the performance of the model may be reduced when faced with a wide range of features.

Rubini et al. [13] presented a comparison of SVM, RF, NB, and Logistic Regression (LR) for categorizing cardiovascular illness. Notably, the Naive Bayes (NB) algorithm performed the worst, with an accuracy of 54.08%, while Random Forest (RF) demonstrated the best accuracy of 84.81%.

In their research, Sharma et al. [14] presented the application of an Optimized Deep Neural Network (DNN) utilizing the Talos model. They further performed a comparative evaluation of its efficacy in comparison to other ML methods, including KNN, SVM, NB, LR, and RF. The experiments were performed using the UCI Cleveland dataset. Leveraging hyper-parameter optimization (Talos) led to a notable accuracy improvement of 90.78%, resulting in the creation of a Keras model.

A hybrid HRFLM approach, which integrates the traits of Random Forest (RF) and Linear Model (LM), was proposed by Mohan et al. [15]. Utilizing this method, the authors employed three association rule mining techniques, such as apriori, predictive, and Tertius, to identify factors associated with cardiovascular disease within the UCI Cleveland dataset. Additionally, this study employs many ML techniques, such as the Generalized Linear Model (GLM), NB, LR, DL, DT, RF, SVM, and Gradient Boosting Tree (GBT). However, the HRFLM model produced the best result with an accuracy rate of 88.4% and a classification error rate of 11.6%. Moreover, a VOTE approach was also introduced in this study, which combines LR and NB. The VOTE technique yields an accuracy of 87.41%.

In a study, Pasha et al. [16] used deep learning techniques along with SVM, KNN, and DT for cardiovascular disease prediction. The dataset used in this research was sourced from Kaggle. However, the performance of these algorithms was found to be less effective when dealing with larger datasets. To overcome this, the authors implemented an Artificial Neural Network (ANN) and achieved 85.24% accuracy.

In their study, Jindal et al. [17] devised a Heart Disease Prediction system that considers 13 clinical parameters. This system incorporates KNN, LR, and RF classifiers as predictive algorithms. In this study, KNN performs more accurately than the other two algorithms, achieving an accuracy rate of 88.52%.

A majority voting ensemble technique that can predict a person's chance of developing heart disease was proposed by Atallah et al. [18]. The study employed the Cleveland dataset for its investigation. Four machine learning classifiers, namely Stochastic Gradient Descent (SGD), RF, LR, and KNN were integrated into the proposed voting ensemble. The ensemble model attained an accuracy of 90%, surpassing the accuracy achieved by each individual classifier.

Islam et al. [19] employed ten distinct machine learning classification models, including SVM, KNN, LR, DT, RF, Extra-Tree classifier, Gradient Boosting, XGBoosting, Adaptive Boosting, and MLP to predict heart disease. Alongside these models, they integrated four feature selection methods (PCA, Chi-2, RFE, and Pearson) to reduce the features. The XGBoosting classifier demonstrated the best performance with the reduced feature set.

Ramesh et al. [20] conducted a predictive analysis using supervised ML approaches like NB, SVM, KNN, DT, LR, and RF on a dataset containing 14 attributes. According to the authors, machine learning techniques in their study performed noticeably better than statistical methods.

## III. Methodology

This section outlines the detailed methodology utilized in this study to develop the proposed model. The voting classifier method is used in this work to create the suggested machine learning model. This model is created by simultaneously utilizing data preprocessing, data visualization, feature scaling, and machine learning's fusion approach. Figure 1 illustrates the overall research methodology framework.

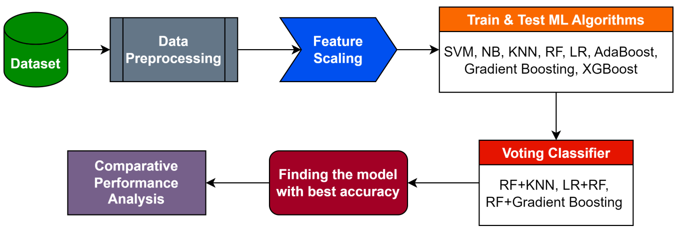


Fig. 1: Proposed framework for heart disease detection

### A. Dataset Description

The "Heart Failure Prediction Dataset", which was obtained from Kaggle [8], is used in this research. By merging five distinct publicly accessible datasets on heart disease, this dataset was developed, utilizing 11 common features. It comprises 918 observations in total, with 12 attributes. Table I provides a comprehensive overview of the dataset.

TABLE I: Dataset Descriptions

| Attributes | Type | Description |
|---|---|---|
| Age | Numerical | The patient's age (in years) |
| Sex | Categorical | Patient's gender (M: Male, F: Female) |
| ChestPainType | Categorical | Chest pain category [Non-Anginal Pain (NAP), Typical Angina (TA), Atypical Angina (ATA), Asymptomatic (ASY)] |
| RestingBP | Numerical | Resting blood pressure measurement (in mm Hg) |
| Cholesterol | Numerical | Serum cholesterol (in mm/dl) |
| FastingBS | Numerical | Fasting blood sugar (1: if FastingBS $>$ 120 mg/dl, 0: otherwise) |
| RestingECG | Categorical | Resting electrocardiogram [Normal: Normal, ST: ST-T wave abnormalities (T wave inversions and/or ST elevation or depression of $>$ 0.05 mV), LVH: probable or definite left ventricular hypertrophy per Estes' criteria] |
| MaxHR | Numerical | Attained maximum heart rate (numerical value between 60 and 202) |
| ExerciseAngina | Categorical | Physical activity-induced angina (N: No, Y: Yes) |
| Oldpeak | Numerical | Compared to rest, exercise-induced ST depression (as assessed by a numeric value in depression) |
| ST_Slope | Categorical | The gradient of the maximal exercise ST segment (Up: upsloping, Down: downsloping, Flat: flat) |
| HeartDisease | Numerical | The target property (1: Heart disease, 0: Normal) |

### B. Data Preprocessing

**One-Hot Encoding:** Machine learning algorithms cannot directly handle categorical values. In the dataset used in this paper, attributes such as Sex, ChestPainType, RestingECG, ExerciseAngina, and ST_Slope contain categorical data. One-hot encoding, a data preprocessing technique, is implemented to make these columns suitable for machine learning models. One-hot encoding transforms categorical data into a binary representation that machine learning models can process.

**Feature Scaling:** Feature scaling is performed to ensure that all numerical features in the dataset have similar scales, which helps prevent some attributes from dominating others during model training. In this research, the standardization method is applied for feature scaling. Equation (1) displays the mathematical representation of the standard scaler, where X represents the standardized value of feature x, the function mean() computes the average value of feature x, and std() calculates its standard deviation.

$$X = \frac{x - \text{mean}(x)}{\text{std}(x)} \tag{1}$$

### C. Classification Algorithms

- **Support Vector Machine (SVM):** SVM is a robust classification algorithm known for finding the optimal hyperplane that maximizes the margin between different classes. Its versatility extends to both linear and non-linear problems, making it a valuable tool in various domains like image recognition and finance.
- **Random Forest (RF):** Random Forest, an ensemble approach, merges numerous decision trees to mitigate overfitting and enhance accuracy. By aggregating predictions from various trees, it enhances model robustness and is particularly effective in complex, noisy datasets. Random Forest achieved good accuracy in this study compared to others will be discussed in the result section.
- **K-nearest Neighbour (KNN):** KNN classifies data points based on their proximity to others. It's simple yet effective for pattern recognition tasks, although its performance can be influenced by the selection of the "k" parameter.
- **Logistic Regression (LR):** A fundamental classification approach, logistic regression models the likelihood of a binary outcome. It's widely used in medical diagnostics and marketing analytics for its simplicity and interpretability.
- **Ada boost classifier (ADB):** AdaBoost is an ensemble technique that combines weak classifiers to create a strong one. It assigns more weight to misclassified samples, iteratively improving performance and adapting well to diverse datasets.
- **Voting Classifier:** The voting classifier is an effective ensemble learning method that aggregates the predictions of numerous separate machine learning models in order to arrive at a conclusive result. It leverages the wisdom of the crowd to improve overall predictive accuracy and is particularly effective when different models bring diverse perspectives to the problem at hand. There are two distinct types of voting classifiers, namely hard voting and soft voting. In hard voting, every base model (classifier) in the ensemble casts a vote for a class, and the ultimate prediction is determined by the class that accumulates the majority of votes. This is most commonly used for classification tasks where the classes are discrete. This

method was used in this study to get the best result. In soft voting, the base models provide probability estimates for each class, and the final prediction is made by averaging these probabilities. This approach can be more effective when models produce probability projections, as it considers the confidence of each classifier in its predictions.

## IV. Experiment and Result Analysis

The coding environment for the proposed study utilized Python as the language for implementing machine learning algorithms, and it was hosted on Google Collaborative. This research uses Python libraries such as Numpy, Pandas, Matplotlib, Seaborn, and more. The dataset was split into a 75:25 ratio for training and testing, respectively.

In the dataset, out of 918 observations, 508 correspond to heart disease cases, while 410 are classified as normal patients. By analyzing the dataset, it is found that male individuals are more prone to cardiac disease than female patients. Figure 2 indicates that heart disease was present in only 50 women, whereas men exhibited nearly 460 cases, signifying an approximately nine times higher difference. Chest pain is

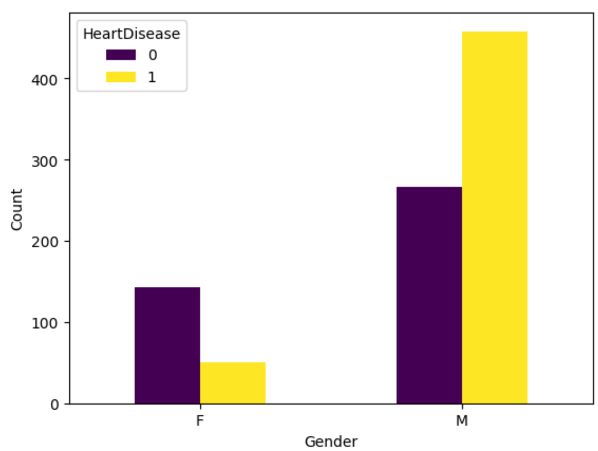


Fig. 2: Distribution of gender types by heart disease

a typical symptom of heart disease that can be very helpful in the diagnosis. Figure 3 illustrates the distribution of various chest pain types and their association with cardiovascular disease. Notably, asymptomatic chest pain (ASY) is more prevalent among individuals with heart disease. In contrast, people without heart disease are more likely to experience atypical angina (ATA) and non-anginal pain (NAP) chest pain. In addition to gender and chest pain, the dataset reveals other risk factors that are strongly connected with cardiac disease.

A number of machine learning (ML) approaches were implemented in this research endeavor to detect the existence of cardiovascular disease in the dataset. The GridSearchCV method is used to find the best hyperparameter for SVM, and it performed the best with poly kernel. The performance metrics for ML algorithms are detailed in Table II. F1 score, precision, recall, and accuracy are computed in order

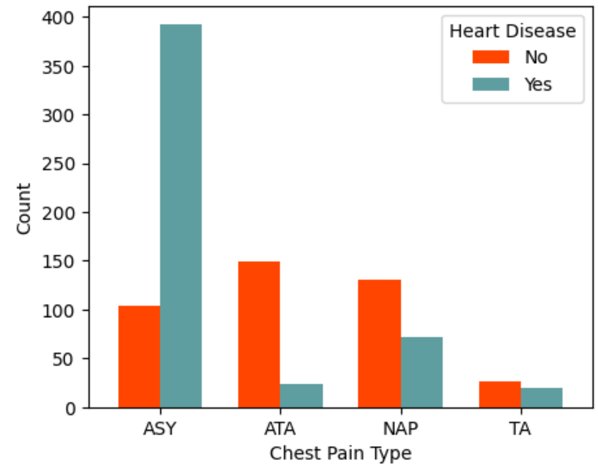


Fig. 3: Association between chest pain types and heart disease incidence

to conduct a comprehensive evaluation of the models. The findings indicate that the RF algorithm exhibited the best level of accuracy, reaching 88.70%. Additionally, it demonstrated notable precision, recall, and F1 score values of 0.92, 0.88, and 0.90, respectively. The NB predictive model showed an accuracy 87.39%, mirroring the SVM's performance. KNN and LR yielded substantial accuracy rates of 87.83% and 86.52%, respectively. In terms of precision, the SVM model leads the way with an impressive score of 0.93 among the algorithms. Simultaneously, ensemble-based boosting methods are implemented, namely Gradient Boost, AdaBoost, and XGBoost. Among these, the Gradient Boosting classifier achieved the best level of accuracy, reaching nearly 87%. All features within the dataset were utilized for the purpose of training and evaluating each of the models.

TABLE II: Comparative Analysis of ML Algorithms Performance for HD Prediction

| Algorithms | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| SVM | 87.39% | 93% | 85% | 89% |
| NB | 87.39% | 92% | 86% | 89% |
| KNN | 87.83% | 91% | 87% | 89% |
| RF | 88.70% | 92% | 88% | 90% |
| LR | 86.52% | 90% | 86% | 88% |
| AdaBoost | 84.78% | 91% | 82% | 86% |
| Gradient Boosting | 86.96% | 92% | 85% | 88% |
| XGBoost | 86.52% | 92% | 84% | 88% |

Subsequently, this study employed the voting ensemble technique to enhance the accuracy of ML algorithms. Remarkably, the ensemble approach involving RF and KNN achieved an impressive accuracy of 90%, exceeding all previously used algorithms. Similarly, the combination of LR and RF, as well as RF and Gradient Boosting, both produced high accuracy scores of 88.26% and 90%, respectively. Hence, it's evident from Table III that the voting classifier technique holds considerable potential for improving ML model performance, which makes it a promising approach for accurate heart disease prediction.

TABLE III: Performance of Voting Ensemble Algorithms

| Voting classifiers | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| RF + KNN | 90% | 92% | 90% | 91% |
| LR + RF | 88.26% | 90% | 89% | 90% |
| RF + Gradient Boosting | 90% | 92% | 90% | 91% |

TABLE IV: Performance Comparison with Previous Works

| Ref | Year | Dataset | HD Prediction Technique | Accuracy |
|---|---|---|---|---|
| Bhatt et al. [11] | 2023 | Cardiovascular Disease dataset from Kaggle (70,000 instances and 12 features) | Multilayer Perceptron (MLP) | 87.28% |
| Mohan et al. [15] | 2019 | UCI Cleveland dataset | VOTE (LR + NB), HRFLM (proposed) | 87.41%, 88.4% |
| Kavitha et al. [21] | 2021 | UCI Cleveland dataset | Hybrid (Decision Tree + Random Forest) | 88.7% |
| Rani et al. [22] | 2021 | UCI Cleveland dataset | Random Forest | 86.6% |
| Hossain et al. [23] | 2023 | Data was gathered from hospitals, clinics, and diagnostic centers in Bangladesh (59 patient's test results with 19 attributes). | Random Forest (with selected features) | 90% |
| Atallah et al. [18] | 2019 | UCI Cleveland dataset | Hard Voting Ensemble Method (SGD + KNN + RF + LR) | 90% |

In order to further contextualize the significance of the above findings, a comparative analysis is carried out with the existing published literature. Table IV demonstrates recent research outcomes in cardiovascular disease prediction. A study by Atallah et al. [18] utilized a hard voting ensemble approach by combining four distinct algorithms (SGD, KNN, RF, LR) to achieve 90% accuracy, while the voting classifier technique in this paper reached the same accuracy using only two algorithms (RF, KNN). Moreover, in [15], a fusion of LR and NB was used for heart disease prediction, resulting in an accuracy of 87.41%. However, a couple of studies [9], [12], [23] have surpassed the performance of this proposed model by utilizing selected features, whereas the research in this paper takes into account all the features within the dataset. Overall, this study's results exhibit a competitive performance in the context of cardiac disease prediction.

## V. Conclusion

An early diagnosis of heart disease offers prompt intervention, better treatment results, and decreased risk of complications. This study evaluates various machine learning algorithms for cardiac disease prediction. This work included a comprehensive assessment of the predictive capabilities of SVM, NB, KNN, RF, LR, AdaBoost, Gradient Boosting, and XGBoost classifiers. However, previous studies predominantly relied on the UCI Cleveland dataset, while this research stands out by integrating four supplementary datasets with the Cleveland dataset. The Random Forest algorithm had the best results with 88.70% accuracy, 0.92 precision, 0.88 recall, and an F1-score of 0.90. In order to augment the precision of the suggested model, the voting classifier technique was utilized, which emerged as the leading approach among the methods used. The ensemble combining RF and KNN, as well as RF and Gradient Boosting through a voting approach, achieved a noteworthy accuracy of 90%. These findings emphasize the transformative potential of machine learning in predicting cardiovascular disease.

## References


[1] S. Robinson, "Cardiovascular disease," in *Priorities for Health Promotion and Public Health*. Routledge, 2021, pp. 355–393.

[2] J. J. Park and D.-J. Choi, "Current status of heart failure: global and korea," *The Korean journal of internal medicine*, vol. 35, no. 3, pp. 487–497, 2020.

[3] World Health Organization. (2023) Cardiovascular diseases (cvds). Accessed: 2023-07-25. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[4] M. T. H. Shawon, S. A. A. Ashrafi, A. K. Azad, S. M. Firth, H. Chowdhury, R. G. Mswia, T. Adair, I. Riley, C. Abouzahr, and A. D. Lopez, "Routine mortality surveillance to identify the cause of death pattern for out-of-hospital adult (aged 12+ years) deaths in bangladesh: introduction of automated verbal autopsy," *BMC public health*, vol. 21, pp. 1–11, 2021.

[5] R. Kones, "Primary prevention of coronary heart disease: integration of new data, evolving views, revised goals, and role of rosuvastatin in management. a comprehensive survey," *Drug design, development and therapy*, pp. 325–380, 2011.

[6] R. Hajar, "Risk factors for coronary artery disease: historical perspectives," *Heart views: the official journal of the Gulf Heart Association*, vol. 18, no. 3, pp. 109–114, 2017.

[7] S. W. P. M. Janosi, Andras and R. Detrano, "Heart Disease," UCI Machine Learning Repository, 1988, DOI: https://doi.org/10.24432/C52P4X.

[8] Fedesoriano, "Heart failure prediction dataset," https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction, 2023.

[9] P. Motarwar, A. Duraphe, G. Suganya, and M. Premalatha, "Cognitive approach for heart disease prediction using machine learning," in *2020 international conference on emerging trends in information technology and engineering (ic-ETITE)*. IEEE, 2020, pp. 1–5.

[10] D. Shah, S. Patel, and S. K. Bharti, "Heart disease prediction using machine learning techniques," *SN Computer Science*, vol. 1, pp. 1–6, 2020.

[11] C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, "Effective heart disease prediction using machine learning techniques," *Algorithms*, vol. 16, no. 2, p. 88, 2023.

[12] C. Boukhatem, H. Y. Youssef, and A. B. Nassif, "Heart disease prediction using machine learning," in *2022 Advances in Science and Engineering Technology International Conferences (ASET)*. IEEE, 2022, pp. 1–6.

[13] P. Rubini, C. Subasini, A. V. Katharine, V. Kumaresan, S. G. Kumar, and T. Nithya, "A cardiovascular disease prediction using machine learning algorithms," *Annals of the Romanian Society for Cell Biology*, pp. 904–912, 2021.

[14] S. Sharma and M. Parmar, "Heart diseases prediction using deep learning neural network model," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 9, no. 3, pp. 2244–2248, 2020.

[15] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE access*, vol. 7, pp. 81 542–81 554, 2019.

[16] S. N. Pasha, D. Ramesh, S. Mohmmad, A. Harshavardhan *et al.*, "Cardiovascular disease prediction using deep learning techniques," in *IOP conference series: materials science and engineering*, vol. 981, no. 2. IOP Publishing, 2020, p. 022006.

[17] H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, “Heart disease prediction using machine learning algorithms,” in *IOP conference series: materials science and engineering*, vol. 1022, no. 1. IOP Publishing, 2021, p. 012072.
[18] R. Atallah and A. Al-Mousa, “Heart disease detection using machine learning majority voting ensemble method,” in *2019 2nd international conference on new trends in computing sciences (ictcs)*. IEEE, 2019, pp. 1–6.
[19] M. R. Islam, M. D. Hoda, M. A. Rashid, S. A. Suha, and M. T. I. Miya, “Data-driven heart disease prediction by ensemble feature selection and machine learning techniques,” in *2022 25th International Conference on Computer and Information Technology (ICCIT)*. IEEE, 2022, pp. 575–580.
[20] T. Ramesh, U. K. Lilhore, M. Poongodi, S. Simaiya, A. Kaur, and M. Hamdi, “Predictive analysis of heart diseases with machine learning approaches,” *Malaysian Journal of Computer Science*, pp. 132–148, 2022.
[21] M. Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai, and R. S. Suraj, “Heart disease prediction using hybrid machine learning model,” in *2021 6th international conference on inventive computation technologies (ICICT)*. IEEE, 2021, pp. 1329–1333.
[22] P. Rani, R. Kumar, N. M. 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, 2021.
[23] M. I. Hossain, M. H. Maruf, M. A. R. Khan, F. S. Prity, S. Fatema, M. S. Ejaz, and M. A. S. Khan, “Heart disease prediction using distinct artificial intelligence techniques: performance analysis and comparison,” *Iran Journal of Computer Science*, pp. 1–21, 2023.