ORIGINAL ARTICLE

# ENHANCING HEART HEALTH: PERFORMANCE ANALYSIS & COMPARISON OF SUPERVISED MACHINE LEARNING ALGORITHMS FOR CARDIOVASCULAR DISEASE PREDICTION

Waseem Javed[1], Muhammad Owais[2]

1. School of Computing & Information Technology Institute of Management Sciences Peshawar, Khyber Pakhtunkhwa, Pakistan.
2. Faculty of Computer Science and Engineering (FCSE), GIKI Swabi, Khyber Pakhtunkhwa, Pakistan.

## ABSTRACT

Heart disease, commonly known as cardiovascular disease, has become a leading cause of death globally. It includes numerous disorders that affect the heart and has been a major cause of death around the world in the previous few decades, about 26 million people effects every year. The prediction and prevention of heart failure is a challenge for cardiologists and cardio-surgeons. The modern lifestyle, poor diet, lack of exercise, and high stress and depression level also increase the rate of cardiovascular disease. Early detection of cardiovascular disease signs and consistent medical monitoring can help decrease in the number of patients and mortality, but however, with limited medical resources and specialist consultants, it is difficult to continuously observe the patient and provide consultation

The healthcare industry holds a substantial amount of data, making machine learning algorithms essential for accurately predicting heart diseases and facilitating informed decision- making. Recent studies have explored the integration of these approaches to create hybrid machine learning algorithms. In this research, some of the data pre-processing techniques, such as eliminating noisy data, eliminating missing data, filling in default values where applicable, and splitting attributes into categories for predictions and decision making across different levels, This project suggests the development of a predictive model to determine the likelihood of individuals having a heart disease, aiming to offer both awareness and diagnostic insights. The accuracy of several techniques, such as Support Vector Machine, Logistic Regression, Random Forest Classifier, Naive Bayes Classifier, and K-Nearest Neighbour, are compared in order to achieve this goal. The goal of this comparison is to determine which model predicts cardiovascular disease the most accurately.

Keywords - Healthcare, Cardiovascular Disease, Machine Learning Algorithms

## INTRODUCTION

Artificial Intelligence (AI) has really made its mark in medicine, particularly when it comes to heart care. These days, digital technologies act like monitoring devices to collect massive amounts of data on cardiovascular disease. Cardiovascular Disease (CVDs) or Heart Failure (HFs) are the primary cause of death rates worldwide[1]. Cardiovascular diseases, including coronary artery disease, arrhythmia, and congenital heart defects, can have a significant impact on the heart and blood vessels[2]. The World Health Organization (WHO) released statistics showing that approximately 17.9 million fatalities globally occur due to cardiovascular disease, making up 32% of all deaths [3]. These days, machine learning algorithms are primarily utilized in healthcare prediction across the globe to enhance the diagnostic process and detect diseases early. The early prediction of disease at its initial stages saves many people's lives worldwide. Even so, cardiovascular disease affects thousands of people each year. The danger of death rates from cardiovascular disease can be decreased if robots are able to predict the disease in its early stages. Heart disease is the leading cause of death globally, despite the heart being an indispensable organ in the human body. Blood flow to the other organs is hampered when the heart isn't working properly. That results in the brain and limbs ceasing to operate, leading to death within mere moments. Cardiovascular disease is among the leading illnesses affecting middle-aged and elderly people, frequently resulting in severe complications within the human body. Although each type of cardiovascular disease, including coronary artery disease, congestive heart failure, and heart attack, necessitates distinct treatments, they may present similar warning signs. Pain in the chest, arm, or below the breastbone are frequently observed symptoms associated with heart disease[4]. The most challenging aspect of the disease is to predicting them, while the average pulse rate falls within the range of 60-100 beats per minute and the intermediate level of blood pressure is commonly recorded as 120/80, these baseline values serve as general benchmarks. Cardiovascular or heart diseases is a most serious health condition that affect the heart vessels The World Health Organization (WHO) projects that by 2030, there would be about 23.6 million deaths worldwide as a result of cardiovascular diseases (CVDs). The leading factors contributing to this rise are expected to be heart disease and stroke. From the last few decades, CVD has been a leading cause of death around the world.

According to the World Heart Federation, 20.5 million people died from the cardiovascular disease in 2021. The CVD recorded 12.1 million deaths in 1990, a significant increase in the total number of deaths worldwide. Currently, heart disease, sometimes referred to as cardiovascular disease, is the leading cause of death in 146 nations for men and 98 countries for women. On the one hand, conventional methods for diagnosing heart disease include conducting an electrocardiogram (ECG), auscultation, and measuring blood pressure, blood sugar, and cholesterol levels. However, these techniques are costly and time consuming. On the other hand, machine learning algo- rithms such as SVM, NB, RFC, and LG offer an alternative approach. These algorithms reduce the time spent processing and enhance prediction performance. This study's goal is to identify an efficient approach by comparing the accuracy of several machine learning algorithms.

The healthcare sector has made significant strides in data mining and machine learning applications in recent years, especially in the field of medical cardiology. The quick growth of medical data has made these tactics effective in a range of healthcare settings. Researchers are concentrating more on finding risk factors and early warning signs because heart disease is still one of the main causes of mortality in developing nations. By efficiently leveraging the enormous quantity of medical data that is already available, the combination of data mining techniques and machine learning algorithms holds tremendous potential for the early detection and prevention of cardiac disease[6-10].

Machine Krumholz et al. used machine learning approaches to predict the hospitalization and death rates of heart failure patients[11]. For two distinct approaches—one involving the forward selection of variables and the other utilizing lasso regularization for variable selection—they employed five methods, including logistic regression, gradient descent boosting (GDB), support vector machines (SVM), and random forests (RF). Over a three-year follow-up period, all of these algorithms were used to train models for assessing the risks of hospitalization for heart failure (HF) and mortality. Five-fold cross-validation was used to validate the models. The best results were obtained using the Random Forest approach, which had a mean C-statistic value of 0.76 for hospitalization and 0.72 for death prediction.

In order to assess CVD in patients undergoing dialysis, Vilasi et al. use machine learning algorithms [12]. The algorithms were tested on datasets from the United States and Italy. The Support Vector Machine (SVM) with RBF Kernel method fared better in the Italian and American datasets, with accuracy rates of 95.25% and 92.15%, respectively. It's crucial to understand, though, that inherent bias in the Italian dataset may affect how accurate the estimates are.

Amin et al. created a hybrid intelligent system for heart disease prediction by employing artificial neural networks (ANN), support vector machines (SVM), k-nearest neighbors (KNN), decision trees, Naive Bayes, and random forests [13]. Three feature selection strategies (Relief, mRMR, and LASSO) improved prediction efficiency: Logistic Regression used 10-fold validation and the relief feature selection algorithm to achieve 89 percent accuracy. According to the study, adding optimization methods based on neural networks could enhance the model's results even further.

Ashok used many methods, like ANN, KNN, SVM, Logistic Regression, Classification Tree, and Naive Bayes, to guess heart problems. As per the[14] logistic regression yielded the highest accuracy rate of 85 percent, accompanied by impressive sensitivity (89 percent) and specificity (81 percent). Further testing of the model with larger datasets is necessary to ensure its reliability.

Stephen et al. used routine clinical data from 378,256 UK citizens to predict cvd using machine learning[15]. Techniques including Random Forest, Logistic Regression, Neural Network, and Gradient Boosting were used; Neural Network produced the best results in terms of accuracy. The interpretation of the Neural Network process is made more challenging due to its inherent complexities. A system utilizing a decision tree and Support Vector Machine (SVM) was created in a study centered on the diagnosis of cardiac illnesses[16]. The idea was to use a person's heart condition to forecast the likelihood of heart failure. The system used SVM for classification and a decision tree for feature selection. With c-svc and an RBF kernel, it achieved an accuracy of 82.35 percent.

In order to identify important risk factors for cardiovascular disease (CVD) in individuals with metabolically-associated fatty liver disease (MAFLD), Drod et al. (2022)[17] employed machine learning approaches. For 191 MAFLD patients, the study comprised subclinical coronary artery disease assessment and blood biochemical analysis. Using several machine learning methods, such as logistic regression, principal component analysis, and uni variate feature ranking, the model determined that the duration of diabetes, plaque scores, and high cholesterol were significant clinical factors. With an AUC of 0.87, the machine learning approach effectively identified high-risk patients at a rate of 85.11 percent and low-risk patients at 79.17 percent. Based on straightforward criteria, the results indicate that ML can be useful in identifying extensive CVD in patients with MAFLD.

Random forest, decision tree, multilayer perceptron (MLP), and XGBoost classification algorithms were used in the study by Bhatt et al.,[18] which assessed their effectiveness using a dataset related to cardiovascular illness. The GridSearchCV approach was utilized for hyper parameter tuning, resulting in MLP achieving a maximum cross-validation accuracy of 87.28 percent. Additionally, it demonstrated impressive scores for recall, precision, F1 score, and AUC. With the exception of random forest and XGBoost, all classifiers had accuracies greater than 86.5 percent following hyperparameter adjustment.

## I. MATERIAL AND METHODOLOGY

Because of its ease of use, capacity to show the maturity of data mining projects, ease of project replication, and enhanced project planning and management, the CRISP-DM technique was chosen for this study. An organized approach to plan a DM project is to use the CRISP-DM methodology, a six-phase hierarchical process that is divided into the following components: business understanding, data understanding, data preparation, modeling, assessment, and deployment to identify and comprehend its primary risk factors in order to address this global issue. The major objective is to develop a system that, a business understanding during this phase, it is important to gain a thorough understanding of the business, including the customer's needs. The focus should be on comprehending the project's objectives and requirements.

Given the alarming number of deaths caused by cardiovascular disease, it has become imperative with the use of medical information gathered from examinations, can reliably forecast whether or not individuals will have cardiovascular diseases (CVDs). This will help to significantly reduce the time it takes to diagnose patients and provide them with immediate and appropriate treatment[20].

### B. Data Understanding

In this phase, one starts with data understanding, which shows how to identify, collect, and analyze the datasets and get the project goals. For this study, a Kaggle dataset on cardiovascular disease was employed[21]. Twelve characteristics total, one of which is a target variable. Table 1 is a representation of the same. For the analysis, people between the ages of 29 and 64 were taken into consideration. In addition, the individual's weight and height are documented. Patients were assigned gender values of 1 for males and 0 for females. The calculation involves determining the diastolic and systolic blood pressures in order to measure the impact. The patients' values of blood pressure and cholesterol fell into three categories: normal, above normal, and severely above normal. When it comes to heart disease, drinking and smoking go hand in hand. Binary values are assigned to these two variables. A number of ”1” indicates that the patient is a ”drinker/smoker,” whereas ”0” indicates that they are ”non- drinker/non-smoker.” '1' for patients who engage in regular physical activity and '0' for everyone else. The target attribute is determined by the presence or absence of cardiovascular disease. It consists of binary values. A ”0” denotes normal, whereas a ”1” indicates those who have been diagnosed with cardiac disease.

TABLE I
CHARACTERISTICS OF THE ATTRIBUTES

| Attribute | Description | Type |
|---|---|---|
| ID | Patient Unique Identifier | Obj |
| Age | Patient Age In Days | Obj |
| Height | Patient Height In Centimetres | Obj |
| Weight | Patient Weight In Kilograms | Obj |
| Gender | Patient Gender (male Or Female | Obj |
| Ap Hi | Systolic Blood Pressure | Medical-exam |
| Ap Lo | Diastolic Blood Pressure | Medical-exam |
| Cholesterol | Patient Cholesterol Level | Medical-exam |
| Gluc | Patient Glucose Level | Medical-exam |
| Smoke | Patient Smoker Or Non Smoker | Subj |
| Alco | Patient Consumes Alcohol Or Not | Subj |
| Active | Patient Physically Active Or Not | Subj |
| Cardio | Presence Of Cvd | Target Value |

There are seventy thousand patient records in the dataset. Three categories were created from the characteristics that were gathered during the patient's medical examination:

. Factual data, or objective data.

. Examination data: the outcomes of health examinations.

. Patient-provided information is referred to as subjective data.

A description of each attribute may be found in Table I.

The study's primary focus is the cardio characteristic, which offers details regarding the existence or lack of cardiovascular disease. If the cardio value is ”0”, it indicates the patient's good health. Conversely, a value of ”1” suggests the presence of cardiovascular disease. It was important to investigate if there were any differences in social class that could impact the learning process of algorithms. According to Figure XYZ, the distribution of the target attribute is fairly balanced, with 50.2% classified as ”yes” and 49.8% classified as ”no”. This translates to 34,820 registers for ”yes” and 34,605 registers for ”no”.

## C. Data Preparation

Real-world data includes noisy and high amounts of missing data. To get around these problems and produce accurate forecasts, these data have been pre-processed. The sequential chart of our suggested model is explained in Figure 1. Noise and missing values are typically present when cleaning the acquired data. For a precise and useful outcome, these data require must be cleared of any noise and missing values to become flustered.

Transformation modifies the data's format from one to con- vert one form to another for easier comprehension. It entails duties related to aggregation, standardization, and smoothing. Integration The data must be integrated before processing because it may come from a variety of sources rather than just one. To get effective results, the Reduction of the acquired data must be prepared because it is complex.

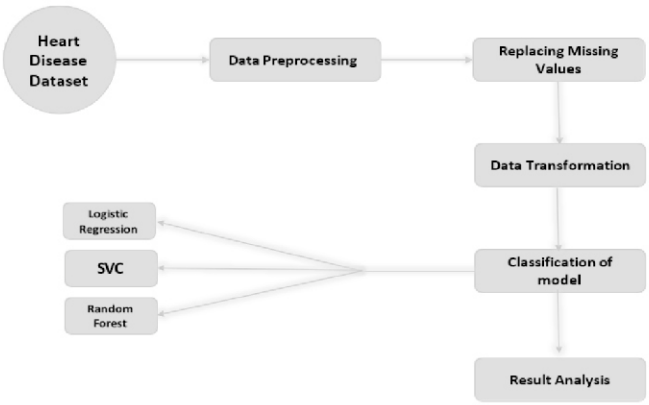


**Fig. 1. Chart of the Proposed Model**

After classification, the data are split into training and test sets, which are subsequently run through a number of algorithms to get the accuracy score.

Initially, data integration and data cleansing were performed. A more comprehensive examination was conducted to identify any instances of duplicate data, missing values, outliers, or inconsistencies. No instances of duplicate data, missing values, or inconsistencies were identified.

Afterwards, it was necessary to advance to the process of converting the data. The attribute ”age” was transformed from days to years using the equation ”age/365.” The majority of characteristics were imported as integers by default. However, the polynomial type is linked to cholesterol, while the binomial type is linked to the characteristics active, alco, cardio, and smoke. These qualities were therefore modified appropriately.

D. Performance Evaluation metrics

How well the categorization algorithms employed in this study has been evaluated using various evaluation metrics, including recall, f1-score, accuracy, and precision. model accurately identifies an individual as healthy when they do not have this condition.

A True Positive (TP) implies that the patient has heart disease and that the model correctly classifies the patient as having heart disease, i.e., the model predicts the same outcome with accuracy.

False Positive (FP) means that the model incorrectly identified a healthy person as having cardiac disease even while the patient does not. A type-1 error is another term for this.

A patient with heart disease is misclassified when a False Negative (FN) is present, indicating that they have heart ailment even though the model suggests they do not.

1. Accuracy: One key indicator of a classification model's correctness is its accuracy. Finding the percentage of correct predictions relative to the total number of guesses is the computation. The following is the accuracy formula: TP + TNATrue Negative (TN) in a confusion matrix denotes that the patient does not have the illness under consideration and that the model has accurately predicted their state of health.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

(1)Regarding heart illness, a true negative (TN) implies that the

2. Precision: A quantitative metric called precision is used to evaluate the correctness or accuracy of positive predictions made by a model.

The computation entails splitting the total of true positives and false positives divided by the number of true positives. The precision formula can be written:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

1. Recall (Sensitivity or True Positive Rate): Recall, some- times referred to as sensitivity or true positive rate, is the ca- pacity of a model to accurately detect all pertinent occurrences of a class. The calculation entails dividing the total number of false negatives and true positives by the number of true positives. This is how the recall formula is expressed:

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

2. F1-Score: The F1 score is a comprehensive measure that takes into account both precision and recall, resulting in a single value that encompasses both false positives and false negatives. The computation involves utilizing the harmonic mean of precision and recall, which is determined by the following formula:

$$F1 = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

# II. MODELING

The first step in this stage involves choosing the appropriate modelling technique. Then, the test design is created and the model is builtLastly, an assessment of the model's performance is made. Eighty percent of the dataset is designated as the training dataset, and the remaining twenty percent is designated as the testing dataset. The training data is used to train a model, which is then used to predict outcomes on the testing data. The testing set's actual labels are compared with the predictions made on the testing set to conduct evaluation. The efficiency of the clustered dataset is assessed using a variety of classifiers, such as logistic regression, random forest classifier, and decision tree classifier. Next, each classifier's performance is evaluated effectively classify patients based on their risk levels, enabling tailored clinical therapies for heart disease

**1. Random Forest:**

The Random Forest algorithm is a technique in supervised learning that combines many decision trees to make predictions by voting, in order to address the constraints of individual decision trees. By training each tree using random data samples and characteristics, this strategy improves accuracy and mitigates overfitting. Random forests demonstrate constant performance on large and diverse datasets, even when there are significant missing values. This is possible because random forests can handle different types of data by using decision tree samples. The algorithm's capacity to combine several decision tree outputs allows for strong generalisation and stability in many predictive modelling tasks, such as classification and regression. The Random Forest is a very versatile and dependable machine learning approach that is well-suited for a wide range of real-world applications where accurate predictions are needed.

**2. Logistic Regression:**

The Logistic Regression method plays a crucial role in the field of predictive modelling for heart disease. Logistic Regression differs from decision trees by specifically targeting probabilistic outcomes. It estimates the probability of heart disease occurrence by considering input features. The model accurately represents the correlation between the dependent variable (presence or absence of heart disease) and independent variables, offering a comprehensive comprehension of risk factors [23]. The algorithm's simplicity and interpretability render it highly beneficial in clinical settings, as it provides significant insights into the likelihood of heart disease incidence and assists in categorising patients based on their risk level, hence enhancing the quality of patient treatment

**3. Decision Tree:**

A flexible and interpretable machine learning approach that is frequently employed in the prediction of cardiovascular disease is decision trees. These trees produce a hierarchical structure that resembles a tree by recursively splitting the dataset according to the most informative attributes. Complex decision boundaries and interactions between patient risk factors are well captured by decision trees. The "max-depth" hyperparameter regulates the tree's depth, enabling users to strike a compromise between overfitting and model complexity. Decision Trees are beneficial for identifying crucial risk factors and providing clear insights into the decision-making process. However, they are susceptible to overfitting, which can be mitigated by tuning hyperparameters. Ensemble methods like Random Forests, which consist of multiple Decision Trees, are often employed to enhance predictive performance and generalization.

In cardiovascular risk assessment, Decision Trees offer transparency and the ability to prioritize risk factors, aiding professionals in reaching well-informed decisions specific to each patient's profile. Consideration of interpretability and fine- tuning of hyperparameters are key aspects when leveraging Decision Trees for predictive modeling in cardiovascular health.

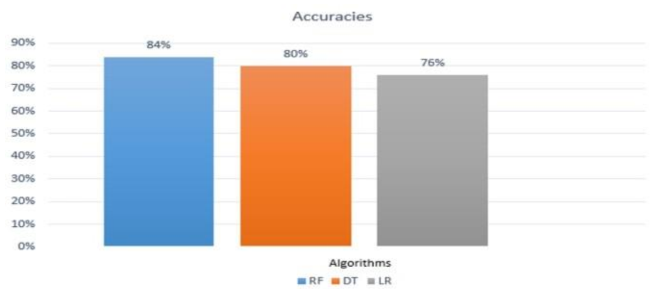


Fig. 2. Accuracies of the Models

## III. RESULTS AND DISCUSSION

The Random Forest (RF) model exhibited the best accuracy of 84(%) among the three classification methods, surpassing both the Decision Tree (DT) and Logistic Regression (LR) models, which attained accuracies of 80(%) and 76(%), re- spectively. The Random Forest's better performance indicates that using an ensemble approach—combining several decision trees—has worked well for identifying complex patterns in the data on cardiovascular risk factors. The Decision Tree model demonstrated its capacity to generate a hierarchical structure that captures complicated decision boundaries, even if it lagged behind the Random Forest model by a small margin and still reached an impressive accuracy of 80(%). With 76(%) accuracy, the easier-to-understand model of logistic regression demonstrated its suitability as a baseline model for cardiovascular disease.

These findings emphasise how crucial it is to take accuracy and model complexity into account when selecting the best algorithm for classifying cardiovascular risk. While the Deci- sion Tree and Logistic Regression models provide insightful analyses of the decision-making process with differing degrees of complexity, the Random Forest's strong performance indi- cates that it is appropriate for jobs requiring precision and generalisation. A thorough grasp of the models' performance in the context of predicting cardiovascular disease will be aided by additional analysis that takes into account evaluation measures other than accuracy as well as potential overfitting or interpretability needs.

## REFERENCES

1. D. Shah, S. Patel, and S. K. Bharti, "Heart disease prediction using machine learning techniques," SN Computer Science, vol. 1, p. 345, 11 2020.
2. Hasan, Ruby, "Comparative analysis of machine learning algorithms for heart disease prediction," ITM Web Conf., vol. 40, p. 03007, 2021. [Online].Available: https://doi.org/10.1051/itmconf/20214003007
3. N. Alotaibi and M. Alzahrani, "Comparative analysis of machine learn- ing algorithms and data mining techniques for predicting the existence of heart disease," International Journal of Advanced Computer Science and Applications, vol. 13, no. 7, 2022.
4. N. Biswas, M. M. Ali, M. A. Rahaman, M. Islam, M. R. Mia, S. Azam, K. Ahmed, F. M. Bui, F. A. Al-Zahrani, and M. A. Moni, "Machine learning-based model to predict heart disease in early stage employing different feature selection techniques," BioMed Research International, vol. 2023, pp. 1–15, 5 2023.
5. J. Azmi, M. Arif, M. T. Nafis, M. A. Alam, S. Tanweer, and G.Wang, "A systematic review on machine learning approaches for cardiovascular disease prediction using medical big data," Medical Engineering Physics, vol. 105, p. 103825, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S135045332200 0741
6. D. Waigi, D. S. Choudhary, D. P. Fulzele, D. Mishra et al., "Predicting the risk of heart disease using advanced machine learning approach," Eur. J. Mol. Clin. Med, vol. 7, no. 7, pp. 1638–1645, 2020.
7. L. Breiman, "Random forests," Machine learning, vol. 45, pp. 5–32, 2001.
8. M. Gietzelt, K.-H. Wolf, M. Marschollek, and R. Haux, "Perfor- mance comparison of accelerometer calibration algorithms based on 3d-ellipsoid fitting methods," Computer methods and programs in biomedicine, vol. 111, no. 1, pp. 62–71, 2013.
9. "Performance comparison of accelerometer calibration algorithms based on 3d-ellipsoid fitting methods," Computer methods and programs in biomedicine, vol. 111, no. 1, pp. 62–71, 2013.

10. K. Vanisree and J. Singaraju, “Decision support system for congenital heart disease diagnosis based on signs and symptoms using neural networks,” International Journal of computer applications, vol. 19, no. 6, pp. 6–12, 2011.

11. S. Angraal, B. J. Mortazavi, A. Gupta, R. Khera, T. Ahmad, N. R. Desai, D. L. Jacoby, F. A. Masoudi, J. A. Spertus, and H. M. Krumholz, “Machine learning prediction of mortality and hospitalization in heart failure with preserved ejection fraction,” JACC: Heart Failure, vol. 8, no. 1, pp. 12–21, 2020.

12 S. Mezzatesta, C. Torino, P. De Meo, G. Fiumara, and A. Vilasi, “A machine learning-based approach for predicting the outbreak of cardiovascular diseases in patients on dialysis,” Computer methods and programs in biomedicine, vol. 177, pp. 9–15, 2019.

13. A. ul Haq, J. Li, M. H. Memon, S. Nazir, and R. Sun, “A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms,” Mob. Inf. Syst., vol. 2018, pp. 3 860 146:1–3 860 146:21, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:56179193

14. A. K. Dwivedi, “Performance evaluation of different machine learning techniques for prediction of heart disease,” Neural Computing and Applications, vol. 29, pp. 685–693, September 2016.

15. S. F. Weng, J. Reps, J. Kai, J. M. Garibaldi, and N. Qureshi, “Can machine-learning improve cardiovascular risk prediction using routine clinical data?” PLoS One, vol. 12, no. 4, p. e0174944, April 2017.

16. N. Louridi, M. Amar, and B. El Ouahidi, “Identification of cardio- vascular diseases using machine learning,” in 2019 7th mediterranean congress of telecommunications (CMT). IEEE, 2019, pp. 1–6.

17. K. Droz˙dz˙, K. Nabrdalik, H. Kwiendacz, M. Hendel, A. Olejarz, A.Tomasik, W. Bartman, J. Nalepa, J. Gumprecht, and G. Y. Lip, “Risk factors for cardiovascular disease in patients with metabolic-associated fatty liver disease: a machine learning approach,” Cardiovascular Dia- betology, vol. 21, no. 1, p. 240, 2022.

18. C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, “Effective heart disease prediction using machine learning techniques,” Algorithms, vol. 16, no. 2, 2023. [Online]. Available: https://www.mdpi.com/1999-4893/16/2/88

19. Saad MN, Jahan E, Ahmed S, Uddin KM. An Ensemble Machine Learning-Based Method to Determine the Presence of Heart Ailment Using Pooled Dataset. In2024 IEEE International Conference on Computing, Applications and Systems (COMPAS) 2024 Sep 25 (pp. 1-6). IEEE.

20. B. Martins, D. Ferreira, C. Neto, A. Abelha, and J. Machado, “Data min- ing for cardiovascular disease prediction,” Journal of Medical Systems, vol. 45, p. 6, 1 2021.

21. A. K. Dwivedi, “Performance evaluation of different machine learning techniques for prediction of heart disease,” Neural Computing and Applications, vol. 29, pp. 685–693, 2018.

22. M. Srivenkatesh, “Prediction of cardiovascular disease using machine learning algorithms,” Int. J. Eng. Adv. Technol, vol. 9, no. 3, pp. 2404– 2414, 2020.

23. S. Ambesange, A. Vijayalaxmi, S. Sridevi, B. Yashoda et al., “Multiple heart diseases prediction using logistic regression with ensemble and hyper parameter tuning techniques,” in 2020 fourth world conference on smart trends in systems, security and sustainability (WorldS4). IEEE, 2020, pp. 827–832.

## Address for Correspondence:

Waseem Javed School of Computing & Information Technology Institute of Management Sciences Peshawar, Khyber Pakhtunkhwa, Pakistan.
@ waseemjavedd@gmail.com
+92-345-9588032

**Received:** March 17, 2024 , **Revised:** June 19, 2024, **Accepted:** Nov 26, 2024

**Contribution of Authors:**

| CONTRIBUTION | AUTHOR ABBREVIATION |
|---|---|
| Conception & Planning of the Research | WJ |
| Active Participation in Methodology | WJ, MO |
| Interpretation Analysis & Discussion | WJ, MO |

**CONFLICT OF INTEREST**

None Declared

GRANT SUPPORT AND FINANCIAL DISCLOSURE

NIL

Authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.