Computing, Information Systems, Development Informatics & Allied Research Journal
Vol. 15 No. 1, 2024 - www.cisdijournal.net

# Performance Evaluation of Machine Learning Model in Predicting Heart Disease Prevalent

**Abiola O. A. Adebisi R. O. Adeyemo A. B.**
[1,2,3]Department of Computer Science
University of Ibadan
Ibadan, Nigeria
**E-mails;** oladimejiarowolo@yahoo.co.uk , abdurrazaq.raheemah@gmail.com
sesanadeyemo2014@gmail.com
[1,2,3]07055535539,08060517262,07016012267

## ABSTRACT

Heart disease stands as the foremost contributor to global mortality in the present era. While a range of medical tests aids in diagnosing various manifestations of heart disease, anticipating its occurrence without resorting to these tests presents a formidable challenge. Machine learning has revealed a potential path for comprehending extensive medical datasets, unearthing concealed insights that elude human observation. Random Forest, Logistic Regression, XG Boost, k-nearest neighbor, and Gradient boosting was used to predict instances of heart diseases based on age and gender in this study. Regardless of the dataset's generalization or gender-specific characteristics, we observed that Random Forest, XG Boost, and K-Nearest Neighbors models exhibit high robustness and minimal sensitivity to hyperparameters. While the prediction accuracy remains consistent among Random Forest, Xgboost, and K-Nearest Neighbors, the crucial features responsible for achieving that accuracy vary among them. This observation indicates that it is important to test different machine learning algorithms when the interest is in prediction and understanding the most important feature(s).

**Keywords:** Machine Learning, Heart Disease, Diagnosis, Predictions, Accuracy

**CISDI Journal Reference Format**
Abiola O. A. Adebisi R. O. Adeyemo A. B. (2024): Performance Evaluation of Machine Learning Model in Predicting Heart Disease Prevalent. Computing, Information Systems, Development Informatics & Allied Research Journal. Vol 15 No 1, Pp 63-72... dx.doi.org/10.22624/AIMS/CISDI/V15N1P5. Available online at www.isteams.net/cisdijournal

## 1. INTRODUCTION

The heart is a crucial part which circulates blood in the body. If it malfunctions, the brain along with several other organs will cease its operation, which can result in death within a few minutes. Heart disease, a formidable public health challenge, claims millions of lives each year and in recent years, data mining techniques have garnered attention as valuable tools for aiding the diagnosis and prediction of heart disease case (Biswas, et. al., 2023). Changes in lifestyle, work-related stress, Age, elevated blood pressure, excess weight, diabetes, metabolic syndrome, family history of heart attacks, not enough exercise, Stress, and poor diet increase the rate of several heart-related diseases. The rapid advancement of technology has revolutionized various fields, and healthcare is no exception. In recent years, integrating machine learning techniques into medical research and practice has opened new avenues for understanding and predicting complex health conditions (Maiga, et.al., 2019).

One such crucial area of focus is the prediction of heart disease prevalence among patients, with specific attention to the variables of gender and age. By harnessing the power of machine learning models, researchers, and healthcare professionals aim to enhance their ability to identify, assess, and potentially mitigate the effect of heart disease by these significant demographic factors. This innovative approach holds the promise of valid and personalized healthcare interventions, which contribute to improved results and a healthier society at large. In this context, this study delves into developing a machine-learning model tailored to heart disease prediction, considering the pivotal variables of gender and age.

Roth et. Al., (2021) reported in his research work that Cardiovascular diseases stand as the foremost source of both economic impact and disease burden globally. These conditions enclose a spectrum of disorders affecting the cardiovascular system. A cardiac condition affects the major blood vessels that nourish the muscles of the heart. These blood vessels develop deposits of cholesterol known as plaque, which restrict the flow of blood to vital areas of the body from Srivastava, et. al., (2022) and the heart itself . Plaque can result in various heart-related diseases if not properly addressed. Heart conditions are often called silent killers, remaining undetected until severe symptoms that endanger life become apparent. The process of diagnosing these illnesses may involve a range of methods such as blood analyses, electrocardiograms (ECGs), or Holter observation.

The vast amount of medical data, spanning various databases, lacks inherent value in isolation. Yet, through integration and analysis using Machine learning, data mining, and, Artificial Intelligence methods, it becomes feasible to extract analytical insights which are not only accurate but also cost-effective. Cardiovascular diseases have risen to become one of the leading contributors to mortality worldwide Projected by the WHO, the total fatalities attributed to CVDs are projected to reach 23.6 million by the year 2030, primarily driven by heart disease and stroke. During the period spanning 2010-2015, the worldwide economic impact of heart diseases was estimated to approach approximately USD 3.7 trillion (Maiga, et.al., 2019).

Heart diseases are rapidly increasing daily, and predicting heart diseases early is alarming and vital. Precision and prompt detection of human heart disease can play a crucial role in averting early-stage heart attacks and enhancing the chances of patient survival. Conventional methods for identifying heart disease are subject to bias and susceptible to differences between examiners. In this context, machine learning techniques prove to be effective tools for identifying and classifying individuals with heart disease and those who are in good health. Cardiovascular disease remains a prominent contributor to global mortality, emphasizing the critical need for accurate and efficient methods to predict its prevalence and risk factors. Over the years, significant progress has been made in understanding the intricate interplay between various demographic factors and the development of heart disease. Among these factors, gender and age have emerged as pivotal determinants that significantly influence an individual's susceptibility to heart-related conditions.

## 2. RELATED WORK

In a world where heart disease stands as a prominent cause of global mortality, Intisar, (2022), delves into the complexities of predicting heart disease, particularly in the absence of extensive medical tests. The study underscores the potential of machine learning to process vast medical datasets, uncovering latent insights that elude traditional observation.

The objective of the study is to examine the utility of machine learning algorithms in forecasting heart disease prevalence, culminating in the development of an optimized predictive model. The performance of the model was evaluated using precision, recall, F1-score, and support. The precision of 0.88, recall of 0.96, and F1-score of 0.92 reflect the efficacy of the model in heart disease prediction, with a support of 46 instances. The field of Artificial intelligence (AI) has significantly impacted healthcare in recent years. Its adoption and application have garnered substantial attention, surpassing other industries. In medicine, (AI) is being utilized for automated diagnostic procedures and continuous patient monitoring, leading to a transformative impact on healthcare delivery and improving patient outcomes (Phan, and Abad, 2022). From analyzing medical images to predicting disease risks, AI continues to shape the future of healthcare. (Chang, et. al., 2022) suggested an Artificial Intelligence Model for Heart Disease Detection Using Machine Learning models. He presented a Python-based application tailored towards healthcare research, highlighting its reliability and versatility in accommodating various health monitoring applications. A pivotal aspect of the research involved the formulation of a random forest classifier algorithm designed to achieve heightened accuracy in identifying heart diseases.

Early Prediction in the Classification of Cardiovascular Diseases (CVD) with Machine Learning, Neuro-Fuzzy, and Statistical Methods was carried out by Taylan, et.al., (2023). Given the complex and nonlinear nature of CVD risk factors, the authors advocated for the use of artificial intelligence tools to detect and classify CVDs effectively. He highlighted the intricate relationships among CVD risk factors, their ill-defined nature, and the potential benefits of using artificial intelligence techniques to enhance prediction and classification. A comprehensive approach that employed machine learning (ML) techniques, neuro-fuzzy methods, and statistical approaches were proposed to predict and classify CVDs.

Early prediction of life-threatening health conditions, particularly heart disease, which has emerged as the primary factor behind mortality over the past decade was proposed by Khan, et.al., (2023). Early detection of heart disease is important for saving lives and preventing complications. Machine Learning was used to improve the reliability and simplicity of heart disease prediction, acknowledging the complexities of predicting a disease with various contributing factors. The research focused on generating and analyzing statistics for each algorithm, including Accuracy, Recall, Precision, and Specificity, to compare their performances. Kasabe, and Narang. (2020) conducted a research that focused on harnessing data mining methods for predicting heart disease. Data mining involves analyzing data from diverse perspectives to derive valuable insights. The approach aids in identifying patterns and risk factors that facilitate accurate disease prediction. Feature engineering was carried out and features pertinent to the heart condition were extracted for classification via machine learning techniques. Key performance metrics like accuracy, precision, and F-measure were computed.

## 3. METHODOLOGY

The methodology involves the data collection, preprocessing of the data collected, identifying and predicting human heart disease using various machine learning models like Random Forest, Naïve-Bayes, KNN, Logistic Regression, Gradient Boosting, and XGBoost. And employing different performance metrics to assess the performance of the heart disease dataset, such as Precision, F1-score, Recall, Confusion matrix, and ROC Curve. Furthermore, Python programming language was utilized for all computational, preprocessing, and visualization tasks to train, validate, and assess the machine-learning algorithms.

Data Collection and EDA

The dataset utilized for the study was acquired from the UCI Machine Learning Repository. It has 6750 instances with 15 features. The collected data are subjected to Exploratory Data Analysis to understand the data. Exploratory Data Analysis (EDA) is the method of utilizing data visualization techniques to gain insights into individual features and their relationships with the target variable. The process started by Importing the necessary libraries like pandas, NumPy, Matplotlib, and Seaborn for manipulating the data, numerical operations, and data visualization. Checking for missing values and data cleaning was performed on the dataset, and categorical data was transformed into numerical variables as shown in Table 3.2.

Table 3.2: Data Transformation

| | Sex | Heart_Disease |
|---|---|---|
| 0 | 1 | 1 |
| 1 | 0 | 0 |
| 2 | 1 | 1 |
| 3 | 1 | 0 |
| 4 | 0 | 0 |

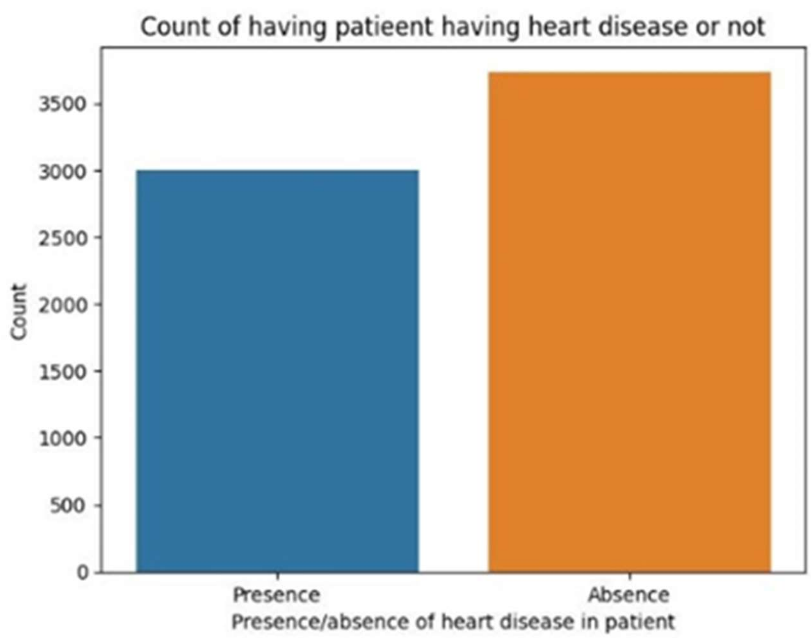


Figure 3.2: Distribution of Heart Disease

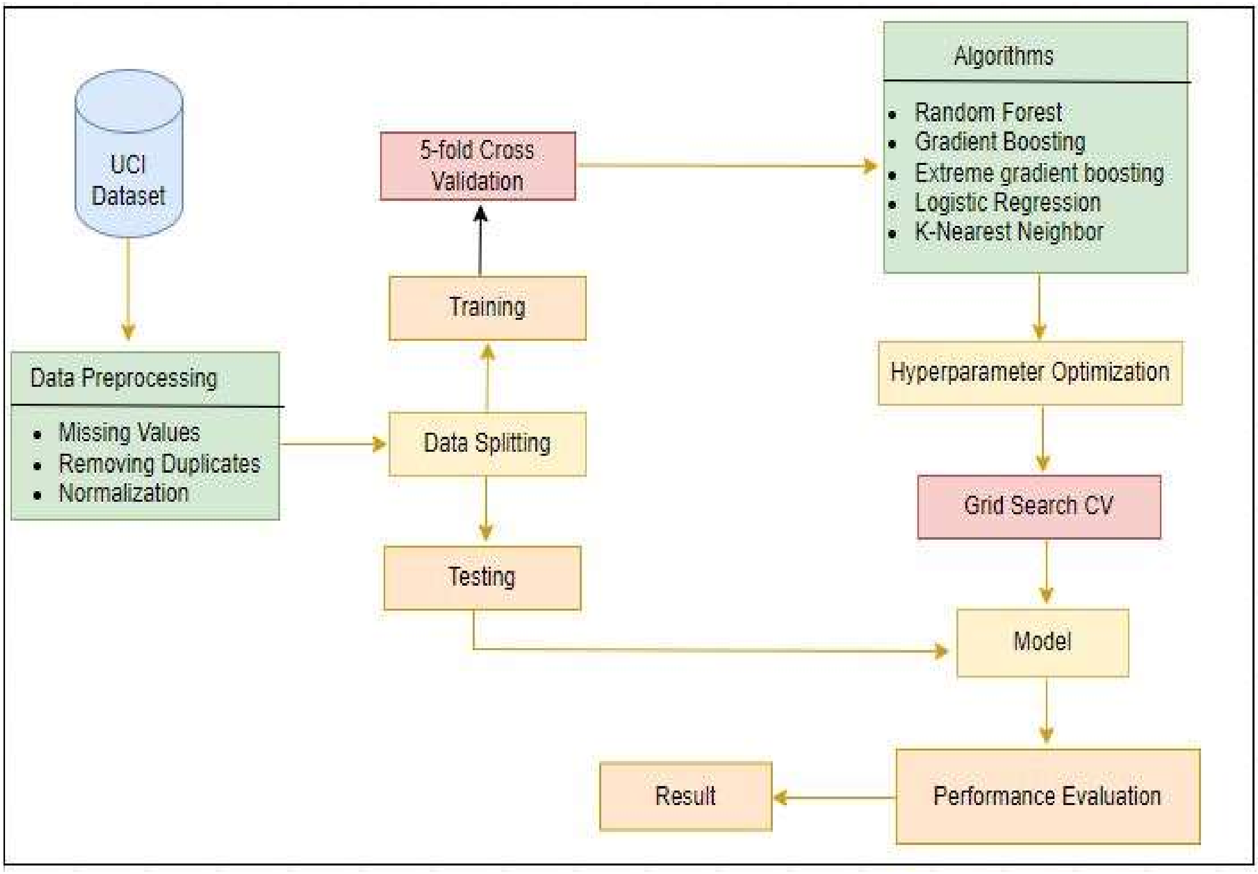


Figure 3.1: Model Framework

### 3.2 Random Forest

Random Forest refers to a machine learning ensemble technique that combines multiple decision trees to form a more effective and accurate model. It uses the principle of "wisdom of the crowd" to improve the overall performance and reduce overfitting compared to a single decision tree. A forest has multiple trees. The strength of a forest is often correlated with the number of trees it contains. As suggested by its name, the random forest comprises a substantial quantity of single decision trees that function collaboratively as an ensemble.

### 3.3 Logistic Regression

A logistic model is a statistical technique employed in scenarios involving binary classification tasks. Its purpose is to predict a binary result variable using one or more predictor variables. Contrary to its name, logistic regression is employed for classification purposes rather than regression. The model is prepared to recognize the ideal parameters that best fit the data and separate the classes by a decision boundary. It's a foundational algorithm in machine learning and is widely used in various fields for tasks like spam detection, medical diagnosis, and more.

### 3.4 K-Nearest Neighbor

K-Nearest Neighbors (K-NN) represents a straightforward yet impactful machine-learning algorithm applicable to both classification and regression assignments. For K-NN regression, the algorithm calculates the average (or another form of aggregation) of the target values of the 'k' nearest neighbors to predict the desired outcome for the newly introduced data point. K-NN relies on the assumption that similar instances in the attribute space are likely to have similar outcomes.

### 3.5 Gradient Boosting

Gradient boosting stands as a widely utilized machine-learning approach employed in both regression and classification assignments. This approach involves utilizing an ensemble technique that merges several modest predictive models, usually in the form of decision trees, to construct a strong and reliable predictive model with enhanced performance. It is a powerful machine learning technique that sequentially builds predictive models by aggregating the predictions of several weak learners. It aims to iteratively correct the errors made by the previous models, gradually improving the model's accuracy. The term "gradient" refers to the optimization process, where the algorithm minimizes a loss function by adjusting the parameters of each successive model based on the negative gradient of the loss about the predicted values.

### 3.6 Linear SVM

Linear SVM is a classification technique designed to look for a hyperplane that best divides different classes in a dataset. Unlike some other algorithms that focus on maximizing the margin between classes, the linear SVM strives to identify a hyperplane that distinguishes the classes while minimizing the misclassification of data points. Linear SVM works by converting the input features into a higher-dimensional space and then finding the hyperplane that best separates the transformed data points. The transformation is done using a linear kernel function. Despite the name "linear," this technique can effectively capture complex relationships in the data.

### 3.7 Naïve Bayes

Naive Bayes constitutes a probabilistic technique for classification grounded in Bayes' theorem. It finds special utility in scenarios like text classification and the analysis of other categorical data types. The "naive" aspect of Naive Bayes refers to the assumption that all attributes are not dependent on one another, given the class. This simplification allows the algorithm to calculate probabilities more efficiently. Naive Bayes computes the likelihood of a specific data point being associated with a particular class by considering the probabilities of its features given that class.
Despite its simplifying assumptions, Naive Bayes often performs surprisingly well, especially in situations where the independence assumption holds reasonably well or when there's a large amount of data available.

## 4. RESULTS AND DISCUSSION

The evaluation results for the model involve using a confusion matrix and various performance metrics such as accuracy, precision, recall, ROC curve, and F1-measure to assess how well the model is performing. The model is built by using Random Forest, K-Nearest Neighbor, Gradient boosting, Linear SVM, Naïve Bayes, and Extreme Gradient boosting.

### 4.1 Result of Random Forest

Random forest obtained an accuracy of 99.99% when trained on the generalized model and when the dataset is split by gender, it also obtained an accuracy of 99.95% for males and 99.89% for females.

It has a ROC curve of 1. The Confusion matrix and ROC curve are shown in Figures 4.1 and 4.2 respectively.

### 4.2 Result of Extreme Gradient Boosting

The extreme gradient boosting algorithm achieved a perfect accuracy of 99.98% when trained on the generalized model. When the dataset was segregated by gender, the accuracy remained exceptional, reaching 99.94% for males and 99.89% for females.
Additionally, the ROC curve for this model exhibits a value of 1, indicative of its strong performance. The figure below displays the detailed classification report.

### 4.3 Result of Logistic Regression

Upon training with the generalized model, the logistic regression algorithm attained an accuracy level of 88.49%. Upon segmenting the dataset by gender, accuracy figures of 87.61% for males and 84.20% for females were achieved. Furthermore, the ROC curve for this particular model demonstrates a value of 0.8, signifying its strong performance. A comprehensive depiction of the classification report's details can be observed in the figure provided below.

### 4.4 Result of Gradient Boosting

After undergoing training using the generalized model, the gradient boosting algorithm achieved an impressive accuracy rate of 99.93%. Upon dividing the dataset by gender, remarkable accuracy rates of 99.92% for males and 99.66% for females were realized. Additionally, the ROC curve associated with this specific model illustrates a value of 1, underscoring its robust performance. The detailed breakdown of the classification report can be thoroughly examined in the provided figure below.

### 4.5 Comparison of model accuracy for generalized models and model by sex group

Table 4.1: Result of the Models Based On A Generalized Model

| | LOGISTIC REGRESSION | RANDOM FOREST | GRADIENT BOOSTING | k-NN | LSVM | XGBOOST | NAÏVE BAYES |
|---|---|---|---|---|---|---|---|
| ACCURACY | 88.49% | 99.99% | 99.93% | 99.97% | 84.86% | 99.98% | 86.86% |
| PRECISION | 0.89 | 1.00 | 1.00 | 1.00 | 0.80 | 1.00 | 0.89 |
| RECALL | 0.90 | 1.00 | 1.00 | 1.00 | 0.97 | 1.00 | 0.88 |
| F1-SCORE | 0.90 | 1.00 | 1.00 | 1.00 | 0.88 | 1.00 | 0.88 |

Table 4.2: Result of the model based on Male

| | LOGISTIC REGRESSION | RANDOM FOREST | GRADIENT BOOSTING | k-NN | LSVM | XGBOOST |
|---|---|---|---|---|---|---|
| ACCURACY | 87.61% | 99.95% | 99.92% | 99.93% | 90.43% | 99.94% |
| PRECISION | 0.89 | 1.00 | 1.00 | 1.00 | 0.90 | 1.00 |
| RECALL | 0.96 | 1.00 | 1.00 | 1.00 | 0.99 | 1.00 |
| F1-SCORE | 0.92 | 1.00 | 1.00 | 1.00 | 0.94 | 1.00 |

Table 4.3; Result of the model based on Female

| | LOGISTIC REGRESSION | RANDOM FOREST | GRADIENT BOOSTING | k-NN | LSVM | XGBOOST |
|---|---|---|---|---|---|---|
| ACCURACY | 84.20% | 99.89% | 99.66% | 99.77% | 69.14% | 99.89% |
| PRECISION | 0.78 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| RECALL | 0.80 | 1.00 | 1.00 | 1.00 | 0.32 | 1.00 |
| F1-SCORE | 0.79 | 1.00 | 1.00 | 1.00 | 0.48 | 1.00 |

## 5. CONCLUSION

Regardless of the dataset's generalization or gender-specific characteristics, we observed that Random Forest, Extreme Gradient Boosting, and K-Nearest Neighbors models exhibit high robustness and minimal sensitivity to hyperparameters tuning. While the prediction accuracy remains consistent among Random Forest, Xgboost, and K-Nearest Neighbors, the crucial features responsible for achieving that accuracy vary among them. This observation indicates that it is important to test different machine learning algorithms when the interest is in prediction and understanding the most important feature(s).

Also, the critical features for diagnosing heart disease differ between males, females, and the generalized model. This suggests that using the entire dataset for prediction may introduce bias especially when the dataset is unbalanced as the case in this study.

## REFERENCES

1. *Ahmed, (2022). A study of heart disease diagnosis using machine learning and data mining (2022). Electronic theses, projects, and dissertations. 1591. Https://scholarworks.lib.csusb.edu/etd/1591*
2. Biswas, N., Ali, M. M., Rahaman, M. A., Islam, M., Mia, M. R., Azam, S., Ahmed, K., Bui, F.M., Al-Zahrani, F. A., & Moni, M. A. (2023). Machine Learning-Based Model to Predict Heart Disease in Early Stage Employing Different Feature Selection Techniques. *BioMed Research International*, 2023, 6864343
3. Maiga, A., Farjah, F., Blume, J., Deppen, S., Welty, V. F., D'Agostino, R. S., ... & Grogan, E. L. (2019). Risk prediction in clinical practice: a practical guide for cardiothoracic surgeons. The Annals of Thoracic Surgery, 108(5), 1573-1582.
4. Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, et al. Global burden of cardiovascular diseases and risk factors, 1990–2019. J Am Coll Cardiol. 2021;77(15) https://doi.org/10.1016/j.jacc.2021.02.039.
5. Srivastava, R., Maji, S., & Panda, T. (2022, December). Machine Learning Techniques in Cardiovascular Disease Prediction. In International Advanced Computing Conference (pp. 199-211). Cham: Springer Nature Switzerland.
6. Chang, V., Bhavani, V. R., Xu, A. Q., & Hossain, M. A. (2022). An artificial intelligence model for heart disease detection using machine learning algorithms. Healthcare Analytics, 2, 100016.
7. Taylan, O., Alkabaa, A. S., Alqabbaa, H. S., Pamukçu, E., & Leiva, V. (2023). Early prediction in the classification of cardiovascular diseases with machine learning, neuro-fuzzy, and statistical methods. Biology, 12(1), 117.
8. Phan, P. H., & Abad, C. L. (2022). *The future of AI in health-care quality: defining the challenges and opportunities moving forward*. International Journal for Quality in Health Care, 34(4), mzac080
9. Khan, A., Qureshi, M., Daniyal, M., & Tawiah, K. (2023). A Novel Study on Machine Learning Algorithm-Based Cardiovascular Disease Prediction. Health & Social Care in the Community, 2023.
10. Kasabe, R., & Narang, G. (2020). Heart Disease Prediction using Machine Learning. *International Journal of Engineering Research & Technology (IJERT)*, *9*(8), 353-355.