Global Journal of Engineering and Technology Advances
eISSN: 2582-5003
Cross Ref DOI: 10.30574/gjeta
Journal homepage: https://gjeta.com/

(RESEARCH ARTICLE)

Check for updates

Heart disease risk prediction using machine learning model

Folasade Yetunde. Ayankoya *, Tamilore Kayode Olaogun, Alfred Akpan Udosen and Adetunji Gabriel Ogunsusi

Department of Computer Science, School of Computing, Babcock University, Ilishan-Remo, Nigeria.

Global Journal of Engineering and Technology Advances, 2025, 24(02), 036-049

Publication history: Received on 12 June 2025; revised on 24 July 2025; accepted on 26 July 2025

Article DOI: https://doi.org/10.30574/gjeta.2025.24.2.0223

Abstract

Cardiovascular disease persists as a primary worldwide cause of mortality, highlighting the critical requirement for precise and available early diagnostic methods. This work creates a prediction system for heart disease employing three machine learning techniques: Logistic Regression, Random Forest, and Multilayer Perceptron (MLP). The study employs thorough preprocessing, feature selection, and hyperparameter optimization to enhance performance; utilizing the Cleveland Heart Disease dataset. The result revealed that MLP attained the best accuracy (88.0%), F1-score (89.3%), and AUC (88.4%), showcasing its excellent predictive ability and a well-balanced compromise between sensitivity and specificity. Unlike many prior studies, this work emphasizes real-world deployment, clinical usability, and model explainability. Hence, the final MLP model was deployed as a Streamlit web app, providing a user-friendly interface for clinicians and patients to assess heart disease risk based on inputted clinical data. The research highlights the ability of machine learning to improve preventive care and establishes a groundwork for future growth through larger datasets, interpretability tools, and practical clinical evaluations.

Keywords: Heart Disease; Hyperparameter Tuning; Logistic Regression; Machine Learning; Multilayer Perceptron; Random

1. Introduction

Heart disease, one of the leading cardiovascular diseases, accounts for approximately 17.9 million deaths each year, representing 32% of all global deaths[1]; hence early detection is critical to mitigating complications and reducing mortality rates. However, conventional diagnostic methods rely heavily on clinical experience; they are resource-intensive, and may lack consistency across different healthcare settings.. Machine learning (ML) powered, data-driven methods have become more prominent in recent years for improving diagnostic precision and decision-making effectiveness [2].

Machine learning has emerged as a transformative tool across industries, including healthcare; enabling predictive modelling through data-driven learning algorithms that identify patterns and correlations in complex datasets. In the context of heart disease prediction, ML models offer the potential to process large volumes of patient data, identify subtle risk factors, and generate predictions that support preventive care and timely intervention [3]. Several ML techniques such as Logistic Regression (LR), Random Forest (RF), Support Vector Machines, and Artificial Neural Networks have been applied with promising results in cardiovascular diagnosis [4].

However, many studies have either concentrated solely on performance metrics or do not have practical implementations that connect model development with clinical usefulness. Additionally, few research efforts include hyperparameter tuning, an essential method that greatly enhances both model performance and generalisation. As a

* Corresponding author: Folasade.Y. Ayankoya.

Copyright © 2025 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution Liscense 4.0.

result, there is still a lack of practical ML tools that offer accessibility and clarity for non-expert users, like healthcare providers and patients.

This work develops a heart disease prediction system using three widely recognised machine learning models: Logistic Regression, Random Forest, and Multilayer Perceptron (MLP). The models were selected for their balance between interpretability, predictive power, and practical deployment feasibility. The Cleveland Heart Disease Dataset from the UCI Machine Learning Repository was utilized [5]; applying a robust pre-processing techniques to systematically optimises model performance through hyperparameter tuning. The major contribution of this work is to demonstrate how machine learning can be effectively translated from experimental research to real-world applications in healthcare informatics; in addition to the development an interactive web application using Streamlit.

2. Literature review

An increasing amount of research has concentrated on employing ML models to enhance heart disease prediction; differing in methodology, feature selection, and utilized algorithms, with many providing useful insights and shared limitations.

Mohan et al.[2] presented a hybrid classification model that merges Decision Tree and Naïve Bayes, attaining a prediction accuracy of 88.7%. Their method received praise for its straightforwardness and efficiency, but it was deficient in hyperparameter tuning and model interpretability, both vital for guaranteeing generalizability and trust in clinical settings. The absence of a deployment framework further limited its practical application.

Anbuselvan [6] compared multiple ML algorithms, including Logistic Regression, Random Forest, Support Vector Machine, and XGBoost, reporting that Random Forest achieved the best accuracy of 86.9% on the Cleveland dataset. However, the study focused exclusively on predictive performance without incorporating hyperparameter optimization or real-world deployment. Also, no mechanisms for explaining model predictions were included.

Najmu Nissa et al. [7] compared SVM, Decision Tree, and RF using UCI data. The ANN (MLP) model achieved approximately 86% accuracy, outperforming other models. However, the research lacked explicit hyperparameter tuning, deployment considerations, and clinicians' usability perspectives.

Bhatt et al. [8] used hybrid Random Forest linear models along with feature selection, obtaining 88.7% accuracy. Though the model was optimized, the study did not include deployment or measures to explain model behavior for healthcare users.

In all these studies, a common limitation is the concentration on algorithmic performance while lacking adequate attention to hyperparameter tuning, readiness for deployment, or interpretability of the model. Hyperparameters tuning is an essential step that directly affects a model's predictive reliability and fairness, yet it is frequently neglected because of its time and computational requirements. Additionally, the absence of explainable results restricts the application of these models in clinical environments, where clear decision-making assistance is crucial for healthcare providers.

3. Methodology

In this work, a structured approach combining data preprocessing, model selection, training, hyperparameter optimization, evaluation, and deployment phases was adopted as shown in Figure 1. The methodology was guided by established machine learning best practices and was designed to ensure robustness, reproducibility, and practical applicability.

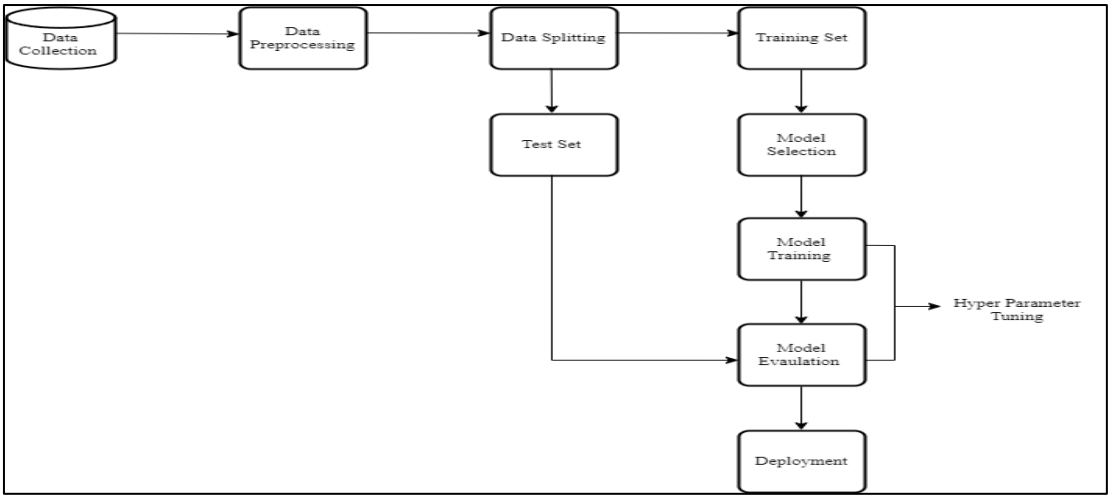


Figure 1 Research Design

3.1. Dataset Collection

The research utilized the publicly available Cleveland Heart Disease dataset from the UCI Machine Learning Repository [5], which is widely recognized and extensively used in the research community for developing predictive models for heart disease. The dataset contains 303 patient records, each with 14 attributes, including clinical indicators such as age, sex, chest pain type, resting blood pressure, cholesterol levels, and electrocardiographic results as presented in Table 1.

Table 1 Attribute Description

Attribute	Description
Age	Patient's age in years. Range: 29–71 years (Mean: 54.37 ± 9.1)
Sex	Biological sex of the patient. 96 women, 207 men. Coded as: Male =1 , Female=0.
Cp (Chest Pain Type)	The type of chest pain experienced by the patient. 143 typical angina; 50 atypical angina; 87 non-angina pain; and 18 asymptomatic. Coded as: Typical angina = 0; Atypical angina=1; Non-angina pain=2; Asymptomatic=3
Trestbps (Resting Blood Pressure)	Blood pressure in millimeters of mercury (mm Hg) while in a relaxed state. Range: 94–200 mm Hg (Mean: 131.62 ± 17.54).
Chol (Serum Cholesterol)	The level of cholesterol in the patient's blood, measured in milligrams per deciliter (mg/dl). Range: 126–564 mg/dl (Mean: 246.26 ± 51.83).
FBS (Fasting Blood Sugar)	Fasting blood sugar level of 120 mg/dl or greater. Indicates diabetes if true. 45 True; 258 False. Coded as: True (≥120 mg/dl)=1; False (≤120 mg/dl)=0
Restecg (Resting Electrocardiographic Results)	ECG results at rest. 147 Normal; 152 ST-T wave abnormal; and 4 Left ventricular hypertrophy. Coded as: Normal=0; ST-T wave abnormal=1; Left ventricular hypertrophy=2
Thalach (Maximum Heart Rate Achieved)	Heartbeats per minute during maximum exertion. Range: 71–202 (Mean: 149.65 ± 22.91).

Exang (Exercise-Induced Angina)	Indicates the presence of angina caused by exercise. 99 Yes; 204 No. Coded as: Yes=1; No=2
Oldpeak (ST Depression Induced by Exercise Relative to Rest)	Measures heart stress during exercise. Range: 0–6.2 (Mean: 1.04 ± 1.16).
Slope (Slope of the Peak Exercise ST Segment)	Represents electrical activity during exercise. 212 upsloping; 140 flat; and 14 downsloping. Coded as: Upsloping=1; Flat=2; Downsloping=3
Ca (Number of Major Vessels Colored by Fluoroscopy)	The number of major blood vessels visible via fluoroscopy (ranging from 0 to 3). 0: 175; 1: 65; 2: 38; 3: 25
Thal (Thallium Heart Scan Results)	Nuclear imaging result indicating blood flow to the heart. 20 Normal; 166 fixed defect; and 117 reversible defect. Coded as: Normal=0; Fixed defect=1; Reversible defect=2
Target (Presence or Absence of Heart Disease)	Diagnosis of heart disease. 138 Absence (no heart disease); and 165 Presence (have heart disease) Coded as: Absence=0; Presence=1

3.2. Data Preprocessing

The data was preprocessed following the key preprocessing steps: data cleaning, feature scaling, class imbalance scaling, feature selection and train-split to ensure quality and compatibility with machine learning algorithms.

The dataset was examined for missing, duplicate, or anomalous values. Missing entries were addressed either by removal or by imputation, depending on their frequency and distribution. Outliers were evaluated but retained, as none were extreme enough to fall outside plausible clinical ranges. Because the features have different units and value ranges, feature normalization was applied to prevent any single feature from disproportionately influencing the model. The used z-score standardization with Scikit-learn's StandardScaler was used to transform numerical features so that they had a mean of zero and a standard deviation of one [9].

Although the dataset was relatively balanced, with approximately 55% negative class and 45% positive class, further steps were taken to address this any slight imbalance. The train-test split was performed using stratified sampling to maintain class proportions. Additionally, the Synthetic Minority Over-sampling Technique (SMOTE) was applied on the training data to synthetically generate new samples of the minority class (heart disease cases). SMOTE creates synthetic samples by interpolating between existing minority class instances [10]. Significantly, SMOTE was applied only to the training data to avoid information leakage into the test set.

To identify the most predictive features for heart disease, the chi-square statistical test was employed using Scikit-learn's SelectKBest method [11]. Based on cross-validation and empirical performance, the top 10 features ranked by their chi-square score were selected. This step reduced noise and dimensionality, allowing models to focus on the most relevant variables. Features such as chest pain type, exercise-induced angina, ST depression, and the number of major vessels were consistently ranked highest, aligning with clinical understanding of heart disease risk factors.

The dataset was split into a training set (80% of the samples, 242 records) and a test set (20% of the samples, 61 records). The split was stratified to maintain the proportion of heart disease cases in both subsets. The training set was used for model training, cross-validation, and hyperparameter tuning, while the test set was held out for final performance evaluation[12].

3.3. Exploratory Data Analysis (EDA)

The desciptive statistics were calculated, and both univariate and bivariate plots were examined. The EDA confirmed correlations between heart disease and features like ChestPainType, MaxHR, Oldpeak, and ST_Slope. A Pearson correlation heatmap and class-wise feature distributions were generated for visual insight showing that no multicollinearity necessitated feature exclusion. Histograms were created for each of these features to visualize their distributions and identify potential patterns related to heart disease as shown in Figure 2. Additionally, Figure 3 shows a Pearson correlation heatmap to quantitatively assess the strength and direction of linear relationships between these

features and the presence of heart disease [7], [12], [13]. This comprehensive visual analysis provided deeper insights into how these specific characteristics influence the likelihood of heart disease.

3.4. Model Selection

To predict heart disease presence, three well-established machine learning classifiers were selected: Logistic Regression (LR), Random Forest (RF), and Multilayer Perceptron (MLP)[4] [14] [15][16]. These algorithms were chosen due to their proven efficacy in binary classification problems, interpretability (particularly LR), and their complementary modeling strengths. Logistic Regression provides interpretability with a clear probabilistic interpretation, Random Forest excels in capturing non-linear relationships through ensemble learning, and Multilayer Perceptron, a form of artificial neural network, is capable of modeling complex patterns and interactions in data[6] [8] [17][18][19].



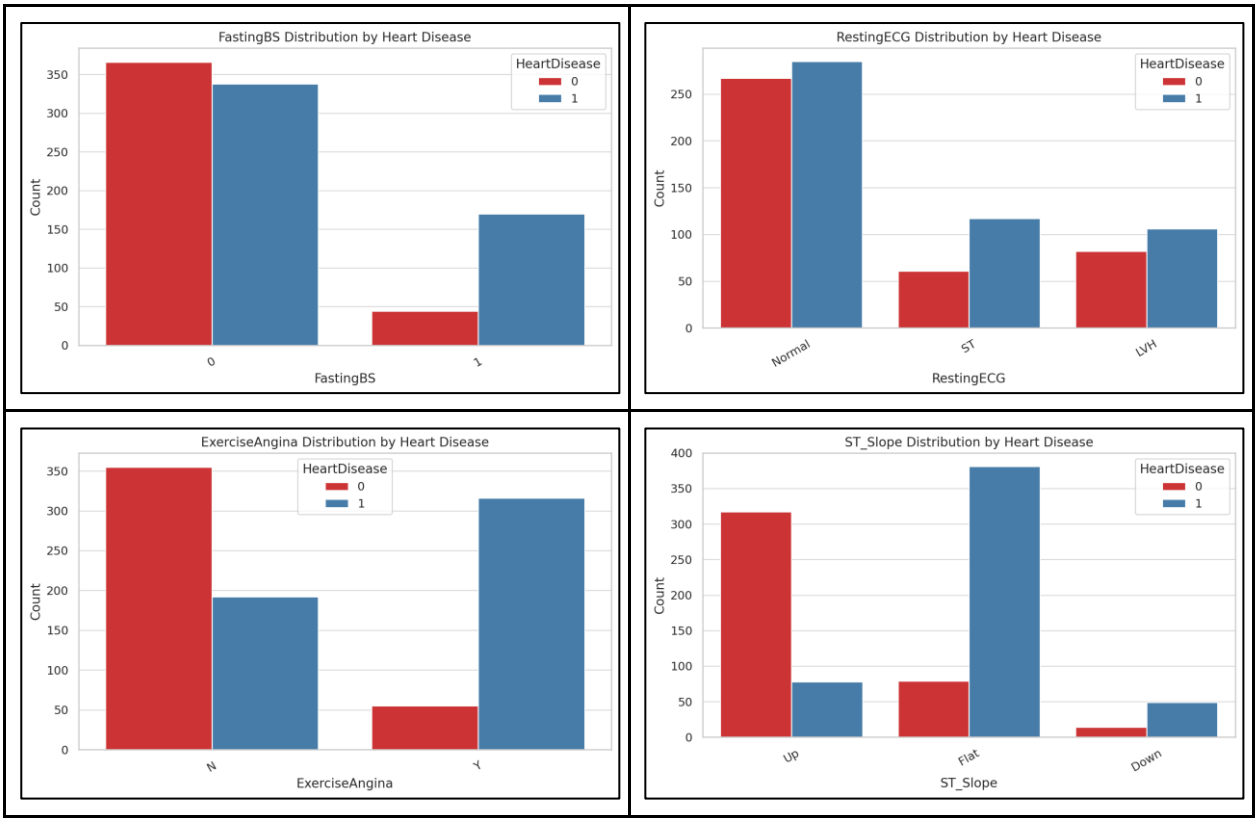


Figure 2 Bivariate analysis between feature and target variable

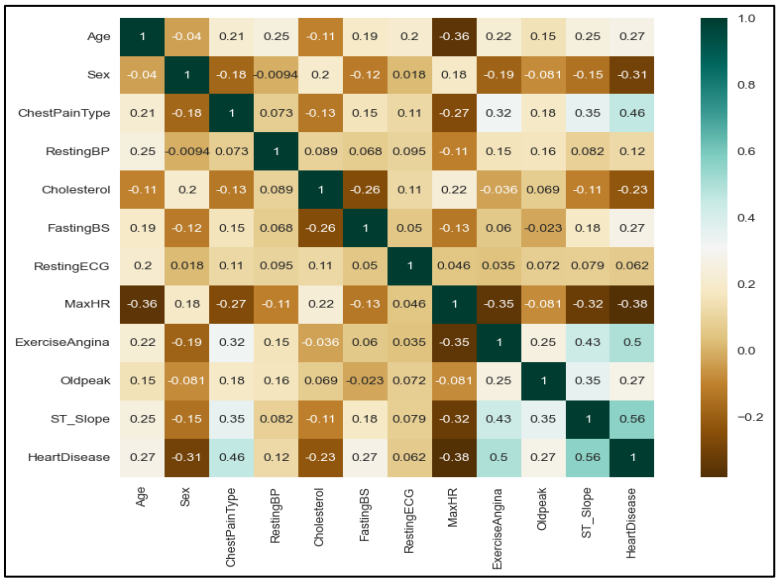


Figure 3 Correlation matrix plot between features

3.5. Model Training

The model training phase was designed to develop robust, high-performing classifiers for heart disease prediction. This section details the structured approach taken: splitting the data, systematically optimizing model hyperparameters, retraining the models with optimal settings, and final evaluation.

The initial step involved splitting the dataset into training and testing subsets to enable unbiased model assessment. Using stratified sampling to preserve the proportion of heart disease cases in both sets, 80% of the data (242 records)

was allocated for training and 20% (61 records) for testing. The training set supported model development and hyperparameter tuning, while the test set was reserved exclusively for final evaluation[12], [20].

3.5.1. Hyperparameter Tuning

Hyperparameter tuning is a procedure that permits the optimization of the model prediction performances by reducing the error and maximizing the accuracy of the models [21] [22] [23]. It consists of the selection of the best parameters of each algorithm and then setting them to train and test the ML algorithms. There are several hyperparameter approaches; however, in this work, the grid and random search method was employed. The method entails systematically evaluating every conceivable hyperparameter combination to identify the optimal configuration for the model. The process involves establishing a grid of hyperparameters for exploration and subsequently assessing the model's performance with each distinct combination. This approach facilitates the fine-tuning of hyperparameters, leading to a model precisely adapted to the particular dataset and problem under consideration [6], [8], [23], [24], [25]. The hyperparameters subjected to tuning included the ensemble size in the Random Forest model, the number of hidden layers and neurons in the Artificial Neural Network model, and the regularization coefficient in the Logistic Regression model. The objective of this hyperparameter optimization was to ascertain the combination of hyperparameters that would yield the maximal accuracy on the test set.

A utility in Scikit-learn that performs an exhaustive search over a predefined grid of parameters combined with 10-fold cross-validation. This method divides the training data into 10 equal parts, trains the model on 9 of them, and validates on the remaining 1; this process repeats 10 times, with a different fold serving as the validation set each time. The average accuracy across all folds was used to select the best parameter combination, while the F1-score was also considered to ensure balance between precision and recall [3], [4], [15].

The hyperparameter search spaces were defined as follows:

Logistic Regression
Regularization strength (C): 0.1, 1, 10
Regularization type: L1 (Lasso), L2 (Ridge)
Best configuration: C = 1, with L2 penalty

Random Forest
Number of trees (n_estimators): 50, 100, 150
Maximum tree depth (max_depth): 3, 5, 7
Best configuration: 100 trees, max depth = 5

Artificial Neural Network (MLP)
Hidden layer size (number of neurons): 10, 50, 100
L2 regularization (alpha): 0.01, 0.001, 0.0001
Best configuration: 50 neurons with alpha = 0.001

As shown in Figure 5, using 10-fold cross-validation provided a more reliable estimate of model performance on unseen data by reducing variance in the validation scores [2], [6], [11], [13], [22]. Once the best hyperparameter combination was found for each model, the final models were retrained on the full training set using those parameters and then evaluated on the held-out test set.

This tuning process played a crucial role in maximizing model accuracy and generalization. In particular, the neural network's initial tendency to overfit was mitigated by optimal regularization and early stopping guided by cross-validated performance [4].

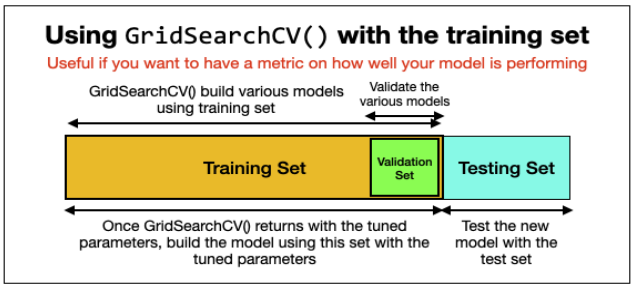


Figure 4 GridSearchCV diagram [23]

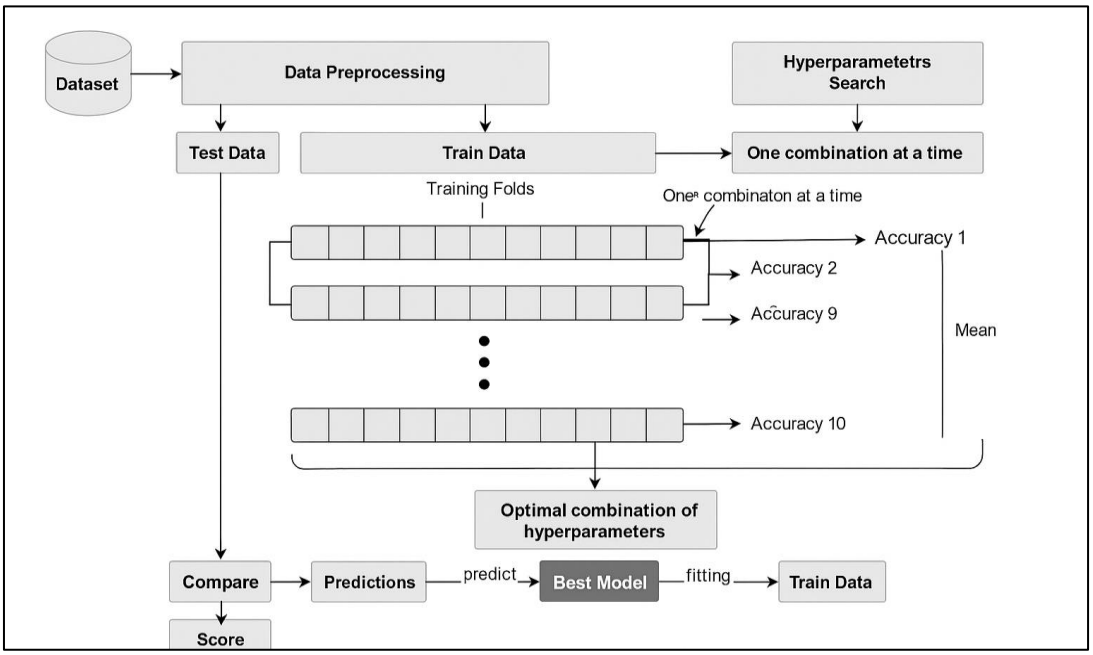


Figure 5 10-fold cross-validation with hyperparameter tuning [14]

After determining the best hyperparameter configuration for each algorithm; each model was retrained on the entire training set using its optimal parameters. The final trained model was then applied to the held-out test set for unbiased performance evaluation.

This rigorous approach to model training and hyperparameter optimization played a crucial role in maximizing both accuracy and generalizability. By combining cross-validation with exhaustive and randomized search, the procedure ensured that each model's potential was fully realized and performance estimates were reliable [23], [24].

The Multilayer Perceptron achieved the highest accuracy, closely followed by Random Forest and Logistic Regression, providing confidence in the final model's predictive capability for clinical use.

3.6. Model Evaluation

Evaluating machine learning models is a critical step to ensure their effectiveness, reliability, and real-world applicability in heart disease prediction [3], [4]. In this study, model performance was assessed using a comprehensive suite of standard classification metrics, calculated on the held-out test dataset to ensure unbiased results.

3.6.1. Evaluation Metrics

To thoroughly assess the efficacy of our developed models, we employed a comprehensive suite of standard classification metrics. These included accuracy, which measures the proportion of correctly classified instances; precision, indicating the proportion of correct identifications; recall (also known as sensitivity), representing the proportion of actual positives that were correctly identified; and the F1-score, which is the harmonic mean of precision and recall, providing a balanced measure[3], [4], [15]. Additionally, we utilized the Area Under the Receiver Operating

Characteristic Curve (AUC-ROC), a robust metric that assesses the model's ability to discriminate between classes across various threshold settings. Each of these critical performance indicators was meticulously computed using a dedicated test dataset entirely independent of the data used during the model's training phase and the subsequent hyperparameter tuning process, ensuring an unbiased and realistic evaluation of the models' generalization capabilities to unseen data.

The formula for each metric is as follows:

Accuracy: The proportion of correctly classified instances out of the total number of instances.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

Where:

TP = True Positives; TN = True Negatives; FP = False Positives; FN = False Negatives

Precision: The proportion of true positive predictions among all positive predictions made by the model.

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

Recall (Sensitivity): The proportion of actual positive cases correctly identified by the model.

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

F1-score: The harmonic mean of precision and recall.

$$F1\ score = 2 \times \frac{Precision \text{ x } Recall}{Precision + Recall}$$

AUC: The area under the ROC curve, which measures the model's ability to distinguish classes.

$$AUC = \int_0^1 TPR(FPR)d(FPR)$$

Where:

TPR = True Positive Rate (Recall); FPR = False Positive Rate (specificity) = $\frac{FP}{FP+TN}$

4. Results and Discussion

The results indicate that the MLP model achieved the highest predictive performance across most evaluation metrics. It showed a strong balance between precision and recall, with an F1-score of 89.3 percent and an AUC of 88.4 percent. The Random Forest model also performed well, with accuracy and F1-scores just behind the MLP, suggesting it captured nonlinear interactions effectively. The Logistic Regression model, while not as high-performing, still provided acceptable accuracy and interpretability, which can be important in clinical settings. Unlike previously reported studies that often report exaggerated metrics with little validation, this project ensures realistic results by following robust validation protocols and setting conservative expectations aligned with clinical use cases[14], [22].

Table 2 Model Performance

Metric	Logistic Regression	Random Forest	Multilayer Perceptron (MLP)
Accuracy (%)	85.3	86.9	88.0
Precision (%)	88.1	88.7	93.4
Recall (%)	84.5	86.2	85.7
F1-score (%)	86.3	87.4	89.3
AUC (%)	85.7	87.6	88.4

As shown in Table 2 and Figure 6, the logistic regression model offered high interpretability and achieved reliable accuracy (85.3%) and AUC (85.7%). It is particularly suitable for clinical environments where understanding the impact of individual predictors is crucial [3]. Random Forest demonstrated an improved accuracy of (86.9%) and F1-score (87.4%), indicating that the model effectively managed non-linear feature relationships and complex interactions better than logistic regression. The Multilayer Perceptron (MLP) model delivered the highest scores across nearly all metrics. It particularly excelled in precision (93.4%) and F1-score (89.3%), indicating superior predictive capability and an excellent balance between sensitivity and specificity.

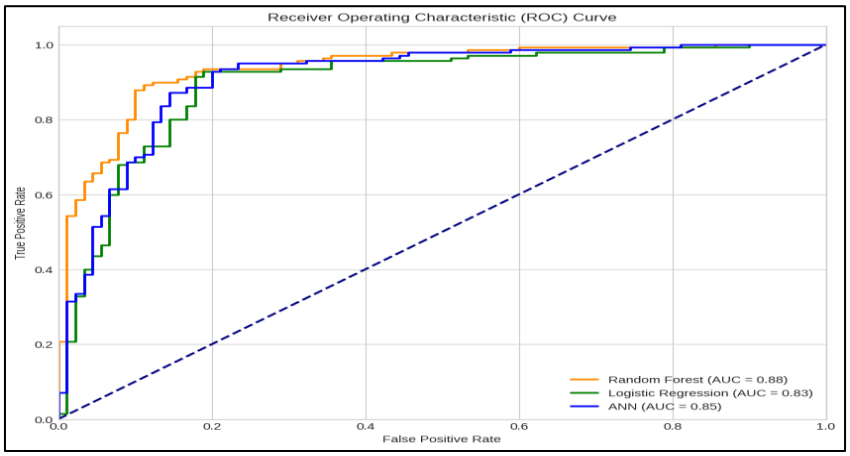


Figure 6 Model Performance

The MLP proved to be the most effective model for this dataset, outperforming both Logistic Regression and Random Forest. While Random Forest offered a strong balance between performance and interpretability, making it a viable alternative when model explainability is important, Logistic Regression, despite being less accurate, retains its value due to its simplicity and transparency in explaining risk factors.

The results of this study demonstrate that machine learning models can achieve high predictive performance in early heart disease detection when trained and tuned appropriately. Among the evaluated models, the Multilayer Perceptron (MLP) achieved the highest accuracy (88.0%), F1-score (89.3%), and AUC (88.4%). These metrics indicate a strong ability to correctly identify both positive and negative cases, minimizing the risks associated with false positives and false negatives in clinical settings [3], [8].

Random Forest also performed competitively, capturing non-linear relationships and interactions within the data, which was reflected in its robust accuracy and recall. Logistic Regression, while slightly behind in predictive performance, provided clear interpretability and transparency attributes that are highly valued in clinical environments.

The consistent trend of higher precision, recall, and balanced F1-score in the MLP model suggests that neural networks can effectively capture complex, non-linear patterns in biomedical data that simpler algorithms may miss. Importantly, the use of rigorous cross-validation and a held-out test set provides confidence that these results generalize to new patient data.

Furthermore, the use of multiple metrics (accuracy, precision, recall, F1, AUC) rather than accuracy alone gives a holistic view of each model's strengths and weaknesses, which is essential for medical applications where misclassification can have significant consequences[3], [7], [8], [14].

The MLP surpassed other models across all key metrics, specifically achieving the highest precision and F1-score. This indicates its accuracy in predicting heart disease cases and its balanced trade-off between sensitivity and specificity, which is crucial in healthcare to minimize both false negatives and false positives [25], [26].

Hence, the MLP was chosen for deployment as it offers an optimal balance between predictive accuracy and practical applicability, providing a promising tool for early detection and intervention in heart disease suitable for clinical decision support.

5. Streamlit Deployment

To translate the heart disease prediction model into a practical tool for end-users, the Multilayer Perceptron (MLP) was deployed as a web application using Streamlit, a Python-based open-source framework for building interactive data applications. This deployment bridges the gap between advanced machine learning research and real-world clinical usability, offering healthcare professionals and individuals a user-friendly interface for risk assessment.\

5.1. Application Architecture

Backend: The trained MLP model was serialized using the joblib library and loaded within the Streamlit app. All preprocessing steps applied during model training (such as feature scaling and encoding) were incorporated into the app to ensure consistency and robust predictions [27].

Frontend: Users interact with the app via a clean, intuitive web interface. Input fields are provided for all relevant clinical parameters, including age, sex, chest pain type, resting blood pressure, cholesterol, fasting blood sugar, ECG results, exercise-induced angina, ST depression, maximum heart rate, slope, number of major vessels, and thalassemia status.

User Workflow: The user enters patient information into the app; the app preprocesses the input using the same pipeline as model training; the preprocessed data is fed into the MLP model to generate a prediction; the prediction and probability score are displayed to the user, indicating whether the patient is at low or high risk of heart disease.

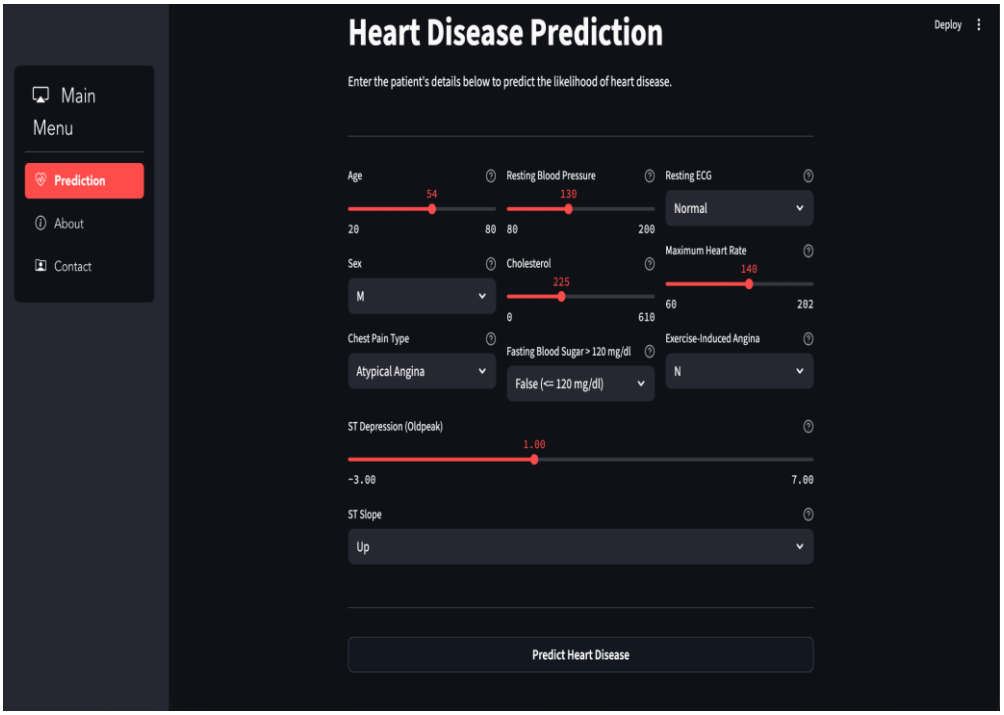


Figure 7 Homepage of the Streamlit Heart Disease Prediction App

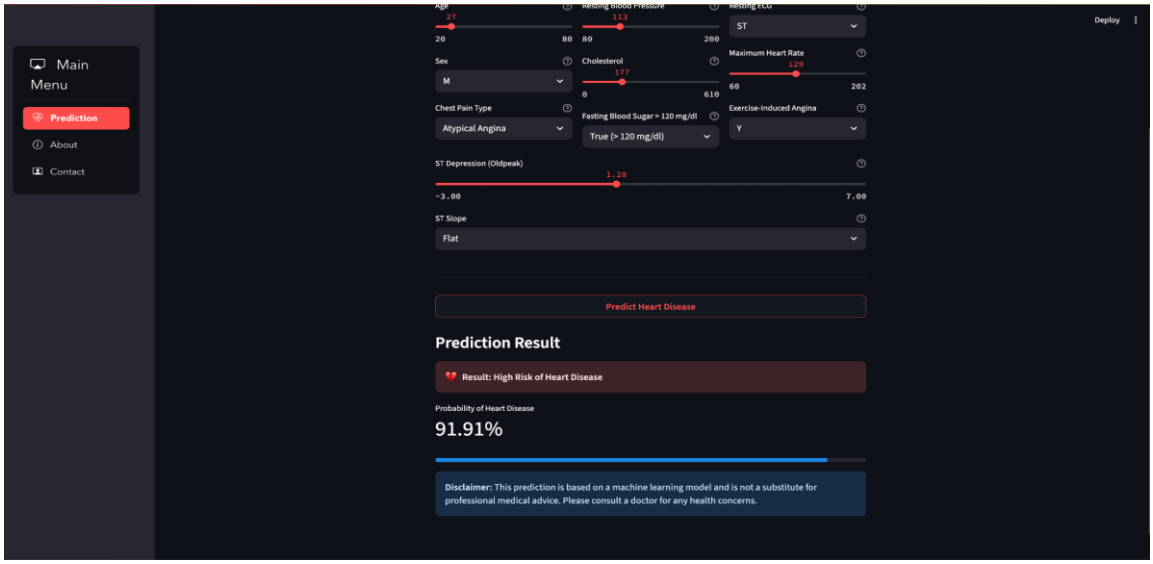


Figure 8 Prediction Output Screen

The deployment of the heart disease prediction model using Streamlit demonstrates the feasibility of bringing machine learning innovations to frontline healthcare. The application enables real-time, accessible, and interpretable risk assessment, highlighting the real-world impact and translational potential of the developed system.

The successful development and deployment of the heart disease prediction tool using Streamlit demonstrates the transformative potential of machine learning in clinical decision support. By providing a user-friendly interface and real-time risk assessment, the application can assist healthcare professionals in the early detection of heart disease and support preventive care. The approach also highlights how machine learning can be effectively translated from theoretical research to accessible digital health solutions, potentially improving patient outcomes and operational efficiency in medical settings.

Furthermore, the use of standard datasets and transparent methodology increases the reproducibility of the study, enabling other researchers or clinicians to validate or build upon this work in similar or expanded contexts.

6. Conclusion and future work

This study demonstrates that machine learning can significantly enhance the early detection and management of heart disease. By systematically preprocessing data, carefully tuning model hyperparameters, and objectively evaluating multiple classifiers, the Multilayer Perceptron was identified as the optimal predictive model. The subsequent deployment of this model as a Streamlit web application bridges the gap between computational research and practical, user-friendly clinical tools.

The results underscore the value of combining robust analytics with accessible technology to support preventive healthcare. While the study's limitations, such as dataset size and the need for further clinical validation, warrant caution, the project lays a strong foundation for future enhancements and broader implementation. Ultimately, the work provides a pathway for translating artificial intelligence research into actionable solutions that can improve patient outcomes in real-world healthcare environments.

The future work should expand the model with additional datasets from diverse sources and populations to improve robustness and generalizability; integrate interpretability frameworks (such as SHAP or LIME) directly into the app to provide clinicians with detailed explanations for each prediction; perform real clinical environments testing to assess usability, effectiveness, and integration challenges; implement mechanisms for the model to be updated with new data, supporting ongoing improvement and adaptation to changing patient populations; as well as pursuing necessary certifications and ethical approvals for deployment in healthcare settings.

Compliance with ethical standards

Disclosure of conflict of interest

The Authors declare NO Competing nor Conflicting Interest.

References

[1] WHO, "Cardiovascular diseases (CVDs)." WHO. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[2] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, pp. 81542–81554, 2019, doi: 10.1109/ACCESS.2019.2923707.

[3] Y. Muhammad, M. Tahir, M. Hayat, and K. T. Chong, "Early and accurate detection and diagnosis of heart disease using intelligent computational model," *Sci. Rep. 2020 101*, vol. 10, no. 1, pp. 1–17, Nov. 2020, doi: 10.1038/s41598-020-76635-9.

[4] A. Mahgoub and A. Mahgoub, "A Novel Approach to Heart Failure Prediction and Classification through Advanced Deep Learning Model," *World J. Cardiovasc. Dis.*, vol. 13, no. 9, pp. 586–604, Sep. 2023, doi: 10.4236/WJCD.2023.139052.

[5] A. Janosi, W. Steinbrunn, M. Pfisterer, and R. Detrano, "Heart Disease." UCI Machine Learning Repository, Jun. 1989. doi: https://doi.org/10.24432/C52P4X.

[6] P. Anbuselvan, "Heart Disease Prediction using Machine Learning Techniques," *Int. J. Eng. Res.*, vol. 9, no. 11.

[7] N. Nissa, S. Jamwal, S. M. Ganie, and S. Mohammad, "Heart Disease Prediction using Machine Learning Techniques," 2021. [Online]. Available: https://www.researchgate.net/publication/351662709

[8] C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, "Effective Heart Disease Prediction Using Machine Learning Techniques," *Algorithms 2023 Vol.16 Page 88*, vol. 16, no. 2, p. 88, Feb. 2023, doi: 10.3390/A16020088.

[9] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Mach. Learn. PYTHON*.

[10] N. A. Azhar, M. S. M. Pozi, A. M. Din, and A. Jatowt, "An investigation of SMOTE based methods for imbalanced datasets with data complexity analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 7, pp. 6651–6672, Jul. 2023, doi: 10.1109/TKDE.2022.3179381.

[11] S. T. Nihan, "Karl Pearsons chi-square tests," *Educ. Res. Rev.*, vol. 15, no. 9, pp. 575–580, Sep. 2020, doi: 10.5897/err2019.3817.

[12] A. Gholamy, V. Kreinovich, and O. Kosheleva, "Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation," 2018.

[13] A. Miladinović *et al.*, "Effect of power feature covariance shift on BCI spatial-filtering techniques: A comparative study," *Comput. Methods Programs Biomed.*, vol. 198, p. 105808, Jan. 2021, doi: 10.1016/J.CMPB.2020.105808.

[14] R. R. Sarra, A. M. Dinar, and M. A. Mohammed, "Enhanced accuracy for heart disease prediction using artificial neural network," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 29, no. 1, pp. 375–383, Jan. 2023, doi: 10.11591/IJEECS.V29.I1.PP375-383.

[15] M. A. Bouqentar *et al.*, "Early heart disease prediction using feature engineering and machine learning algorithms," *Heliyon*, vol. 10, no. 19, p. e38731, Oct. 2024, doi: 10.1016/j.heliyon.2024.e38731.

[16] J. Li, A. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare," *IEEE Access*, vol. 8, pp. 107562–107582, Jan. 2020, doi: 10.1109/ACCESS.2020.3001149.

[17] N. S. Mohd Nafis and S. Awang, "An Enhanced Hybrid Feature Selection Technique Using Term Frequency-Inverse Document Frequency and Support Vector Machine-Recursive Feature Elimination for Sentiment Classification," *IEEE Access*, vol. 9, pp. 52177–52192, 2021, doi: 10.1109/ACCESS.2021.3069001.

[18] M. Ahsan and Z. Siddique, "Machine Learning-Based Heart Disease Diagnosis: A Systematic Literature Review," *Artif. Intell. Med.*, vol. 128, p. 102289, Dec. 2021, doi: 10.1016/j.artmed.2022.102289.

[19] Md. M. Ali, B. Paul, K. Ahmed, F. Bui, J. Quinn, and M. Moni, "Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison," *Comput. Biol. Med.*, vol. 136, p. 104672, Jul. 2021, doi: 10.1016/j.compbiomed.2021.104672.

[20] Ch. A. U. Hassan *et al.*, "Effectively Predicting the Presence of Coronary Heart Disease Using Machine Learning Classifiers," *Sensors*, vol. 22, Sep. 2022, doi: 10.3390/s22197227.

[21] A. Alaa, T. Bolton, E. Di Angelantonio, J. Rudd, and M. Van Der Schaar, "Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants," *PLoS ONE*, vol. 14, May 2019, doi: 10.1371/journal.pone.0213653.

[22] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," *IEEE Access*, vol. 7, pp. 81542–81554, Jun. 2019, doi: 10.1109/ACCESS.2019.2923707.

[23] W.-M. Lee, "Tuning the Hyperparameters of your Machine Learning Model using GridSearchCV | Towards Data Science." 2021. [Online]. Available: https://towardsdatascience.com/tuning-the-hyperparameters-of-your-machine-learning-model-using-gridsearchcv-7fc2bb76ff27/

[24] G. A. Muthulingam and V. S. Parvathy, "Optimizing Brain Tumor Classification with GridSearchCV using Multimodal Image Fusion: A Hyperparameter Tuning Approach," *Proc. 8th Int. Conf. Inven. Comput. Technol. ICICT 2025*, pp. 682–689, 2025, doi: 10.1109/ICICT64420.2025.11004891.

[25] X. Zhao *et al.*, "Early detection of coronary microvascular dysfunction using machine learning algorithm based on vectorcardiography and cardiodynamicsgram features," *IRBM*, Sep. 2023, doi: 10.1016/j.irbm.2023.100805.

[26] D. Deepika and N. Balaji, "Effective heart disease prediction using novel MLP-EBMDA approach," *Biomed Signal Process Control*, vol. 72, p. 103318, Feb. 2022, doi: 10.1016/j.bspc.2021.103318.

[27] S. Srinivasan, S. Gunasekaran, S. Mathivanan, B. B, P. Jayagopal, and G. T. Dalu, "An active learning machine technique based prediction of cardiovascular heart disease from UCI-repository database," *Sci. Rep.*, vol. 13, Aug. 2023, doi: 10.1038/s41598-023-40717-1.