



ARTICLE

A Federated Learning Approach for Cardiovascular Health Analysis and Detection

Farhan Sarwar¹, Muhammad Shoaib Farooq¹, Nagwan Abdel Samee^{2,*}, Mona M. Jamjoom³ and Imran Ashraf^{4,*}

¹Department of Computer Science, School of System and Technology, University of Management and Technology, Lahore, 54000, Pakistan

²Department of Information Technology, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, 11671, Saudi Arabia

³Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, 11671, Saudi Arabia

⁴Department of Information and Communication Engineering, Yeungnam University, Gyeongsan, 38541, Republic of Korea

*Corresponding Authors: Nagwan Abdel Samee. Email: nmabdelsamee@pnu.edu.sa; Imran Ashraf. Email: imranashraf@ynu.ac.kr

Received: 25 January 2025; Accepted: 11 June 2025; Published: 30 July 2025

ABSTRACT: Environmental transition can potentially influence cardiovascular health. Investigating the relationship between such transition and heart disease has important applications. This study uses federated learning (FL) in this context and investigates the link between climate change and heart disease. The dataset containing environmental, meteorological, and health-related factors like blood sugar, cholesterol, maximum heart rate, fasting ECG, etc., is used with machine learning models to identify hidden patterns and relationships. Algorithms such as federated learning, XGBoost, random forest, support vector classifier, extra tree classifier, k-nearest neighbor, and logistic regression are used. A framework for diagnosing heart disease is designed using FL along with other models. Experiments involve discriminating healthy subjects from those who are heart patients and obtain an accuracy of 94.03%. The proposed FL-based framework proves to be superior to existing techniques in terms of usability, dependability, and accuracy. This study paves the way for screening people for early heart disease detection and continuous monitoring in telemedicine and remote care. Personalized treatment can also be planned with customized therapies.

KEYWORDS: Heart disease prediction; medical data; federated learning; machine learning

1 Introduction

Cardiovascular disease is among the leading causes of death around the world. People with heart conditions often have irregular blood circulation, which can cause heart failure [1]. The estimates of the World Health Organization (WHO) indicate approximately 17.9 million deaths due to cardiovascular diseases each year of which 80% are due to coronary artery disease and strokes, with low- and middle-income countries being the most affected. Climate change is a major global issue, affecting our lives in several ways. One serious concern is how it might impact heart health. Changing climate conditions can lead to more cases of heart disease and make existing cases worse. Factors like extreme temperatures and air pollution, which are linked to climate change, can increase the risk of heart problems and other negative health effects.

Climate change leads to many environmental changes that have a negative influence on the heart and worsen heart disease. Extreme heat, air pollution, allergens, and asthma are only a few among many factors



that are linked to climate change and can potentially affect the heart. With the rise in global temperatures, heat waves have become common and intense. Such excessive heat can put extra strain on the heart, raise blood pressure, and elevate existing heart conditions. It can also lead to heat-related illnesses, especially in people who already have heart problems [2]. Climate change is also linked to air quality, increasing harmful pollutants like fine particulate matter (PM_{2.5}), ozone, and nitrogen dioxide. These pollutants are reported to elevate the risk of heart disease [3]. Rise in temperature and rainfall patterns can also affect the spread of allergens like pollen and mold. These allergens can lead to several allergic reactions and breathing problems. It can worsen certain heart ailments, particularly in people with asthma and similar respiratory conditions [4]. Machine learning (ML) uses advanced algorithms to uncover hidden patterns and relationships in data, which can help predict and prevent heart disease, especially as climate change affects health risks [5]. Recent progress in artificial intelligence (AI) has also helped doctors diagnose patients faster and treat them more effectively [umer2023heart]. For instance, the study in [6] shows how ML can analyze large sets of patient data to make accurate predictions about heart disease.

Federated learning (FL) is a modern ML method that keeps data secure by not moving it from its source. Instead, users train their models locally using their data. These locally trained models are then sent to a central server, where they are combined to create a more accurate global model [7]. This process helps solve common issues with traditional machine learning, such as privacy concerns, security risks, and data-sharing limitations since the data never leaves the organization's control. Because FL only transfers small model updates instead of full datasets, it also reduces network traffic and system load [8]. In healthcare, where privacy and data security are crucial, FL has become a valuable tool for building safe and private models, especially for predicting heart disease [9]. As the world increasingly focuses on fighting climate change, combining FL with climate research could lead to powerful and proactive ways to protect heart health from environmental risks. This study presents a new method for identifying people at risk of heart disease using FL. It's especially well-suited for handling sensitive health information since it allows data to be processed locally in compliance with privacy laws.

FL is advantageous over traditional centralized learning, particularly in the case of sensitive applications. Centralized systems are prone to privacy issues, data security breaches, and regulatory compliance [10]. On the contrary, FL uses local devices and only shares encrypted model updates, thereby ensuring increased security and privacy [11,12]. FL is more suitable for healthcare and medical applications due to enhanced privacy, trust, and security [9]. In addition, due to the diverse range of real-world applications, it has better model generalizability and reduced demographic bias [13]. Concerning communication, it reduces bandwidth demands because it only transmits model parameters, which also improves computation capabilities [14]. Personalized models are also possible using FL, where the model is locally fine-tuned for a better fit of local data [15]. In short, FL offers a privacy-first, scalable, and highly effective alternative to traditional centralized methods, unlocking the ability to build smarter, fairer, and safer AI systems without compromising user trust.

The main goal here is to find a reliable and cost-effective way to identify those suffering from heart disease. The proposed approach is important because its accessible, practical, and can play a key role in timely detecting and preventing the disease. This approach focuses on applying proper preprocessing techniques, and exploring both supervised and semi-supervised learning models to improve heart disease prediction.

This study introduces a new method based on FL that combines the strengths of several ML algorithms while maintaining high accuracy. The approach uses FL to improve data access, speed up data processing, and ensure privacy and security. Initially, three ML models were trained on individual computers and then sent to a central server for further optimization. This helped build a global model capable of accurately identifying people at risk of heart disease. The study used various ML techniques such as FL, k-nearest

neighbor (KNN), XGBoost, support vector classifier (SVC), extra trees (ET) Classifier, random forest (RF), and logistic regression (LR) to predict whether or not a person has heart disease.

This article is formatted in the following way. In [Section 2](#), we talk about the related work on heart disease detection. In [Section 3](#), the content and methodology of this article are briefly reviewed. [Section 4](#) presents a summary of the experiment's results, as well as, a discussion of the results. The conclusion is provided in [Section 5](#).

2 Related Work

Heart disease is one of the main causes of death. Patients with this illness frequently develop heart failure as a result of inadequate blood circulation. To foresee the onset of heart disease, Araujo et al. [16] present an ML model. The data from the UCI repository was used including a 918-case, 12-feature dataset on cardiovascular disease. Using the Python scikit-learn package and the free Weka environment for knowledge analysis tools, the authors applied several ML models to the dataset. Ten-fold cross-validation is used to confirm these results. Ensemble approaches give higher accuracy than stand-alone classifiers, even though Weka and scikit-learn in Python produced relatively identical results. In terms of accuracy, KNN, and LR produced the best results, whereas Bagging ($i = 100$, RF) in scikit-learn and Python provided the best results with an 85.36% accuracy while using Weka, stacking showed the best accuracy of 87.24%.

Ansarullah et al. proposed a model in [17] for the initial prediction of cardiac disease by integrating multiple ML models. The dataset contains 5776 records, of which 3031 (52.5%) are healthy and 2745 (47.5%) have heart disease. The study develops five ML classifiers for cardiovascular disease prediction including decision tree (DT), Naive Bayes (NB), support vector machines (SVM), RF, and KNN. The model with the highest AUC-ROC of 0.85 is the RF model which shows that RF is very good at distinguishing between sick and healthy patients.

The study [18] proposed an approach that used a genetic algorithm in conjunction with an RF classifier to predict heart failure. The model uses 918 records with 12 features to provide a diagnosis of heart failure. The data used to build the model was sourced from the Kaggle platform. RF, DT, and NB models were used to develop the model in the Google Colab Notebook. The second iteration has a far higher F1 score (0.90789) than the DT, single RF, and NB (all of which have scores of 0.85034). In addition, it is reported that the best recall was achieved by a single algorithm, RF. The first result of 0.89937 is the most reliable. These findings suggest that the modified RF model might be used for an accurate prediction of cardiovascular arrest.

A stacking model was proposed to predict cardiovascular disease by Liu et al. [19]. The study made use of the Heart Dataset, a collection of records from various clinics and hospitals in Cleveland, Hungary, Sweden, Long Beach, VA, and Stalog (heart) that may be used to approximately estimate a person's risk of cardiovascular disease and heart attack. The Heart Dataset was normalized and divided 80:20 for training and testing, respectively. Eleven features were chosen from the five datasets for training. Several ML models have been developed to predict heart attacks and cardiovascular diseases including KNN, SVM, LR, RF, ET, gradient boosting (GB), DT, XGBoost, LightGBM, Catboost, multilayer perceptron (MLP), and stacking. Accuracy in cardiac settings improved to 89.86% with stacking and in cardiac events to 84.625%.

An ML approach comparing several algorithms was created by Pooja Anbuselvan [20]. The Rapid Miner tool performed more precisely than both Matlab and Weka. The authors employed a range of data mining and ML approaches to assess large datasets and aid in the precise prediction of heart disease. The supervised learning models LR, NB, SVM, KNN, DT, RF, and the ensemble technique are investigated to provide a performance comparison. In comparison with other algorithms, RF is shown to provide the best accuracy at 86.89%. Along the same lines, the study [21] looked into several ML models to categorize cardiac conditions.

The research was done to find out how accurate ML algorithms are at predicting heart disease, using the UCI repository dataset. Among SVM, KNN, linear regression, and DT, KNN showed the highest accuracy. It was also deduced that it could be made more effective by combining multiple models and fine-tuning its parameters.

An ensemble approach is presented in [22] where SVM is coupled with a convolutional neural network (CNN) and optimization is carried out using brave-hunting optimization (BHO). The CNN is adopted for feature extraction while SVM optimized using BHO is used for classification. An accuracy of 94.89% is reported for the proposed approach. Similarly, the study [23] makes use of an artificial neural network (ANN) for feature sensing and RF for classification. The study also utilizes a dedicated sensor to get sensory data from individuals. Experimental results from extensive experiments indicate an accuracy higher than 90.0% for the proposed approach.

Along the same directions, reference [24] leveraged various ML and DL models to formulate an ensemble approach. The ensemble model uses gated recurrent unit (GRU) and LeNet models. In addition, the explainable AI model SHAP is also used to explain model outcomes. The study reported an 88% accuracy with a higher precision of 91%.

The results of previous research on the outlook for patients with heart illness are summarized below, and a comparison of the different researcher's results is shown in Table 1.

Table 1: Comparison of the heart disease results

References	Year	Algorithms	Accuracy
Araujo et al. [16]	2022	LR, Stacking (KNN, LR), DT, KNN, RF, SVM	84.27, 87.24, 85.06, 85.28, 85.23, 85.01
Ansarullah et al. [17]	2022	DT, KNN, SVM, NB, RF	81.0, 70.0, 82.0, 84.0, 69.0
Siregar et al. [18]	2022	RF, DT, NB	F1 score, RF at second iteration has the highest F1 score is 90.7
Liu et al. [19]	2022	KNN, SVM, LR, RF, ET, Stacking	71.74, 72.46, 88.41, 87.68, 88.04, 89.86
Pooja et al. [20]	2020	KNN, SVM, LR, RF, DT, NB, XGBoost	58.83, 73.77, 75.41, 86.89, 77.05, 77.05, 78.69
Singh et al. [21]	2020	SVM, DT, LR, KNN	83.0, 79.0, 78.0, 87.0
Naeem et al. [23]	2024	RF	90.2%
Khan et al. [24]	2024	GRU, LeNet, Proposed	85.0%, 84.0%, 88.0%
Tambe et al. [22]	2025	Ensemble (CNN+SVM)	94.89%

From the discussed literature it is observed that ML models have been the first choice of researchers for cardiovascular disease detection. It is found that KNN, SVM, LR, DT, and RF are widely utilized in this regard. In addition, stacking models comprising KNN and LR are also used for better accuracy, as shown in [16] which reported an 87.24% accuracy using the stacking model. Similarly, reference [19] showed an 89.86% accuracy with the stacking model. The study [17] reports the best accuracy of 84.0% with the NB model while reference [25] shows an 85.0% accuracy using SVM. These studies show the potential of using ML for automated detection of cardiovascular detection.

3 Materials and Methods

The fundamental goal here is to create a system that, using a database of heart disease, can forecast heart disease in response to climate change. Medical attributes like fasting blood sugar, cholesterol, electrocardiogram, sex, and seven more likely variables are used to develop a similar framework. Several ML algorithms are applied to this dataset. Fig. 1 illustrates the workflow of the adopted methodology.

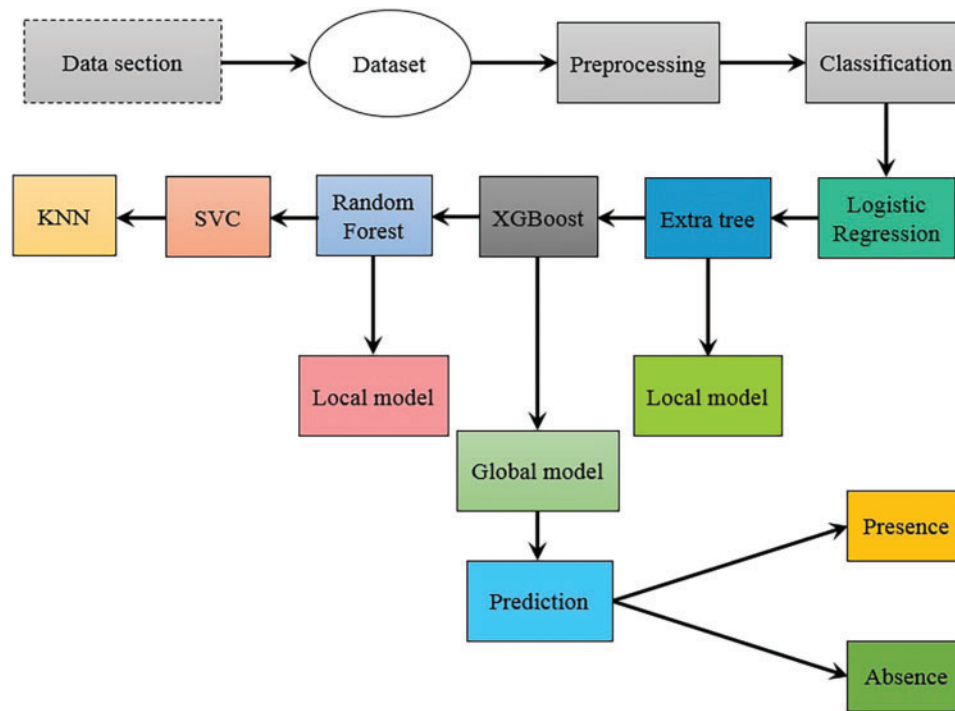


Figure 1: An illustration of a flow chart utilizing ML methods

3.1 Dataset

The dataset, which was collected using Kaggle's [26], has both discrete and continuous characteristics. Sex, the nature of the chest pain, fasting blood pressure, exercise-induced angina, old peak, and ST slope are among the distinctive aspects listed. Your age, resting blood pressure, lipid levels, maximum oxygen uptake, and cholesterol levels are all numbers in the same category. The '1' stands for men, whereas the number 0 is for women. The data set contained 508 records that matched the alternative and 410 records that matched the null hypothesis. The dataset has two categories and 12 characteristics. Further details of the dataset are given in Table 2.

Table 2: Description of the dataset

Feature	Type	Description
Age	C	Age in years [27–77]
Sex	D	0 = female, 1 = male
Chest pain type	D	Chest pain type: 0 = asymptom (ASY), 1 = atypical angina (ATA), 2 = non-anginal pan (NAP), 3 = typical angina (TA)

(Continued)

Table 2 (continued)

Feature	Type	Description
Resting BP	C	Resting blood pressure in (mm Hg) [0, 200]
Cholesterol	C	Cholesterol in mg/dl [0, 603]
Fasting BS	D	Fasting blood sugar normal [100–125], 1 = true, 0 = false
Resting ECG	D	Resting electrocardiogram: Left ventricular hypertrophy (LVH) = 0, Normal = 1, ST = 2
Max HR	C	Maximal heart rate [60, 202]
Exercise	D	N = 0, Y = 1
Angina		
Old peak	D	Old peak = [−2.6, 6.2]
ST Slope	D	The slope of the peak exercise segment: flat sloping = 1, up sloping = 2, down sloping = 0
Heart disease	D	Heart disease prediction: No = 0, Yes = 1

C shows ‘Continuous’, and D shows ‘Discrete’ type variable

3.2 Dataset Preprocessing

In the data mining process for disease prediction and diagnosis, data quality is crucial because low-quality data can lead to incorrect predictions. The preprocessing was applied to clean up the data and increase the predicting accuracy of models. Preprocessing includes tasks like cleaning and normalizing. Because different measuring units are employed, the data normalization is applied to normalize values between 0 and 1.

The dataset consists of 918 rows and 12 columns, and we have two types of data: numerical and categorical. The dataset is discrete, and continuous, and presents a classification challenge. First, we encoded the categorical data and then checked the null values. After that, checking the outliers and replacing them with the mean values is carried out. The division of the category into two classes is done afterward (0-non-heart disease, 1-heart disease). We divide the dataset into dependent and independent variables. The target of the dataset is to predict heart disease.

3.3 Choice of ML Models over Deep Learning

The choice of selected ML models is based on several factors. RF and ET classifiers are selected because they perform well on data related to lab values, like structure data or tabular data. RF can work well with smaller datasets contrary to most DL models like deep neural networks (DNNs) that require larger datasets. Similarly, the ET classifier performs the best when used with smaller and medium-sized datasets.

RF is easy to interpret and less sensitive to hyperparameters while ET requires minimal hyperparameter tuning. In addition, both RF and ET are faster to train and test and robust to noise and outliers. In the case of the imbalanced dataset, RF and ET are better choices than DL models. DNNs, on the other hand, offer better results with natural language text and images and work well with larger datasets where RF and ET are not suited well. While RF and ET classifiers work similarly, the ET classifier differs concerning randomness in splitting. ET classifier uses more randomness which provides lower variance, thereby leading to faster training than RF.

XGBoost also provides better results on tabular datasets and requires minimal training compared to DL models. Even with fewer resources, XGBoost can offer better outcomes. It is also robust to missing values and outliers and works well even with imbalanced datasets.

3.4 ML Models

ML models categorize/classify data into various classes based on their attributes or qualities related to a particular class. In the discipline of supervised learning, it is essential to train a model using a labeled dataset to identify patterns and make predictions on unseen data. The model learns to recognize correlations between input features and target labels, allowing it to generalize and classify new cases properly. There are different types of models such as DTs, SVM, LR, and FL, each with its own set of pros and limitations. The choice of a particular model depends on the nature of the data and the task at hand [27].

3.4.1 Logistic Regression

A popular supervised ML model is LR which is called a sibling of linear regression. LR addresses the binary classification problem by applying a logistic function to linear regression predictions and translating them into binary classification's discrete values (i.e., zero and one). It uses the following sigmoid function [28].

$$Z = \frac{1}{1 + e^{-z}} \quad (1)$$

where Z is the sigmoid function.

3.4.2 Extra Tree

The ET model is an ensemble model that works similarly to RF. Similar to RF, it is based on multiple DTs, however, it offers more randomness than RF. Instead of using sampling with the replacement used by the RF, it uses the whole dataset. In addition, it follows a random split approach, contrary to RF's best-split approach. Due to random split, it can work fast compared to RF. Moreover, it has less susceptibility to overfitting and it can perform well with high-dimensional datasets [29]. ET is used for feature selection. Feature selection is the process of automatically choosing the best attributes. Processing unnecessary features can decrease the model's accuracy and lengthen processing time. The use of the ET model for feature selection resulted in a considerable reduction in the classification execution time in this study.

3.4.3 Random Forest

RF is one of the often-used techniques for accurately classifying high volumes of data. The core concept is to join weak classifiers to create a powerful classifier. Using this ensemble learning strategy for generating models, a large number of DTs are produced during training. Then the final class is predicted based on the predictions of these individual DTs [30]. It consists of multiple tree predictors, depending on the results of independent sets of values from a random vector that has the same distribution among all the trees in the forest [31]. While training a group of classifiers, the following equation provides a concise description of RF.

$$D = (x_i, y_i)^n_i \quad (2)$$

Entropy is give by Eq. (3), where $H(S)$ stands for entropy:

$$H(S) = \sum_{x \in X} p(X) \log_2 \frac{1}{p(x)} \quad (3)$$

How to calculate information gain is shown in the following equation, where $IG(S, A)$ is the information gain.

$$IG(S, A) = H(S) - \sum_{i=1}^n p(X) * H(X) \quad (4)$$

3.4.4 K-Nearest Neighbor

KNN is a data classification model that prioritizes the most common label by selecting the K closest neighbors to the test sample based on the average distance between the test sample and all of the training data [32].

$$d(x, y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2} \quad (5)$$

where x and y are two points in the Euclidean n space and x_i and y_i are the Euclidean vectors beginning at the space origin.

This approach is widely used in classification rather than regression due to its simplicity and speed [6]. To use KNN, the first critical parameter is the k value. The second parameter is the distance function, such as Euclidean, Manhattan, and Minkowski. KNN has several advantages, including simple to implement and simple to interpret. However, for KNN to be effective, the dataset must be constrained; otherwise, computational time will be high [33].

3.4.5 Support Vector Classifier

The SVC is another popular model that can be used to classify heart patients. SVC builds many hyperplanes to separate the classes and then selects the one with the largest distance between positive and negative data points [34]. Class assignment for the new sample is based on a distance function.

$$y = b + w_1 * x_1 + w_2 * x_2 \quad (6)$$

where b is the slope, w_1 are the weights in the coefficient are identified, and x_1 below are the input variables.

Once constructed, the hyperplane can be used to make predictions. The following is how the hypothesis function h is expressed.

$$h(x_i) = \begin{cases} +1 & \text{if } w \cdot x + b \leq 0 \\ -1 & \text{if } w \cdot x + b < 0 \end{cases} \quad (7)$$

3.4.6 XGBOOST

The gradient boosting model known as XGBoost has been developed to be very effective, adaptable, and portable. It offers scalability with higher speed without losing performance. It incorporates L1 and L2 regularization layers to reduce overfitting. It works sequentially where new trees are improved using the errors of previous trees. Optimization is carried out using parallel processing, tree regularization, missing value management, and pruning to avoid bias or overfitting [20].

3.5 Hyperparameter Selection

Finding the best hyperparameter is essential in ML models to obtain optimal results for heart disease detection. In this study, each model is optimized using a set of parameters. The range of values for these

parameters was selected from the existing studies on the same topic and then optimized using the Grid search method. Table 3 provides the details of the hyperparameter for all ML models.

Table 3: Details of models' parameters

Model	Hyperparamters
LR	C = 1.0, penalty = 'l2', solver = 'liblinear', max_iter = 1000
ET	n_estimators = 100, min_samples_split = 2, min_samples_leaf = 1, max_features = 'sqrt', bootstrap = False
RF	n_estimators = 200, max_depth = 10, min_samples_split = 5, min_samples_leaf = 4, max_features = 'sqrt', bootstrap = True
KNN	n_neighbors = 5, weights = 'uniform', algorithm = 'auto', leaf_size = 30, distance = Euclidean
SVC	C = 1.0, kernel = 'rbf', gamma = 'scale', probability = True, class_weight = 'balanced'
XGBoost	n_estimators = 100, learning_rate = 0.1, max_depth = 4, min_child_weight = 1, gamma = 0, subsample = 0.8, colsample_bytree = 0.8, scale_pos_weight = 1, use_label_encoder = False, eval_metric = 'logloss'

3.6 Federated Learning

The FL approach to ML has attracted a great deal of interest from the healthcare industry. Because it allows training ML models without disclosing identifying information about specific patients, it has proven to be particularly useful for speedy responses in the medical sector. Healthcare personnel need to be able to construct and improve ML models, but data security and privacy are important concerns. By keeping the data on local devices, FL ensures that patient privacy is maintained [8,35].

FL offers enhanced privacy by training ML models on local devices which are decentralized units. This way, data transmission to a central server is not needed which enhances security. In addition, processing the data at local devices mitigates threats to data breach and unauthorized access [11]. In the case of FL, only model updates are communicated which enhances data protection. FL makes use of techniques like secure aggregation and differential privacy which can further enhance user confidentiality [36]. Thus, FL empowers organizations to leverage large-scale data insights while minimizing privacy risks, making it an attractive solution for sensitive domains like healthcare and finance.

FL involves transferring locally constructed data models from several devices to a central server to develop a global model. This is achieved using a client-server architecture, which ensures that the privacy of local data models is not compromised. FL thus offers several benefits, including reducing network overload, addressing latency issues, and ensuring the security and anonymity of sent data.

In this study, a client-server approach is adopted where the models are trained on local clients. FL uses iterative execution local training on local devices while the aggregation is done on the server side. The method described above was applied to develop a system for the prediction of heart disease carriers to demonstrate the value of FL in the healthcare sector. The development of a global model involved the usage of three local models, including RF, ET, and XGBoost, in which we used XGBoost as a global model. Through the process of averaging the results of multiple DTs, RF is an ensemble learning technique that reduces overfitting and increases accuracy to generate predictions. In contrast, the ET is similar to an RF but with more uncertainty, employing random feature thresholds for different types of trees, and less variance. The XGBoost gradient boosting algorithm combines DTs to iteratively restore errors, producing extremely accurate and effective

models. By applying these ML techniques, the proposed framework for heart disease carrier prediction may accurately predict an individual's carrier status.

In this proposed model, we apply local models from various clients and divide the dataset into client 1, client 2, and client 3. For client 1, we apply an RF model, an ET as client 2, and XGBoost as client 3. To ensure optimal performance Grid SearchCV is used to optimize hyperparameters on client datasets, and the best-tuned models are chosen. Hyperparameter grids for the RF and ET models are defined. Next, we used the RF and ET models to create a voting classifier ensemble. Moving forward on the preprocessed dataset, we trained an XGBoost classifier as a global model. The purpose of this global model is to catch patterns in all client datasets and offer a more comprehensive view. To facilitate collaboration, we create a global dataset by joining the datasets from many clients. After that, using the global dataset allows for training the global model since it takes into account data from every client. We make predictions utilizing the global model, valuing its capacity for reliable and accurate prediction-making.

4 Experiment and Results

Heart disease caused by climate change is analyzed and predicted using a range of ML models in this study. By putting people with similar symptoms together, analytics-based solutions can offer an accurate prognosis of disease.

4.1 Evaluation Matrix

The ROC curve, F1 score, Mathew's correlation coefficient (MCC), and accuracy score can be used to assess a model's performance on both positive and negative data. Accuracy represents the percentage of correctly classifier samples.

Accuracy is one of the criteria used to evaluate ML models. The percentage of correctly predicted events that our model predicts is its accuracy. The following formula can be used to determine accuracy in terms of positive and negative samples.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

Precision also called the positive predicted value (PPV), is defined as the proportion of true positive samples divided by the proportion of false positive samples.

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

Recall also called the true positive rate (TPR), is obtained by dividing the total number of correct predictions by the total number of original correct samples.

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

The harmonic mean of sensitivity and precision is called the F1 score.

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (11)$$

The importance of each quadrant of the confusion matrix is proportionate to the positive and negative components of the dataset, an increase in MCC necessitates success in all four quadrants.

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}} \quad (12)$$

This study aims to identify the ML model that leads to the highest performance in terms of multiple metrics, including area under the curve (AUC), F1 score, MCC, precision, recall, and accuracy.

4.2 Correlation Matrix

Fig. 2 shows how the different components are correlated. Each cell depicts the relationship between two variables. Correlation matrices are used to summarize data and as inputs for more complicated investigations, as well as diagnostics for such analysis. Eleven attributes of dataset range from x1 to x11. Each sample contains a label y1 with the digits 1 and 0, indicating the presence or absence of disease.

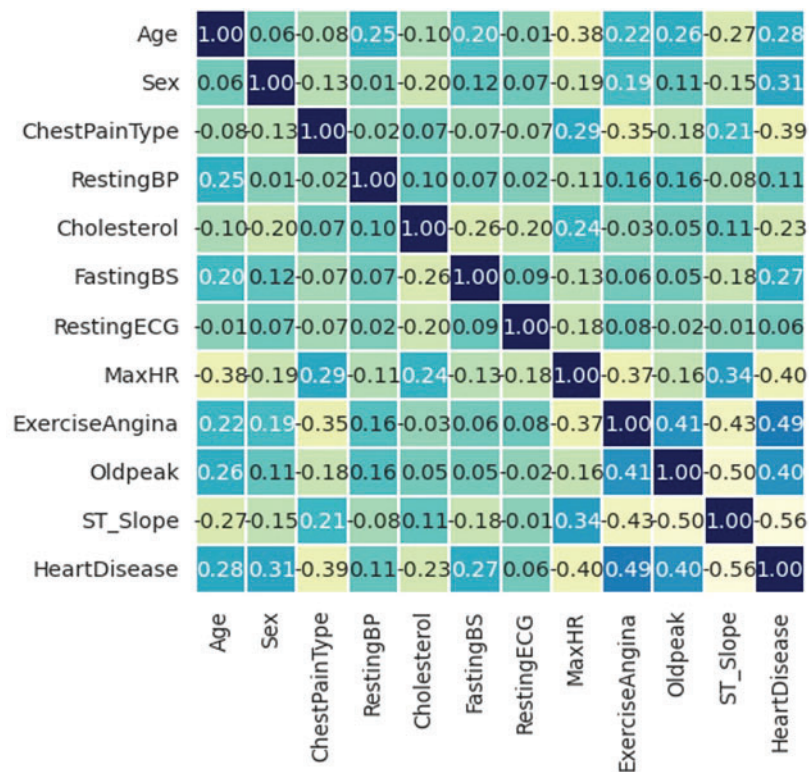


Figure 2: Correlation matrix of the dataset

4.3 Results

A precise range of -1 to $+1$ defines the correlation coefficient. As the name suggests, correlation analysis evaluates the relationship between two variables. There must be a link between the associated objects to do a correlation study. The correlation coefficient is a statistical tool used to quantify the strength of the relationship between two variables. The correlation coefficient is produced by multiplying the two variances by their respective averages using the product-difference approach, which is particularly helpful for calculating the linear single correlation coefficient.

Table 4 compares the performance of ML models. XGboost achieves the best accuracy of 90.20, while its precision, recall, and F1 scores are 0.89, 0.94, and 0.92, respectively. The RF has an 87.5 accuracy, 87.0 precision, and 87.6 recall, with an MCC of 74.6, and an 87.2 F1 score for classification. The accuracy, precision, recall, MCC, and F1 scores for the ET classifier are 88.5, 88.1, 88.9, 77.0, and 88.4, respectively. Scores for the SVC are 85.3, 84.8, 85, 69.9, and 84.9 for accuracy, precision, recall, MCC, and F1, respectively. While the accuracy, precision, recall, MCC, and F1 scores for the KNN classifier are 85.8, 85.4, 85.8, 71.2, and 85.5, respectively. LR has accuracy, precision, recall, MCC, and F1 scores of 84.7, 84.4, 85.2, 69.7, and 84.6, respectively. Results show that the XGBoost obtained the best performance with the highest accuracy of 90.20%. Class-wise results of the best performing XGBoost are given in **Table 5** indicating that the model obtained a similar accuracy for class 0 and class 1, which is 93.40% and 94.03%, respectively. Precision, recall, and F1 scores vary slightly for both classes, however, results regarding these metrics are good.

Table 4: Results for machine learning models

Classifier	Accuracy	Precision	Recall	F1 score
LR	84.7	84.4	85.2	84.6
ET	88.5	88.1	88.9	88.4
RF	87.5	87.0	87.6	87.2
KNN	85.8	85.4	85.8	85.5
SVC	85.3	84.8	85.0	84.9
XGBoost	90.20	0.89	0.94	0.92

Table 5: Results for federated learning

Class	Accuracy	Precision	Recall	F1 score
0	93.40	0.93	0.96	0.94
1	94.03	0.96	0.91	0.93

Results regarding the number of correct predictions (CP) and false predictions (FP) are displayed in **Table 6**. Results show that the LR has the lowest number of FP with only 16 false predictions, shared by the SVC with the same number of FP, while the highest number of false predictions are made by the KNN classifier. The XGBoost has a total of 20 false predictions. SVC and XGBoost also have the lowest number of false negatives with 6 and 7 false negatives, respectively.

Table 6: Results for federated learning

Model	TP	FP	TN	FN	CP	FP
LR	88	4	100	12	188	16
ET	77	11	87	29	164	40
RF	69	19	92	24	161	43
SVC	100	10	88	6	188	16
KNN	81	7	34	82	115	89
XGBoost	75	13	109	7	184	20

In this study, we used various ML algorithms to predict heart disease caused by climate change. A graphic depiction used to assess a binary classification model's effectiveness is carried out using the receiver operating characteristic (ROC) curve, as shown in Fig. 3. At various classification levels, it shows how the TPR and the false positive rate (FPR) trade-off. When evaluating each possible elimination for a test or set of tests, ROC curves are frequently used to visually represent the relationship between clinical sensitivity and specificity. The questioned test's use is indicated by the area under the ROC curve. The ROC is constantly in the range of 0 and 1.

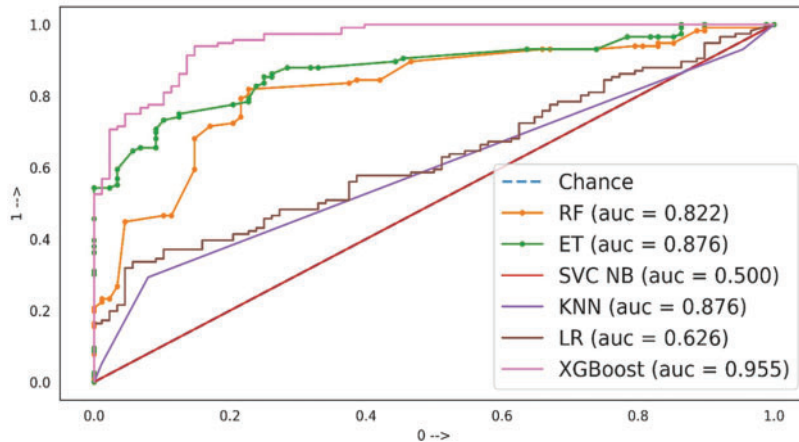


Figure 3: Models evaluation using AUC-ROC curve

An AUC range of 0.9 to 1.0 shows the best results indicating the model's capability to discriminate between positive and negative classes. In this case, the XGBoost classifier shows the highest AUC of 95.5%. No other model can get this AUC. On the other hand, an AUC between 0.8 to 0.9 is indicative of good performance but shows that the model can be further improved for a better AUC. The RF, ET classifier, and KNN have AUC in this range indicating that these models misclassified several samples from negative and positive classes. An AUC falling between 0.6 to 0.7 indicates a fair performance, as indicated by the performance of the LR model which has a 0.626 AUC score. SVC AUC shows poor performance which is between 0.5 to 0.6 and it can be said that it is similar to random guessing.

4.4 Computational Complexity

The XGB model uses DTs over multiple iterations and its computation complexity is approximately $O(n \log n)$ with n indicating the number of samples. The overall complexity of XGB for building multiple trees is $O(T \cdot n \log n)$ where T indicates the number of trees.

The computational complexity of FL is $O(C \cdot T \cdot n_c \log n_c)$ where C , n_c and T refer to the number of clients, number of samples per client, and number of trees per client. Combining XGB and FL complexity becomes $O(C \cdot T \cdot n_c \log n_c)$ which is the client-side training complexity. This complexity varies concerning the number of clients, as well as, the number of samples from each client. FL adds computational complexity to the models since it adds a distribution layer, so for larger deployments it can lead to higher computational costs.

4.5 Results for Statistical *t*-Test

To evaluate the statistical significance of the results with and without FL, we performed a *t*-test. It has a null hypothesis (H_0) and an alternate hypothesis (H_a), where the former states that there is no statistical significance between the results while the latter justifies the statistical significance between two results. Using the performance of the XGBoost classifier with and without FL, we performed a *t*-test. The results show a *p*-value of 0.038 which is less than 0.5 and rejects the null hypothesis. It means that the difference between the performance of the XGBoost using FL and without FL is statistically significant at 95% confidence level.

4.6 Discussion

Climate change has arisen as a global threat, affecting a variety of human health issues, including cardiovascular health. The prediction of heart disease caused by climate change has become a major field of study. ML approaches offer a robust set of tools for analyzing complex connections between climate variables and cardiovascular health outcomes. ML algorithms can detect hidden patterns and associations that influence the occurrence and severity of heart disease by using historical climatic data, health records, and other pertinent parameters. During the trials, many classification models were tried, each of which produced results that were not compatible with the expected class. FL, XGboost, RF, KNN, LR, ET and SVC are all examples of popular ML techniques. The FL approach was used, and it resulted in an accuracy of 94.03 percent.

In FL, an ML model is trained across distributed computers or servers, utilizing data from these clients. It may be used for issues of classification and regression. To improve the model's performance on challenging problems, ensemble learning is used. It works quickly and efficiently with massive datasets. The results of both extra trees (88.5% accuracy) and random forests (87.5% accuracy) demonstrate a model with extraordinarily excellent data fit. Contrary to the random forest, the KNN method is inefficient and prone to mistakes. Maintain an LR prediction accuracy of 84.7% or above for peace of mind. But the accuracy isn't quite as high as with FL. For comparing the relative values of two discrete classes, LR is the most reliable metric to use. In addition, we may draw from a database where the only possible replies are yes or no. Both the initial projections and the performance of the ongoing value have a high likelihood of materializing. The reasons for its popularity are discussed, even though its accuracy is lower than FL in this case. SVC achieves a high level of accuracy 85.3%, to be exact. SVC's mathematical skills are unrivaled. Nevertheless, its accuracy is inferior to that of the Random Forest classifier. Although FL has a 94.03% confidence interval, FL is the method of choice because it can be adjusted to improve data accessibility, speed data collection, and guarantee data security. Unfortunately, overtraining rarely works. FL provides the most precise forecasts by a large margin.

4.7 Performance Comparison

To analyze the performance of the proposed approach, we compared its performance with existing models of cardiovascular disease detection. Existing studies [16,17,20,25] have utilized a variety of ML models, both stand-alone and ensemble models, for disease detection. For example, study [19] utilized stacking to obtain an 89.86% accuracy in contrast to [16] which utilized stacking of KNN and LR to obtain an 87.24% accuracy. A random forest model is implemented in [18] to achieve a very good accuracy of 90.7% for cardiovascular disease detection. In comparison to these studies, the proposed FL approach provides a better accuracy of 94.03%, as shown in Table 7.

Table 7 indicates a better performance of the current approach involving FL. The better performance of the FL-based approach can be attributed to more than one factor. FL has more access to real-world data from diverse devices and learning from this diverse data helps the model get a better fit, thereby leading to better performance [11]. In the case of centralized ML models, overfitting can occur while FL offers the model to

learn from varied environments which increases their generalization and reduces overfitting. In addition, it also helps the FL model reduce dataset bias [37]. FL involves continuous learning from local data which improves model personalization.

Table 7: Comparison of the heart disease results

Reference	Year	Algorithms	Accuracy
[16]	2021	Stacking (KNN, LR)	87.24%
[17]	2022	Naive Bayes	84.0%
[18]	2022	Random forest	90.7%
[25]	2018	Logistic regression	85.0%
[19]	2022	Stacking	89.86%
[20]	2020	Random forest	86.89%
[21]	2020	KNN	87.0%
[23]	2024	Random forest	90.2%
[24]	2024	Ensemble (GRU+LeNet)	88.0%
Proposed	2025	FL+XGBoost	94.03

4.8 Practical Applications

The proposed approach can be used for the following real-world applications.

- The proposed FL-based approach can be used for early detection of health-related diseases and timely decisions can be made for better patient outcomes. In addition, personalized medication plans can be formulated.
- FL-based approaches allow collaborative training of models across various medical institutions with enhanced data privacy. It indicates that the proposed approach can be used by multiple hospitals to develop a joint predictive model involving data from wearable devices for heart disease detection. In addition, the predictive model can be fine-tuned to get data-specific results on local devices without affecting the general performance of the global model.
- FL uses real-time data from patients and in telemedicine, this data can help in early detection of various problems related to the heart. Continuous monitoring of user data from wearable gadgets is likely to enhance patient survival chances in the case of emergencies.

4.9 Limitations and Future Work

While the proposed approach offers better results, the following aspects are not explored in this study.

- The experiments are performed with a comparatively smaller dataset, which is why ML models were selected over DL models. Data augmentation is an attractive solution in this regard and further experiments can be performed involving larger datasets and data augmentation.
- One major challenge is the real-world deployment of FL-based approaches. In the real world, networks comprise heterogeneous devices, diverse communication patterns, as well as, various data requirements from clients which can affect the model's performance.
- Large-scale dataset adds further complexity as accessing them in real-time is challenging due to privacy constraints which might compromise the generalizability of the global model.

- As previously discussed, FL adds a distribution layer which increases the computational complexity of FL-based approaches. Computational complexity is also affected by the number of clients and the size of datasets. Deploying FL on edge devices with limited resources has practical limitations.

5 Conclusions

Climate change is rapidly being recognized as a potential danger to heart disease and cardiovascular health. Climate change-related rising global temperatures, more frequent and intense heatwaves, and poor air quality can all have a direct impact on cardiovascular risk factors and increase the incidence of heart-related incidents. The use of machine learning techniques in the prediction of heart disease caused by climate change has proven its potential to address significant public health challenges. We were able to acquire useful insights into the complicated relationships between climatic conditions and cardiovascular health outcomes by utilizing the power of machine learning algorithms. Given that the study's objective was to examine the efficacy of machine learning models in a federated learning approach to detect cardiovascular disease, the proposed FL-based model attained a good accuracy of 94.03%. The privacy protection offered by a federated learning technique is one of its key benefits. Patient information is maintained securely within the local data centers of each participating institution, so it is not necessary to divulge it to other companies or individuals which improves data security and confidentiality, two elements that are essential in the medical sector. The multifaceted method used in this study, which included varied climate measures, health data, and socioeconomic factors, gave a comprehensive knowledge of the influence of climate change on heart disease. The versatility of the predictive models ensures that they remain relevant and sensitive to the dynamic and evolving nature of climatic patterns and health trends. The findings of this study can be used to improve public health preparedness and conduct targeted interventions to counteract the negative effects of climate change on cardiovascular health. The incorporation of machine learning techniques creates a strong and forward-thinking framework for monitoring and managing climate-related health concerns.

Acknowledgement: The authors would like to express their gratitude to Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R104), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Funding Statement: This research was funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R104), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Author Contributions: Farhan Sarwar: conceptualization, data curation, writing—the original draft; Muhammad Shoaib Farooq: formal analysis, conceptualization, writing—the original draft; Nagwan Abdel Samee: methodology, funding acquisition, software; Mona M. Jamjoom: visualization, project administration, investigation; Imran Ashraf: supervision, validation, writing—review and editing. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The dataset used in this study is publicly available on the following link: <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction> (accessed on 10 June 2025).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Bui AL, Horwich TB, Fonarow GC. Epidemiology and risk profile of heart failure. *Nature Rev Cardiol.* 2011;8(1):30–41. doi:10.1038/nrcardio.2010.165.

2. de Bont J, Nori-Sarma A, Stafoggia M, Banerjee T, Ingole V, Jaganathan S, et al. Impact of heatwaves on all-cause mortality in India: a comprehensive multi-city study. *Environ Int.* 2024;184:108461. doi:10.1016/j.envint.2024.108461.
3. Brook RD, Rajagopalan S, Pope III CA, Brook JR, Bhatnagar A, Diez-Roux AV, et al. Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the American Heart Association. *Circulation.* 2010;121(21):2331–78. doi:10.1161/cir.0b013e3181dbecel.
4. Khraishah H, Alahmad B, Ostergard RL, AlAshqar A, Albaghdadi M, Vellanki N, et al. Climate change and cardiovascular disease: implications for global health. *Nature Rev Cardiol.* 2022;19(12):798–812. doi:10.1038/s41569-022-00720-x.
5. Ljungman PL, Mittleman MA. Ambient air pollution and stroke. *Stroke.* 2014;45(12):3734–41. doi:10.1161/strokeaha.114.003130.
6. Pourhomayoun M, Shakibi M. Predicting mortality risk in patients with COVID-19 using machine learning to help medical decision-making. *Smart Health.* 2021;20(2):100178. doi:10.1016/j.smhl.2020.100178.
7. Shaheen M, Farooq MS, Umer T, Kim BS. Applications of federated learning; taxonomy, challenges, and research trends. *Electronics.* 2022;11(4):670. doi:10.3390/electronics11040670.
8. Farooq MS, Tehseen R, Qureshi JN, Omer U, Yaqoob R, Tanweer HA, et al. FFM: flood forecasting model using federated learning. *IEEE Access.* 2023;11:24472–83. doi:10.1109/access.2023.3252896.
9. Li T, Sahu AK, Talwalkar A, Smith V. Federated learning: challenges, methods, and future directions. *IEEE Signal Process Magaz.* 2020;37(3):50–60. doi:10.1109/msp.2020.2975749.
10. McMahan B, Moore E, Ramage D, Hampson S, Arcas BA. Communication-efficient learning of deep networks from decentralized data. In: *Artificial intelligence and statistics*. Westminster, UK: PMLR; 2017. p. 1273–82.
11. Kairouz P, McMahan HB, Avent B, Bellet A, Bennis M, Bhagoji AN, et al. Advances and open problems in federated learning. *Found Trends® in Mach Learn.* 2021;14(1–2):1–210. doi:10.1561/22000000083.
12. Rana A, Chakraborty C, Sharma S, Dhawan S, Pani SK, Ashraf I. Internet of medical things-based secure and energy-efficient framework for health care. *Big Data.* 2022;10(1):18–33. doi:10.1089/big.2021.0202.
13. Zhao Y, Li M, Lai L, Suda N, Civin D, Chandra V. Federated learning with non-iid data. *arXiv:1806.00582.* 2018.
14. Bonawitz K, Eichner H, Grieskamp W, Huba D, Ingerman A, Ivanov V, et al. Towards federated learning at scale: system design. In: *Proceedings of the Second Conference on Machine Learning and Systems, SysML 2019; 2019 Mar 31–Apr 2*. Stanford, CA, USA; 2019. p. 374–88.
15. Chen F, Luo M, Dong Z, Li Z, He X. Federated meta-learning with fast convergence and efficient communication. *arXiv:1802.07876.* 2018.
16. Araujo M, Pope L, Still S, Yannone C. GR-130-prediction of heart disease with machine learning techniques. In: *Digital commons*. Kennesaw, GA, USA: Kennesaw State University; 2021.
17. Ansarullah SI, Saif SM, Kumar P, Kirmani MM. Significance of visible non-invasive risk attributes for the initial prediction of heart disease using different machine learning techniques. *Comput Intell Neurosci.* 2022;2022(7):9580896. doi:10.1155/2022/9580896.
18. Siregar MU, Setiawan I, Akmal NZ, Wardani D, Yunitasari Y, Wijayanto A. Optimized random forest classifier based on genetic algorithm for heart failure prediction. In: *2022 Seventh International Conference on Informatics and Computing (ICIC); 2022 Dec 8–9; Denpasar, Indonesia.* p. 1–6.
19. Liu J, Dong X, Zhao H, Tian Y. Predictive classifier for cardiovascular disease based on stacking model fusion. *Processes.* 2022;10(4):749. doi:10.3390/pr10040749.
20. Anbuselvan P. Heart disease prediction using machine learning techniques. *Int J Eng Res Technol.* 2020;9:515–8.
21. Singh A, Kumar R. Heart disease prediction using machine learning algorithms. In: *2020 International Conference on Electrical and Electronics Engineering (ICE3); 2020 Feb 14–15; Gorakhpur, India.* p. 452–7.
22. Tambe PM, Shrivastava M. Hybrid brave-hunting optimisation for heart disease detection model with SVM coupled deep CNN. *Int J Intel Informat Database Syst.* 2025;17(1):92–123. doi:10.1504/ijiids.2025.10068157.
23. Naeem AB, Senapati B, Bhuva D, Zaidi A, Bhuva A, Sudman MSI, et al. Heart disease detection using feature extraction and artificial neural networks: a sensor-based approach. *IEEE Access.* 2024;12(2):37349–62. doi:10.1109/access.2024.3373646.

24. Khan H, Javaid N, Bashir T, Akbar M, Alrajeh N, Aslam S. Heart disease prediction using novel ensemble and blending based cardiovascular disease detection networks: ensCVDD-Net and BICVDD-Net. *IEEE Access*. 2024;12:109230–254. doi:10.1109/access.2024.3421241.
25. Dwivedi AK. Performance evaluation of different machine learning techniques for prediction of heart disease. *Neural Comput Applicat*. 2018;29(10):685–93. doi:10.1007/s00521-016-2604-1.
26. Fedesorian. Heart failure prediction dataset; 2022 [Internet]. [cited 2024 Feb 23]. Available from: <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>.
27. Hastie T, Tibshirani R, Friedman JH, Friedman JH. The elements of statistical learning: data mining, inference, and prediction. Vol. 2. Cham, Switzerland: Springer; 2009.
28. Yang J, Guan J. A heart disease prediction model based on feature optimization and smote-XGBoost algorithm. *Information*. 2022;13(10):475. doi:10.3390/info13100475.
29. Sharaff A, Gupta H. Extra-tree classifier with metaheuristics approach for email classification. In: *Advances in computer communication and computational sciences: Proceedings of IC4S 2018*. Cham, Switzerland: Springer; 2019. p. 189–97.
30. Jain A. Machine learning techniques for medical diagnosis: a review. In: *2nd International Conference on Science, Technology and Management*; 2015 Sep 27; New Delhi, India. p. 2449–59.
31. Alotaibi SS, Almajid YA, Alsaahali SF, Asalam N, Alotaibi MD, Ullah I, et al. Automated prediction of Coronary Artery Disease using Random Forest and Naïve Bayes. In: *2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*; 2020 Oct 17–18; Depok, Indonesia. p. 109–14.
32. Srivastava A. Impact of k-nearest neighbour on classification accuracy in KNN algorithm using machine learning. In: *Advances in smart communication and imaging systems: Select proceedings of MedCom 2020*. Cham, Switzerland: Springer; 2021. p. 363–73.
33. Basha YH, Nassif AB, AlShabi M. Predicting heart failure disease using machine learning. In: *Smart biomedical and physiological sensor technology XIX*. Vol. 12123. Bellingham, DC, USA: SPIE; 2022. p. 75–84.
34. Mehmood A, Farooq MS, Naseem A, Rustam F, Villar MG, Rodríguez CL, et al. Threatening URDU language detection from tweets using machine learning. *Appl Sci*. 2022;12(20):10342. doi:10.3390/app122010342.
35. Tehseen R, Farooq MS, Abid A. A framework for the prediction of earthquake using federated learning. *PeerJ Comput Sci*. 2021;7(7):e540. doi:10.7717/peerj-cs.540.
36. Bonawitz K, Ivanov V, Kreuter B, Marcedone A, McMahan HB, Patel S, et al. Practical secure aggregation for privacy-preserving machine learning. In: *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*; 2017 Oct 30–Nov 3. Dallas, TX, USA. p. 1175–91. doi:10.1145/3133956.3133982.
37. Smith V, Chiang CK, Sanjabi M, Talwalkar AS. Federated multi-task learning. In: *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, USA. 2017 [cited 2025 Jun 10]. Available from: https://proceedings.neurips.cc/paper_files/paper/2017/file/6211080fa89981f66b1a0c9d55c61d0f-Paper.pdf.