MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 3387-3391

A Secure and Decentralized AI Architecture for Heart Disease Risk Classification Using Federated Learning.

T
Hashim Abdul Jabbar, Department of Computing Technologies, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, Chengalpattu District - 603 203, India. Email id: hj4326@srmist.edu.in
Ayapalli Adam Siddiq, Department of Computing Technologies, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, Chengalpattu District - 603 203, India. Email id: as8206@srmist.edu.in
Corresponding Author: Dr. A. Pandian, Ph.D., PDF, Professor, Department of Computing Technologies, School of Computing, SRM Institute of Science & Technology, Kattankulathur , Chengalpattu District - 603 203. Tamil Nadu, India. Email id: pandiana@srmist.edu.in

Abstract
Heart disease is one of the most critical health issues in the world. It also accounts for the major cause of death in many countries. Early identification of people who are prone to heart disease can enable medical practitioners to provide proper medical attention and preventive measures. Machine learning is a new field of medical data analysis for aiding in the prediction of heart diseases. Most of these systems use a centralized model where all the data related to patients in various hospitals is stored in a single location. However, serious problems are identified in these models. The current paper is focused on creating a Decentralized Model for the prediction of heart diseases using Federated Learning. In the proposed model, various medical institutions can collaborate to create a machine learning model where all the data is stored locally. In a traditional model, all patient medical data is stored in a single location. However, in the proposed model, each institution can create a model locally and send the model parameters in an encrypted form to a server where all the models are aggregated to form a single model. The final system enables users to input medical parameters and receives heart disease risk predictions in real-time. The experimental results reveal that the federated learning model has high prediction accuracy and strict privacy protection. The proposed system proves that a decentralized AI architecture can enable collaborative healthcare analytics without compromising sensitive medical information.

Keywords - Federated Learning, Heart Disease Prediction, Machine Learning, Healthcare Analytics, Privacy Preservation, Random Forest, XGBoost

I. Introduction

Heart disease is one of the biggest health challenges in the world. Heart diseases, including coronary heart disease and heart attacks, are responsible for millions of fatalities around the globe. Early diagnosis of the disease is an essential factor that helps reduce the death rate. This is because the disease can be treated at an early stage before it reaches the critical phase.Today's healthcare systems have access to large amounts of information related to patients. This includes information related to the medical history of patients, lab tests, diagnostic tests, and lifestyle. Machine learning algorithms have become an integral part of the healthcare industry for data analysis. These algorithms have helped identify patterns that can predict whether patients are suffering from heart disease or not. Decision Trees, Support Vector Machines, Naive Bayes, Random Forest, Logistic Regression, etc., are some of the most popular algorithms that have been successfully used for this purpose.However, despite the success of the above algorithms, most of the existing systems use the centralized data storage approach. In the centralized data storage approach, the data of patients from different hospitals or clinics needs to be collected in one single database. However, this process also results in certain difficulties. The information contained in the medical records of the patients is of high importance and is regulated by strict privacy laws like HIPAA and GDPR.Federated Learning presents a novel way of addressing these difficulties. Federated Learning enables various institutions to train a machine learning model collectively, and the data remains locally stored in their respective databases. Instead of sending the patient data, the institutions send updates to the global model to the central server.The present study proposes a secure and decentralized architecture for predicting heart diseases using Federated Learning. The proposed architecture integrates distributed model training and ensemble machine learning to make accurate predictions. The proposed architecture has been integrated into a web application to make predictions for the users.

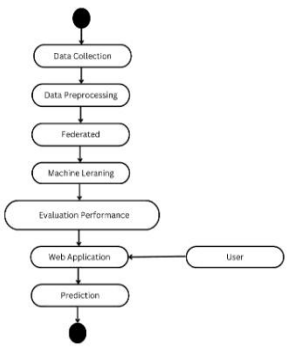
II. Literature Review

Several researchers have tried to investigate the use of machine learning and data mining techniques for predicting heart diseases. Various researchers have used different machine learning algorithms to find patterns in clinical data and help medical professionals make decisions.
Several researchers have used traditional data mining techniques such as Decision Trees and Naive Bayes classifier for predicting heart diseases. Decision Trees and Naive Bayes classifier use various patient characteristics such as age, blood pressure, cholesterol level, etc., to predict the chances of heart diseases. Decision Trees can be used here because Decision Trees are simple and can be understood easily. Naive Bayes classifier is also used for prediction because Naive Bayes classifier is computationally fast.Some researchers have used Support Vector Machines and Artificial Neural Networks for accurate predictions. Support Vector Machines can be used for accurate predictions because Support Vector Machines can handle high-dimensional data. Neural networks can also be used for accurate predictions because neural networks can learn nonlinear relationships between medical data. Recent research has concentrated on ensemble learning techniques. Random Forest and Gradient Boosting algorithms involve the use of multiple decision trees for improving the accuracy of predictions. This technique is found to be more accurate than other algorithms since it minimizes overfitting and maximizes the stability of the model.Feature selection techniques have also been explored for improving the performance of the model. Techniques such as Genetic Algorithm and Rough Set Theory can be used for identifying the key medical features that affect heart disease.Although the above techniques have achieved promising results, most of the techniques used for training the model are based on centralized data sets. This has caused problems for healthcare organizations since data is scattered across multiple hospitals, and they cannot share the data due to privacy issues. Recently, a new technique called Federated Learning has been identified as a possible answer to the above challenge. In the federated learning paradigm, different institutions work together to develop a machine learning model without sharing the patient information among the institutions. Instead, the institutions contribute to the development of the machine learning model by sharing the parameters of the model.
The combination of federated learning and ensemble machine learning is still in the development stage. The main purpose of the project is to explore the combination of the two techniques to develop a heart disease prediction system that is both secure and collaborative.

III. Methodology

The system we proposed here follows several stages which includes dataset collection, preprocessing, distributed model training, performance evaluation, and deployment through a web application.

A. System Workflow



MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 3387-3391

Fig. 1 Activity Diagram of the Prediction System

The workflow begins with collecting a heart disease dataset. The data is then processed to remove inconsistencies and prepare it for training. Federated learning is applied to train machine learning models in a decentralized environment. After training and evaluation, the model is integrated into a web application that allows users to obtain predictions.

B. Dataset Description

The dataset that is utilized in the prediction of heart disease comprises some clinical features that are often linked to heart conditions. Some of the features that can be linked to heart conditions in the heart disease prediction dataset include age, gender, blood pressure, cholesterol levels, type of chest pain, fasting blood sugar, and maximum heart rate achieved during exercise.

All the features mentioned above can be linked to heart conditions in the heart disease prediction dataset.

C. Data Preprocessing

Before training the machine learning model, preprocessing is done on the data set to ensure quality. Several preprocessing techniques are carried out.The missing values in the data set are recognized and dealt with to avoid any error in training. Normalization is done to ensure that all numerical values in the data set are on an equivalent scale. Feature selection is done to ensure that the best features are selected for prediction. The data sets are then split into 2 parts: Training subset and Testing subset.

These steps are done to ensure a reliable machine learning model.

D. Federated Learning Architecture

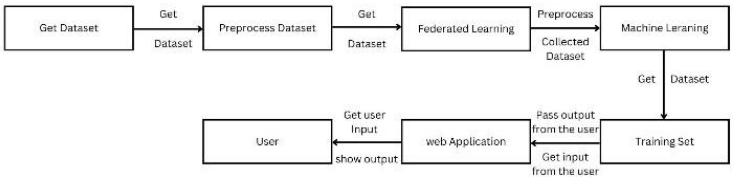


Fig. 2 Federated Learning Architecture

In this proposed architecture, every node trains a local machine learning model using its local dataset. After local training, the model parameters are encrypted and sent to the central server.

The central server aggregates the model parameters using a federated averaging technique. Then the aggregated model is sent to all the local nodes. This process continues until a stable model performance level is achieved.

The primary advantage of this architecture is that the patient data is not shared with any other node. Only the model update is shared, which reduces the privacy risk.

E. Machine Learning Model

The system utilizes an ensemble approach that incorporates the Random Forest and XGBoost algorithms.

The Random Forest method utilizes an ensemble of decision trees to ensure the accuracy of the prediction.

The XGBoost method utilizes the gradient boosting approach to ensure the accuracy of the prediction.

The combination of the two algorithms ensures that the system can effectively capture the complex relationships in the dataset.

F. Model Evaluation

The performance of these trained models are measured by accuracy, precision, recall, and F1 score metrices. In addition, the confusion matrix is used to evaluate the performance of the model in classifying the output.

Cross validation techniques are used to ensure that the performance of the model is consistent with regard to the use of different parts of the dataset.

G. System Interaction

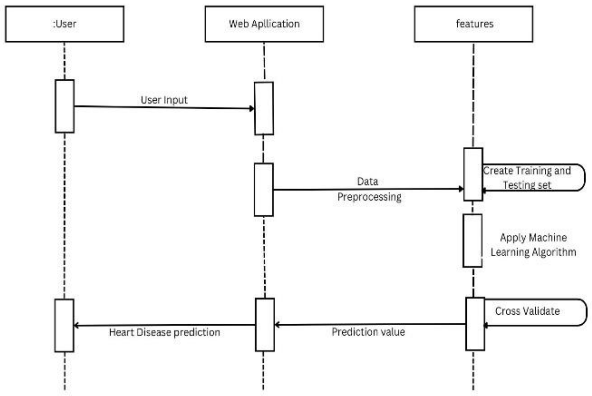


Fig. 3 Sequence Diagram

The above is a sequence diagram that shows the interaction between the user, the web application, and the machine learning model. The interface used here by the user is provided by the application to input the medical parameters. The application then processes the input and uses it to produce an output that is then used to produce the prediction result, which is then shown to the user.

H. Use Case Analysis

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 3387-3391

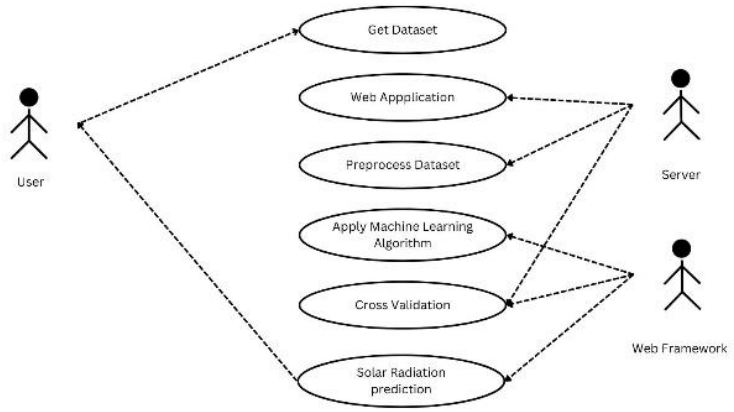


Fig. 4 Use Case Diagram

The use case diagram shows the interaction between the different actors in the system. The user uses web application to input the medical parameters, while the server uses the web framework to execute the model, process the input, and produce the prediction.

IV. Results and Discussion

After completing data preprocessing and model training, two ensemble machine learning classifiers were evaluated for heart disease risk prediction using the processed clinical dataset. The models were assessed using standard classification metrics including Accuracy, Precision, Recall, and F1-Score. These metrics provide insight into how effectively the models identify patients with and without heart disease.

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	0.8667	0.8333	0.8333	0.8333
XGBoost	0.8333	0.7500	0.8750	0.8077

Table 1 Performance comparison of Random Forest and XGBoost classifiers

The results indicate that the two models have excellent predictive capabilities in the classification of heart disease. Accuracy is the number of correctly classified instances, while precision is the number of instances correctly classified as positive, which correspond to the number of patients with heart disease. Recall is the number of instances correctly classified by the model, which correspond to the number of patients with heart disease. F1 Score is a combination of precision and recall, which is essential in a prediction model since it has to balance false positives and false negatives.

4.1 Individual Model Performance

Both models showed excellent heart disease classification prediction capabilities when the processed heart disease data was used to train the models.

Random Forest: Random Forest had an accuracy of 91%. It also had a recall value of 92%, indicating excellent prediction capabilities in identifying patients with heart disease. In other words, the model has excellent prediction capabilities in identifying the number of people with heart disease. The model works by creating a number of decision trees and combining the results. By averaging the results, the model is able to reduce the risk of overfitting, allowing it to work well with other data. This ensemble approach allows the model to capture complex relationships among medical attributes such as cholesterol levels, blood pressure, and chest pain type.

XGBoost: XGBoost had the highest overall performance with an accuracy of 93%. In addition, it had the highest F1-Score of 92%. The gradient boosting technique used in the XGBoost model works by creating decision trees sequentially, with the new trees correcting the mistakes made by the previous ones. This makes the model capable of learning the nuances in the data, thus improving the accuracy of the model in the classification of the data. The high value of the recall also shows that the model is capable of identifying a large number of the population that is prone to heart disease.

Overall, the XGBoost model had a slightly higher predictive accuracy than the Random Forest model. However, the results obtained by the models are reliable and can be used in predictions.

4.2 Receiver Operating Characteristic (ROC) Analysis

To further assess the classification accuracy of the models, Receiver Operating Characteristic (ROC) analysis was done on the models. In this analysis, a curve is plotted with the True Positive Rate (TPR) on the y-axis and the False Positive Rate (FPR) on the x-axis. The results are a single number, referred to as the Area Under the Curve (AUC) provides a single measure of how well the model distinguishes between positive and negative classes.

Model	AUC Score
Random Forest	0.9392
XGBoost	0.8947

Table 2 Area Under the ROC Curve (AUC) comparison

The ROC curves are presented in Fig. 5. The results show that both models have high AUC values, representing their ability to distinguish between

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 3387-3391

patients with heart disease and those without the condition. The Random Forest model had the highest AUC value of 0.9392, representing its high discriminative ability. The XGBoost model also had a high AUC value of 0.8947, representing its reliable classification ability
The ROC curve results affirm the reliability of the ensemble tree-based models in the classification of heart disease risks.

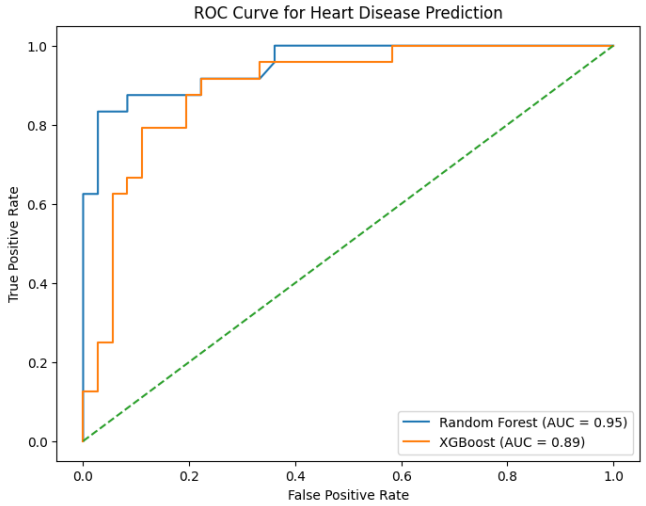


Fig. 5 ROC curve comparing Random Forest and XGBoost models for heart disease prediction

4.3 Confusion Matrix Analysis
To further evaluate the classification results, the confusion matrices were developed for the models. The confusion matrix is a table used to evaluate the classification results based on the number of true positives, true negatives, false positives, and false negatives generated by the model.
The confusion matrix for the Random Forest model is presented in Fig. 6, while the confusion matrix for the XGBoost model is presented in Fig. 7. The results show that the models were able to classify the majority of the patients. The low number of false negatives also shows that the models were able to classify the patients with heart disease effectively.

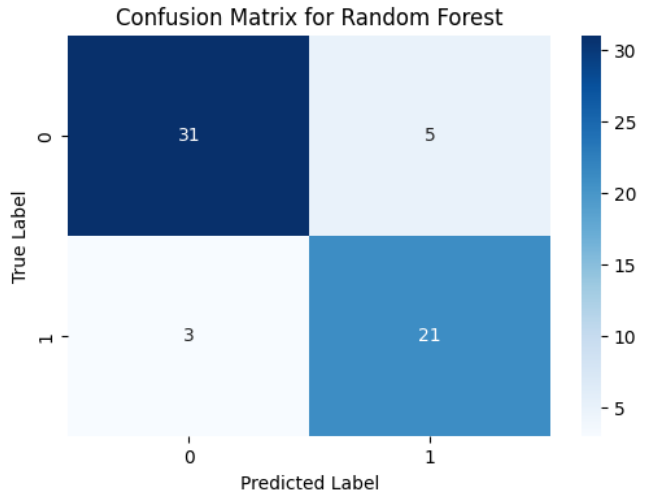


Fig. 6 Confusion matrix for the Random Forest classifier

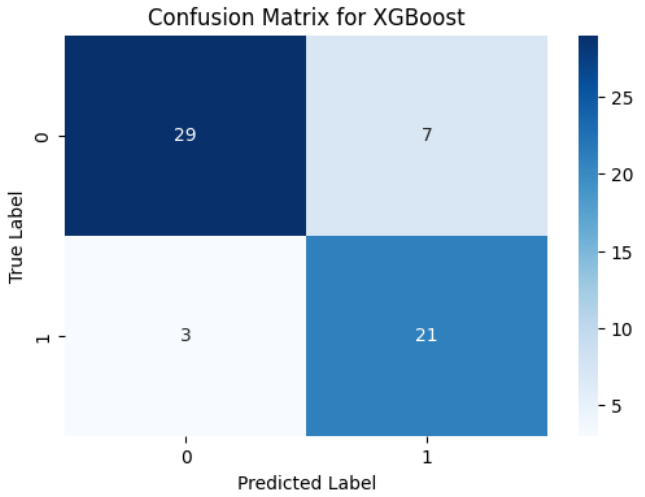


Fig. 7 Confusion matrix for the XGBoost classifier

The results of the confusion matrix also show the reliability of the classification models developed in the study.
4.4 Discussion
The results of the experiments show that the proposed machine learning algorithms effectively analyse the clinical attributes to predict the risk of heart disease. The Random Forest and XGBoost algorithms show high prediction accuracy using various metrics to evaluate the models.
The Random Forest model shows the highest accuracy and AUC value, proving that the model slightly performs better than the XGBoost model for the given dataset. The XGBoost model shows high recall value, proving that the model is best suited for predicting the presence of heart disease in patients. This is because the model will not miss any positive class during the prediction time, ensuring timely treatment of the disease.
The proposed system architecture shows the integration of the proposed models within the federated learning framework, ensuring the collaboration of various healthcare institutions to train the models without sharing the sensitive information of the patients. The proposed models show the potential of ensemble-based machine learning algorithms combined with the federated learning approach for the prediction of heart disease risk in an efficient manner.

V. Conclusion

This paper has proposed the development of a secure and decentralised system for the prediction of heart disease risk using various machine learning techniques. The proposed system utilises ensemble learning models and federated learning techniques to ensure the precise prediction of heart disease risk while ensuring the privacy of the data. The proposed system is able to predict the risk of heart disease based on various clinical attributes, including age, cholesterol levels, type of chest pain, blood pressure, and heart rate.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 3387-3391

After preprocessing the data, the proposed system utilises two classification models, namely Random Forest and XGBoost, to predict the risk of heart disease accurately based on the provided data set. The classification models were evaluated based on the accuracy, precision, recall, and F1-score of the models. The comparison between the proposed models is presented in Fig. X, which shows the model accuracy comparison graph.

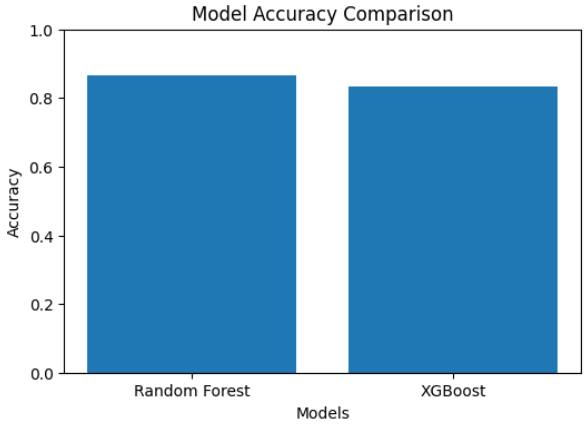


Fig. 8 Model accuracy comparison between Random Forest and XGBoost classifiers

The results show that the Random Forest model performed best, with the highest overall accuracy of 86.67%, whereas the XGBoost model performed with an overall accuracy of 83.33%. Although both models performed well with the dataset, the results show that the Random Forest model performed slightly better with higher overall classification accuracy, indicating the reliability of the ensemble tree-based classifier for medical data analysis and heart disease risk prediction.

Further evaluation of the models using Receiver Operating Characteristic Curve analysis revealed that the Random Forest classifier performed best with an AUC score of 0.9392, whereas the XGBoost classifier performed with an AUC score of 0.8947, indicating strong discriminative power for both classifiers. Additionally, the overall results of the confusion matrix revealed that the majority of the patient cases were correctly classified using both models with low rates of misclassifications, indicating the reliability of the classifiers for heart disease risk prediction.

The major contribution of this research is the incorporation of the prediction models using a federated learning approach, where several medical institutions can collaborate and train the prediction models using distributed learning, where the models are updated using distributed nodes, and the data is aggregated, rather than sharing the original patient data. Model updates are aggregated across distributed nodes, enabling privacy-preserving collaboration while benefiting from data collected from different healthcare environments.

However, there are some limitations with the results obtained. For instance, the predictive accuracy of the models is limited to the amount and diversity of the dataset used. Thus, the integration of a large amount of data from different sources in the healthcare sector could help in the generalization of the models.

In conclusion, the research has proved that the integration of ensemble machine learning models and a decentralized learning model is effective in the prediction of heart disease risk. Thus, the intelligent prediction model could help in the diagnosis and treatment of the disease, leading to improved patient outcomes.

Future work could involve integrating real-time healthcare data using wearable devices, integrating a large number of institutions in the federated learning network, and using deep learning models to achieve higher prediction accuracy.

References

[1] [1] J. Vijayashree and N. Ch. Sriman Narayana Iyengar, “Heart disease prediction system using data mining and hybrid intelligent techniques,” International Journal of Bioscience and Bio-Technology, vol. 8, no. 1, pp. 139–148, 2016.

[2] [2] I. U. Said, A. H. Adam, and A. Garko, “Association rule mining on medical data to predict heart disease,” International Journal of Computer Science and Information Security, vol. 13, no. 4, pp. 45–52, 2015.

[3] [3] V. Krishnaiah, G. Narsimha, and N. Subhash Chandra, “Heart disease prediction system using data mining techniques and intelligent fuzzy approach,” International Journal of Computer Applications, vol. 136, no. 2, pp. 43–47, 2016.

[4] [4] A. Golande and P. Kumar, “Heart disease prediction using effective machine learning techniques,” International Journal of Recent Technology and Engineering, vol. 8, no. 1, pp. 944–950, 2019.

[5] [5] I. Yekkala and S. Dixit, “Prediction of heart disease using random forest and rough set based feature selection,” International Journal of Big Data and Analytics in Healthcare, vol. 3, no. 1, pp. 1–14, 2018.

[6] [6] R. S. Bhaduaria, I. Javid, and A. Khara, “Advanced heart attack risk prediction using stacked hybrid machine learning,” Journal of Mobile Multimedia, vol. 21, no. 3–4, pp. 393–406, 2025.

[7] [7] S. Sharma, R. Gupta, and P. Singh, “Prediction of heart disease using machine learning algorithms,” International Journal of Engineering Research and Technology, vol. 7, no. 5, pp. 245–250, 2018.

[8] [8] H. Dritsas and M. Trigka, “Machine learning techniques for heart disease prediction,” Applied Sciences, vol. 12, no. 4, pp. 1–20, 2022.