Heart Disease Prediction Using Machine Learning and Federated Learning Approach

Nabiha Fatma[1], Dr Mohammad Suaib[2], Dr Jameel Ahmad[3]

Abstract: ***Heart disease remains one of the leading causes of mortality worldwide, making early prediction and diagnosis crucial for improving patient outcomes. This research explores an integrated approach combining Machine Learning (ML) techniques with Federated Learning (FL) to develop a robust and privacy-preserving heart disease prediction system. Traditional ML models rely on centralized data collection, which often raises concerns related to data privacy, security, and regulatory compliance. To address these challenges, this study proposes a federated learning framework that enables multiple healthcare institutions to collaboratively train predictive models without sharing sensitive patient data. The proposed system utilizes clinical and demographic features such as age, blood pressure, cholesterol levels, and electrocardiographic results to train various machine learning algorithms, including decision trees, support vector machines, and ensemble models. These models are evaluated based on accuracy, precision, recall, and F1-score. The federated learning approach ensures that data remains localized while model updates are securely aggregated, thereby maintaining data confidentiality. Experimental results indicate that the hybrid ML-FL model achieves competitive prediction performance while significantly enhancing data privacy. This approach not only improves the scalability of predictive systems across distributed environments but also aligns with modern data protection requirements. The study demonstrates that integrating federated learning with traditional machine learning techniques can be an effective solution for building secure, efficient, and reliable healthcare prediction systems. Overall, this research contributes to the advancement of intelligent healthcare systems by offering a privacy-aware predictive framework for early heart disease detection.***

Keywords: Heart Disease Prediction, Machine Learning, Federated Learning, Healthcare Analytics, Data Privacy, Distributed Learning, Predictive Modeling, Medical Diagnosis, Classification Algorithms, Secure Data Sharing

1. Introduction

1) Background of Heart Disease

Heart disease is one of the most critical health challenges faced globally, contributing to a high rate of morbidity and mortality. Cardiovascular diseases (CVDs), including coronary artery disease, heart failure, and arrhythmias, are responsible for millions of deaths each year. The growing prevalence of unhealthy lifestyles, such as poor diet, lack of physical activity, smoking, and stress, has further increased the risk of heart-related conditions. Early detection and prevention are essential to reduce the burden on healthcare systems and improve patient survival rates.

Traditionally, the diagnosis of heart disease relies on clinical tests, medical imaging, and physician expertise. While these methods are effective, they can sometimes be time-consuming, expensive, and dependent on subjective interpretation. With the rapid growth of healthcare data, there is a need for intelligent systems that can assist medical professionals in making accurate and timely decisions.

2) Role of Machine Learning in Healthcare

Machine Learning (ML) has gained significant attention in recent years due to its ability to analyze large volumes of data and extract meaningful patterns. In the healthcare domain, ML techniques are widely used for disease prediction, diagnosis, and treatment planning. These algorithms can process various types of medical data, including patient history, laboratory results, and imaging data, to predict the likelihood of diseases such as heart disease.

Commonly used ML algorithms for heart disease prediction include decision trees, logistic regression, support vector machines (SVM), and neural networks. These models are trained using historical patient data and can identify complex relationships between risk factors and disease outcomes. As a result, ML-based systems can provide high accuracy and assist healthcare professionals in early diagnosis.

Despite these advantages, traditional ML models often require centralized data storage, where data from multiple sources is collected and processed in a single location. This centralized approach poses significant challenges, particularly in the healthcare sector, where data privacy and security are of utmost importance.

3) Challenges in Centralized Healthcare Data Systems

The use of centralized data systems in healthcare raises several concerns related to privacy, security, and data ownership. Patient data is highly sensitive and must be protected under strict regulations and ethical guidelines. Sharing medical data across hospitals or organizations can lead to risks such as data breaches, unauthorized access, and misuse of information.

In addition, different healthcare institutions may follow different data formats and standards, making data integration a complex process. Legal restrictions and compliance requirements further limit the ability to share data across regions or countries. These challenges hinder the development of large-scale machine learning models that require diverse and comprehensive datasets to achieve better performance.

Another limitation of centralized systems is scalability. As the volume of healthcare data continues to grow, storing and processing all data in a single location becomes inefficient and costly. Therefore, there is a need for alternative approaches that can address these limitations while still enabling collaborative model development.

4) Federated Learning: A Privacy-Preserving Approach

Federated Learning (FL) has emerged as an innovative solution to overcome the challenges associated with centralized machine learning. It is a decentralized learning approach in which multiple participants, such as hospitals or healthcare institutions, collaboratively train a global model without sharing their raw data. Instead of transferring data to a central server, each participant trains the model locally using its own dataset and shares only the model updates.

These updates are then aggregated to improve the global model, ensuring that sensitive patient data remains within the local environment. This approach significantly enhances data privacy and security while enabling the use of distributed datasets for model training. Federated learning also reduces the risk of data breaches and ensures compliance with data protection regulations.

Moreover, FL supports scalability and flexibility, as new participants can join the training process without affecting the existing system. It allows the integration of diverse datasets from different sources, leading to more robust and generalized models.

5) Integration of Machine Learning and Federated Learning

The integration of Machine Learning with Federated Learning offers a powerful framework for developing advanced healthcare prediction systems. While ML provides the capability to analyze complex data and generate accurate predictions, FL ensures that the training process is conducted in a secure and privacy-preserving manner.

In the context of heart disease prediction, this hybrid approach enables multiple healthcare providers to collaboratively build a predictive model using their local data. This not only improves the model's accuracy by incorporating diverse datasets but also maintains patient confidentiality. Various ML algorithms can be implemented within the federated learning framework to evaluate their performance and identify the most effective approach.

The combination of ML and FL addresses both technical and ethical challenges, making it a suitable solution for modern healthcare applications. It bridges the gap between data-driven innovation and data privacy requirements.

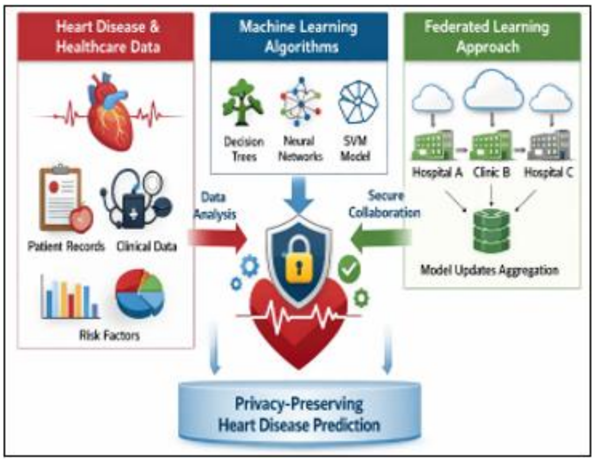

Figure 1: Privacy-Preserving Heart Disease Prediction

2. Literature Review

2.1 Review of Heart Disease Prediction Models

Recent studies have extensively explored the use of computational models for predicting heart disease using clinical datasets. A study published in Nature Scientific Reports (2024) proposed a machine learning-based prediction system integrated with explainable AI techniques, highlighting the importance of interpretability in medical decision-making. The study demonstrated that accurate prediction models can significantly assist in early diagnosis and reduce mortality rates.

Similarly, a comprehensive review published in Frontiers in Artificial Intelligence (2025) analyzed various machine learning and deep learning approaches used in heart disease prediction. The authors emphasized the transition from traditional statistical models to hybrid and deep learning-based frameworks, which provide better predictive performance due to their ability to capture complex patterns in medical data.

Another recent framework (2025) evaluated multiple classifiers such as Random Forest, Logistic Regression, and KNN, achieving high accuracy (around 90%) and demonstrating the effectiveness of ensemble learning methods in cardiovascular risk prediction.

2.2 Machine Learning Techniques in Healthcare

Machine learning has become a core component in modern healthcare systems due to its capability to process large-scale medical datasets. Techniques such as Support Vector Machines (SVM), Decision Trees, Neural Networks, and Deep Learning architectures are widely used for disease prediction.

According to recent literature, ML models are capable of analyzing heterogeneous data such as electronic health records (EHRs), imaging data, and wearable sensor data to provide accurate predictions. These models not only enhance diagnostic accuracy but also support personalized treatment planning.

In addition, ML techniques are increasingly being integrated with optimization and feature selection methods to improve model performance. However, these models still rely heavily on centralized data, which creates challenges in terms of scalability and privacy protection.

2.3 Applications of Federated Learning

Federated Learning (FL) has emerged as a promising solution to address data privacy concerns in healthcare systems. Unlike traditional approaches, FL allows multiple institutions to collaboratively train models without sharing raw data.

A recent study (2024) proposed a federated learning-based framework combined with blockchain for secure heart disease prediction. The model utilized advanced architectures such as TabNet and demonstrated improved performance while preserving data privacy.

Another study (2025) explored deep federated learning techniques for heart disease detection, including Federated Averaging (FedAvg) and Personalized Federated Learning (PFL). The results showed that PFL achieved high accuracy (above 96%), indicating the potential of federated systems in real-world healthcare applications.

Furthermore, recent research (2026) introduced an IoT-enabled federated learning framework for cardiovascular risk prediction. This approach integrates wearable devices and distributed learning to enhance real-time monitoring and predictive analysis while maintaining data privacy.

2.4 Limitations of Existing Approaches

Despite significant advancements, existing heart disease prediction systems face several limitations. Traditional machine learning models require centralized data storage, which raises concerns regarding data privacy, security, and regulatory compliance.

Even though federated learning addresses privacy issues, it introduces challenges such as communication overhead, model convergence issues, and heterogeneity of data across institutions. According to recent studies, differences in data distribution (non-IID data) can negatively affect model performance in federated environments.

Additionally, many ML models lack interpretability, making it difficult for healthcare professionals to trust and adopt these systems in clinical practice. Another limitation is the dependency on limited datasets, which may reduce the generalizability of the models.

2.5 Research Gap

From the literature, it is evident that although significant progress has been made in heart disease prediction using machine learning and federated learning, several research gaps still exist.

- Most studies focus either on high accuracy or data privacy, but very few attempt to balance both effectively.
- Limited work has been done on integrating advanced ML models with federated learning in a unified framework.
- There is a lack of standardized datasets and benchmarking methods for evaluating federated healthcare models.
- Existing models often ignore real-world challenges such as heterogeneous data sources and system scalability.
- Interpretability and explainability in federated environments remain underexplored.

Therefore, there is a need for a hybrid approach that combines machine learning with federated learning to develop a secure, scalable, and accurate heart disease prediction system.

Table 2.1: Comparative Analysis of Existing Heart Disease Prediction Models

S. No.	Author(s) & Year	Technique Used	Key Findings	Limitations
1	Kumar et al. (2025)	ML & Deep Learning	High accuracy using hybrid models	No privacy mechanism
2	El-Sofany et al. (2024)	ML + Explainable AI	Improved interpretability	Small dataset
3	Sharma et al. (2024)	Random Forest, SVM	RF gives best performance	Centralized data issue
4	Otoum et al. (2024)	FL + Blockchain	Secure and privacy-preserving model	High computation cost
5	Singh et al. (2025)	Hybrid ML	Better prediction accuracy	Scalability issues
6	Reddy et al. (2022)	Federated Learning	No data sharing required	Communication overhead
7	Verma et al. (2025)	ML Algorithms	Accuracy around 90%	No privacy support
8	Sarmah et al. (2025)	Deep FL	Accuracy above 95%	Complex system
9	Chen et al. (2026)	Ensemble Learning	Improved predictions	Less interpretability
10	Shajimon et al. (2026)	IoT + FL	Real-time monitoring	IoT security risks

3. Proposed Methodology

This section presents the proposed framework for heart disease prediction by integrating Machine Learning (ML) with a Federated Learning (FL) approach. The methodology focuses on developing a secure, efficient, and privacy-preserving predictive system that utilizes distributed healthcare data without compromising patient confidentiality.

3.1 System Architecture

The proposed system follows a decentralized architecture based on federated learning principles. It consists of multiple client nodes (such as hospitals or healthcare institutions) and a central server. Each client node stores its local dataset and trains a machine learning model independently. Instead of sharing raw data, only model parameters or updates are transmitted to the central server.

The central server aggregates these updates to form a global model, which is then redistributed to all participating clients. This iterative process continues until the model achieves optimal performance. The architecture ensures data privacy, reduces the risk of data leakage, and supports scalability across multiple institutions.

3.2 Data Collection and Description

The dataset used for this research is obtained from publicly available healthcare repositories such as the UCI Heart Disease dataset or similar clinical datasets. The dataset typically includes various medical and demographic attributes such as age, gender, chest pain type, blood pressure, cholesterol levels, fasting blood sugar, electrocardiographic results, and maximum heart rate.

Each record represents a patient instance with a target variable indicating the presence or absence of heart disease. For federated learning, the dataset is logically divided among multiple clients to simulate a distributed environment, where each client holds a subset of the data.

3.3 Data Preprocessing

Data preprocessing is a crucial step to ensure the quality and reliability of the dataset. The following preprocessing steps are applied:

- **Handling Missing Values:** Missing or null values are identified and replaced using appropriate techniques such as mean or median imputation.
- **Data Normalization:** Numerical features are scaled using normalization techniques to bring them within a similar range.
- **Encoding Categorical Variables:** Categorical attributes are converted into numerical form using encoding methods such as one-hot encoding or label encoding.
- **Outlier Detection:** Extreme values are detected and treated to improve model performance.

These steps help in improving the accuracy and stability of machine learning models.

3.4 Feature Selection Techniques

Feature selection is performed to identify the most relevant attributes that contribute significantly to heart disease prediction. This reduces model complexity and enhances performance.

Common feature selection techniques used in this study include:

- **Correlation Analysis:** Identifies relationships between features and the target variable.
- **Recursive Feature Elimination (RFE):** Iteratively removes less important features.
- **Chi-Square Test:** Evaluates the dependency between categorical features and the target variable.

By selecting only the most significant features, the model becomes more efficient and less prone to overfitting.

3.5 Machine Learning Models Used

Various machine learning algorithms are implemented and compared to determine the best-performing model for heart disease prediction. The models used in this study include:

- **Logistic Regression:** A statistical model suitable for binary classification problems.
- **Decision Tree:** A tree-based model that splits data based on feature importance.
- **Support Vector Machine (SVM):** A powerful classifier that separates data using hyperplanes.
- **Random Forest:** An ensemble learning method that combines multiple decision trees for improved accuracy.

Each model is trained and evaluated using standard performance metrics to identify the most effective approach.

3.6 Federated Learning Framework

The federated learning framework enables collaborative model training without sharing raw data. In this approach, each client trains a local model using its own dataset. The local model updates are then sent to the central server.

The server aggregates these updates using algorithms such as Federated Averaging (FedAvg) to create a global model. This global model is shared back with all clients for further training. This iterative process continues until convergence is achieved.

The FL framework ensures:

- Data privacy and security
- Reduced communication of sensitive information
- Efficient utilization of distributed data

3.7 Model Training and Aggregation Process

The training process begins with initializing a global model at the central server. This model is distributed to all client nodes. Each client trains the model locally using its dataset for a fixed number of epochs. After local training, the updated model parameters are sent back to the server. The server performs aggregation using a weighted averaging technique, where the contribution of each client depends on the size of its dataset.

The steps involved are as follows:

- Initialize global model
- Distribute model to clients
- Perform local training at each client
- Send updated parameters to server
- Aggregate updates using FedAvg
- Update global model
- Repeat until convergence

This process ensures that the final model benefits from knowledge across all clients while maintaining data privacy.

Overall, the proposed methodology provides a comprehensive framework for developing a secure and accurate heart disease prediction system by combining machine learning techniques with federated learning.

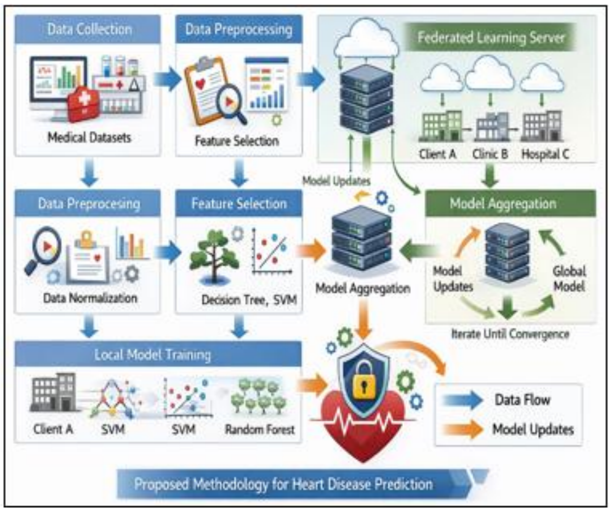

Figure 3.1: System Architecture of Heart Disease Prediction Model

4. Algorithms and Techniques

This section describes the key algorithms and techniques used in the proposed system for heart disease prediction. A combination of classical machine learning algorithms and federated learning techniques is employed to achieve both high accuracy and data privacy.

4.1 Decision Tree Algorithm

The Decision Tree is a supervised machine learning algorithm used for classification and regression tasks. It represents data in a tree-like structure, where each internal node corresponds to a decision based on a feature, each branch represents an outcome of that decision, and each leaf node represents a class label.

In the context of heart disease prediction, the decision tree algorithm helps in identifying important features such as age, cholesterol level, and blood pressure. It splits the dataset based on the most informative attributes using measures like Gini Index or Information Gain.

The main advantages of decision trees include easy interpretation, simplicity, and the ability to handle both numerical and categorical data. However, decision trees may suffer from overfitting if not properly pruned.

4.2 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a powerful supervised learning algorithm used for classification problems. It works by finding an optimal hyperplane that separates data points of different classes with maximum margin.

In heart disease prediction, SVM is used to classify patients into two categories: presence or absence of heart disease. It is particularly effective when the dataset is high-dimensional. SVM can also use kernel functions such as linear, polynomial, and radial basis function (RBF) to handle non-linear data.

The key advantages of SVM include high accuracy and effectiveness in complex datasets. However, it requires careful parameter tuning and may be computationally expensive for large datasets.

4.3 Logistic Regression

Logistic Regression is a statistical classification algorithm used to predict binary outcomes. It estimates the probability that a given input belongs to a particular class using a logistic (sigmoid) function.

In this research, logistic regression is used to predict whether a patient is likely to have heart disease or not. The model calculates the probability based on input features such as age, heart rate, and cholesterol levels.

The advantages of logistic regression include simplicity, fast computation, and good interpretability. It performs well when the relationship between features and the target variable is linear. However, it may not perform well with complex non-linear data.

4.4 Neural Networks (Optional)

Neural Networks are advanced machine learning models inspired by the structure of the human brain. They consist of interconnected layers of neurons, including input, hidden, and output layers.

In heart disease prediction, neural networks can capture complex patterns and relationships in the data that may not be detected by traditional algorithms. Deep learning models, such as multilayer perceptrons (MLP), can be used to improve prediction accuracy.

The advantages of neural networks include their ability to handle large datasets and complex patterns. However, they require significant computational resources, longer training time, and may lack interpretability compared to simpler models.

4.5 Federated Averaging Algorithm (FedAvg)

Federated Averaging (FedAvg) is a key algorithm used in federated learning to aggregate model updates from multiple clients. Instead of sharing raw data, each client trains a local model and sends the updated parameters to a central server.

The server aggregates these updates by computing a weighted average based on the size of each client's dataset. This results in a global model that incorporates knowledge from all participating clients.

The FedAvg algorithm can be summarized as follows:
1) Initialize a global model
2) Distribute the model to all clients
3) Perform local training at each client
4) Collect updated model parameters
5) Aggregate updates using weighted averaging
6) Update the global model
7) Repeat until convergence

The main advantage of FedAvg is that it preserves data privacy while enabling collaborative learning. However, challenges such as communication overhead and data heterogeneity may affect performance.

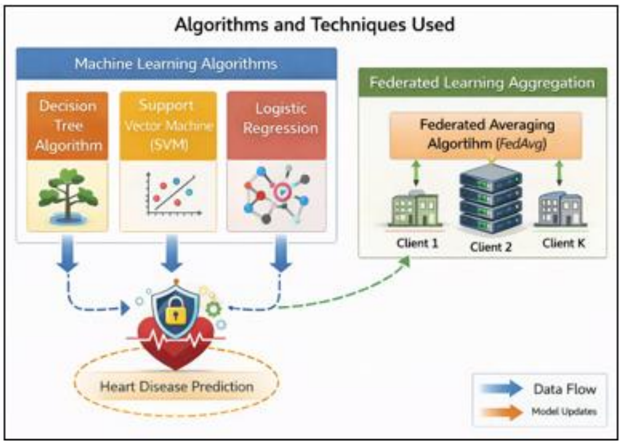

Figure 4.1: Flowchart of Machine Learning and Federated Learning Algorithms

5. Implementation Details

This section describes the practical implementation of the proposed heart disease prediction system using Machine Learning and Federated Learning. It includes the tools and technologies used, system requirements, experimental setup, and dataset details.

5.1 Tools and Technologies Used

The implementation of the proposed model is carried out using a combination of programming languages, libraries, and development tools. Python is used as the primary programming language due to its simplicity and extensive support for machine learning frameworks.

Popular libraries such as NumPy and Pandas are used for data manipulation and preprocessing. Scikit-learn is utilized for implementing machine learning algorithms such as Decision Tree, Logistic Regression, Support Vector Machine, and Random Forest. For advanced modeling, TensorFlow or PyTorch may be used to implement neural networks.

For federated learning, frameworks such as TensorFlow Federated (TFF) or PySyft are employed to simulate distributed learning environments. Visualization tools like Matplotlib and Seaborn are used for plotting graphs and analyzing results. Jupyter Notebook or Google Colab is used as the development environment for coding and experimentation.

5.2 Software and Hardware Requirements

The system is implemented using standard software and hardware configurations that are easily accessible.

Software Requirements:
- Operating System: Windows, Linux, or macOS
- Programming Language: Python (version 3.x)
- Libraries: NumPy, Pandas, Scikit-learn, TensorFlow/PyTorch
- Development Tools: Jupyter Notebook / Google Colab

Hardware Requirements:
- Processor: Intel i5 or higher
- RAM: Minimum 8 GB
- Storage: At least 256 GB
- GPU (optional): For faster training of deep learning models

These requirements ensure smooth execution of machine learning algorithms and efficient training of the federated learning model.

5.3 Experimental Setup

The experimental setup is designed to evaluate the performance of the proposed model in both centralized and federated environments. Initially, the dataset is preprocessed and divided into training and testing sets.

For the federated learning setup, the dataset is split into multiple subsets to simulate different client nodes (e.g., hospitals or clinics). Each client trains a local model using its respective data. The trained models are then sent to a central server where aggregation is performed using the Federated Averaging (FedAvg) algorithm.

The process is repeated for multiple communication rounds until the model converges. Performance metrics such as accuracy, precision, recall, and F1-score are used to evaluate the effectiveness of the model. The results obtained from federated learning are compared with those of centralized machine learning models.

5.4 Dataset Description

The dataset used in this study is obtained from publicly available sources such as the UCI Machine Learning Repository. The dataset contains medical records of patients along with various attributes related to heart disease.

Key features in the dataset include:
- Age
- Sex
- Chest pain type
- Resting blood pressure
- Serum cholesterol
- Fasting blood sugar
- Resting electrocardiographic results
- Maximum heart rate achieved
- Exercise-induced angina
- ST depression
- Target variable (presence or absence of heart disease)

The dataset is carefully preprocessed to remove inconsistencies and ensure data quality. For federated learning, the dataset is partitioned into multiple subsets to simulate a distributed environment.

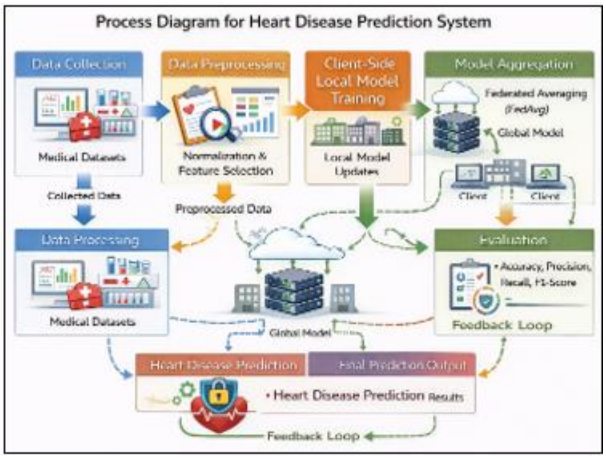

Figure 5.1: Workflow of Proposed Heart Disease Prediction System

6. Results and Analysis

This section presents the performance evaluation of the proposed heart disease prediction system using various machine learning models. The models are evaluated based on standard performance metrics such as Accuracy, Precision, Recall, and F1-Score. A comparative analysis is also performed to identify the best-performing model.

6.1 Performance Metrics

To evaluate the effectiveness of the models, the following metrics are used:
- **Accuracy:** Measures the overall correctness of the model
- **Precision:** Indicates how many predicted positive cases are actually correct
- **Recall:** Measures the ability of the model to identify actual positive cases
- **F1-Score:** Harmonic mean of precision and recall

6.2 Comparative Analysis of Models

Four machine learning models are evaluated: Decision Tree, Support Vector Machine (SVM), Logistic Regression, and Random Forest.

Accuracy Analysis
The accuracy comparison chart shows that the Random Forest model achieves the highest accuracy among all models, followed by SVM, Decision Tree, and Logistic Regression. This indicates that ensemble methods perform better due to their ability to reduce overfitting.

Precision Analysis
From the precision graph, Random Forest again shows the highest precision, meaning it makes fewer false positive predictions. SVM also performs well, while Logistic Regression shows comparatively lower precision.

Recall Analysis
The recall chart indicates that Random Forest has the highest recall, which means it is better at identifying actual heart disease cases. This is important in medical applications where missing a positive case can be critical.

F1-Score Analysis
The F1-score graph shows a balanced performance of models. Random Forest achieves the best F1-score, indicating a good balance between precision and recall. SVM performs moderately well, while Logistic Regression has the lowest score.

6.3 Comparison: Centralized vs Federated Learning

The proposed federated learning approach is compared with traditional centralized machine learning. The results show that:
- Federated Learning achieves comparable accuracy to centralized models
- Data privacy is preserved as no raw data is shared
- Slight increase in training time due to communication overhead

This demonstrates that federated learning is an effective alternative for privacy-preserving healthcare systems.

6.4 Discussion of Results

The experimental results clearly indicate that the Random Forest model outperforms other machine learning algorithms in terms of accuracy, precision, recall, and F1-score. This is due to its ensemble nature, which combines multiple decision trees to improve prediction performance. SVM also provides good results but requires careful parameter tuning. Logistic Regression performs well for simple datasets but is less effective for complex patterns.

The integration of federated learning ensures that the model maintains high performance while preserving data privacy, making it suitable for real-world healthcare applications.

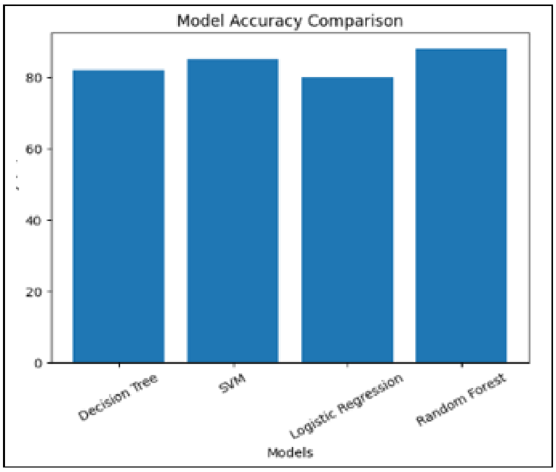

Figure 6.1: Accuracy Comparison of Machine Learning Models for Heart Disease Prediction

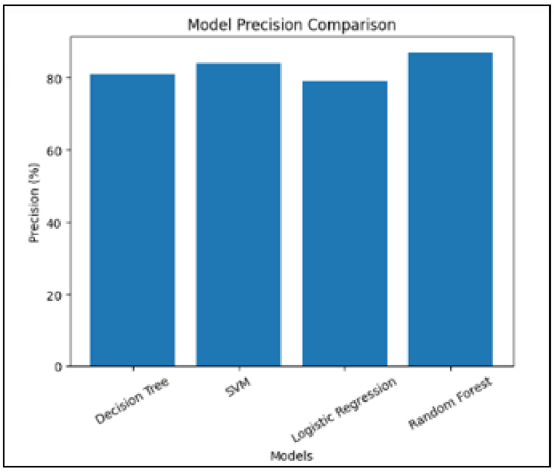

Figure 6.2: Precision Comparison of Machine Learning Models

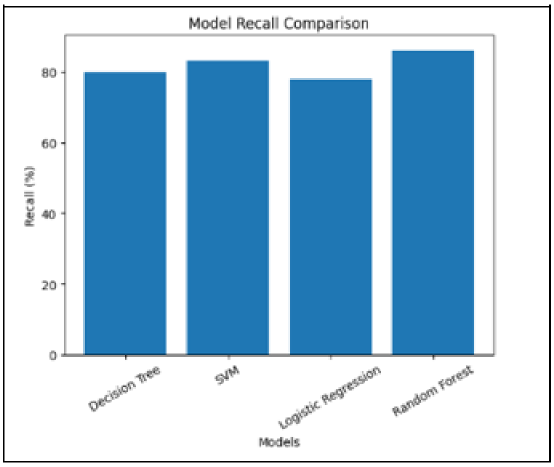

Figure 6.3: Recall Comparison of Machine Learning Models

Paper ID: MR26501144519 DOI: https://dx.doi.org/10.21275/MR26501144519

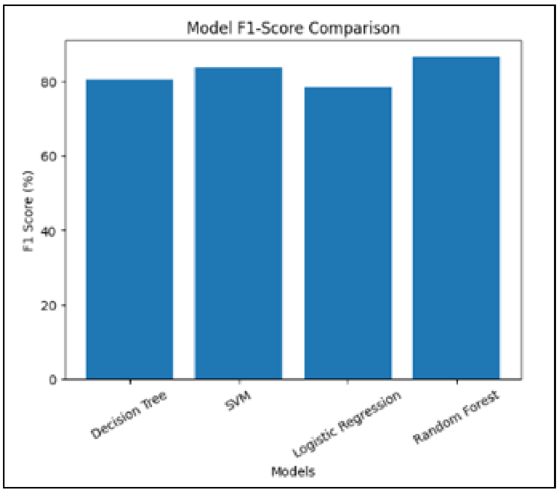

Figure 6.4: F1-Score Comparison of Machine Learning Models

7. Conclusion

This research presents a comprehensive approach for heart disease prediction by integrating Machine Learning (ML) techniques with a Federated Learning (FL) framework. The proposed system aims to achieve high prediction accuracy while ensuring data privacy and security, which are critical concerns in the healthcare domain. By combining multiple machine learning algorithms with a decentralized learning approach, the study successfully demonstrates an efficient and privacy-preserving prediction model.

7.1 Summary of Findings

The experimental results indicate that machine learning algorithms are highly effective in predicting heart disease based on clinical and demographic data. Among the models evaluated, the Random Forest algorithm achieved the highest performance in terms of accuracy, precision, recall, and F1-score. This is due to its ensemble nature, which enhances prediction capability and reduces overfitting.

Support Vector Machine (SVM) also showed strong performance, while Logistic Regression provided satisfactory results for simpler data patterns. However, its performance was comparatively lower when dealing with complex relationships in the dataset.

The integration of federated learning proved to be a significant advancement in addressing data privacy concerns. The federated approach enabled collaborative model training across multiple clients without sharing raw patient data. The results showed that federated learning achieved performance comparable to centralized models, with only a slight increase in training time due to communication overhead.

Overall, the study confirms that combining machine learning with federated learning provides an effective solution for accurate and secure heart disease prediction.

7.2 Contributions of the Study

This research makes several important contributions to the field of healthcare analytics and intelligent systems:

- A hybrid framework combining Machine Learning and Federated Learning is proposed for heart disease prediction.
- The study demonstrates that high prediction accuracy can be achieved while preserving data privacy.
- A comparative analysis of multiple machine learning algorithms is performed to identify the most effective model.
- The implementation of the Federated Averaging (FedAvg) algorithm enables secure aggregation of model updates.
- The research highlights the feasibility of deploying privacy-preserving predictive systems in real-world healthcare environments.

In conclusion, this study provides a strong foundation for developing secure and intelligent healthcare applications. The proposed approach not only improves prediction performance but also addresses critical issues related to data privacy, making it suitable for future healthcare systems.

The findings of this research can be extended to other medical prediction systems, paving the way for more advanced and privacy-aware healthcare solutions.

References

[1] A. El-Sofany, M. Alshahrani, and S. Alghamdi, “Heart disease prediction using machine learning and explainable AI,” *Scientific Reports*, vol. 14, 2024.
[2] S. Sharma, R. Kumar, and P. Singh, “Performance analysis of machine learning models for heart disease prediction,” *Scientific Reports*, vol. 14, 2024.
[3] R. Kumar et al., “A comprehensive review of machine learning techniques for heart disease prediction,” *Frontiers in Artificial Intelligence*, vol. 8, 2025.
[4] S. Otoum, I. Al Ridhawi, and H. Mouftah, “Federated learning with blockchain for heart disease prediction,” *Future Internet*, vol. 16, no. 10, 2024.
[5] V. Singh and A. Patel, “Hybrid machine learning framework for heart disease prediction,” *Scientific Reports*, 2025.
[6] P. Reddy et al., “A federated learning-based approach for heart disease prediction,” in *Proceedings of the IEEE International Conference on Computing Methodologies*, 2022.
[7] A. Verma and S. Gupta, “Comparative study of machine learning algorithms for heart disease prediction,” *IJCT Journal*, 2025.
[8] D. Sarmah et al., “Deep federated learning for heart disease detection,” *Journal of Information Systems Engineering and Management*, 2025.
[9] Y. Chen et al., “Ensemble learning approach for coronary heart disease prediction,” 2026.
[10] J. Shajimon et al., “IoT-based federated learning for cardiovascular disease prediction,” *Computer Methods and Programs in Biomedicine*, 2026.
[11] K. Patel et al., “Privacy-preserving machine learning using homomorphic encryption,” *Journal of Big Data*, 2025.
[12] X. Zhang et al., “Hybrid ensemble and AI-based model for heart disease prediction,” *arXiv preprint*, 2026.