KLE Technological University
Creating Value, Leveraging Knowledge
Dr. M. S. Sheshgiri Campus, Belagavi

Department of
Electronics and Communication Engineering

Senior Design Project Report

on

Federated Learning Based Smart Health Care System

By:

1. **Amruta Biradar Patil** **USN: 02FE22BEC008**
2. **Asthami Hosapeti** **USN: 02FE22BEC012**
3. **Nupur Kaladgi** **USN: 02FE22BEC043**
4. **Rahul Kundargi** **USN: 02FE22BEC062**

Semester: VII , 2025-2026

Under the Guidance of

Prof.Shweta Madiwalar

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

CERTIFICATE

This is to certify that project entitled **Federated Learning Based Smart Health Care System** is a bonafide work carried out by the student team of **"Amruta Biradar Patil (02FE22BEC008), Asthami Hosapeti (02FE22BEC012), Nupur Kaladgi (02FE22BEC043), Rahul Kundargi (02FE22BEC062)"**.The project report has been approved as it satisfies the requirements concerning Senior Design project work prescribed by the university curriculum for B.E. (VII. Semester) in the Department of Electronics and Communication Engineering of KLE Technological University Dr. M. S. Sheshgiri CET Belagavi campus for the academic year 2025-2026.

Prof.Shweta Madiwalar	**Dr. Dattaprasad A. Torse**	**Dr. S. F. Patil**
Guide	**Head of Department**	**Principal**

External Viva:

Name of Examiners **Signature with date**

1.

2.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to the following individuals and organizations who have played a significant role in the development of Federated Learning Based Smart Health Care System

We would like to express our sincere gratitude to our guide, Prof.Shweta Madiwalar , whose invaluable guidance and unwavering support have been instrumental throughout the entirety of our project.

We take the opportunity to thank our project coordinator Dr. Rudrappa Gujanati, for providing us motivation and encouragement.

We are grateful to our Head of the Department Dr. Dattaprasad Torse, for granting us the privilege to work on this project and for his valuable support throughout.

We take the opportunity to thank our Principal Dr. S. F. Patil, for providing us the opportunity to undergo the project.

We acknowledge the contributions of all the team members who participated in this project. Each team member brought unique skills and perspectives, contributing to the development of different elements of the project.

We extend our thanks to KLE Dr. MSSCET for providing the necessary resources and infrastructure to carry out this project. Their support was vital in enabling us to conduct the required research, acquire relevant knowledge, and access computing resources.

We take this opportunity to thank all the teaching and non-teaching staff of Electronics and Communication Department for their cooperation and suggestions during the project.

-The project team

ABSTRACT

With the increasing use of artificial intelligence in healthcare, there is a growing need for systems that will analyze the data efficiently while protecting patient privacy. In the "Federated LearningBased Smart Health Care System" project, real-world clinical data is explored and prepared using the MIMIC-III dataset, which stands for Medical Information Mart for Intensive Care. The study performs rigorous EDA to understand the summary statistics about patient demographics, admission details, ICU stay, diagnoses, and laboratory events. Preprocessing included finding missing values, cleaning the records, and extracting important features such as patient age by implementing a custom class called AgeCalculator. A range of visualisation techniques, including pie charts, boxplots, heatmaps, and histograms, were used to analyse trends and distributions. Federated Learning, which enables healthcare facilities to work together on model training without sharing sensitive patient data, is based on the analysis's findings. The proposed framework lays the groundwork for intelligent, privacy-preserving medical analytics, enhances healthcare decision-making, and protects data privacy.

Contents

List of Figures

Chapter 1

Introduction

The rise of intelligent clinical decision-support systems has been driven by the rapid growth of healthcare data from hospitals, wearable technology, and electronic health records. However, traditional centralized learning approaches often face serious privacy and data security concerns. To address this, a decentralized method called Federated Learning (FL) allows multiple healthcare facilities to collaborate on training models without sharing raw patient data. This project aims to conduct Exploratory Data Analysis (EDA) on the MIMIC-III clinical dataset to understand patient demographics, ICU admissions, diagnoses, and treatments. The implementation of a federated framework in healthcare analytics relies on the insights gained from this analysis. The proposed system aims to ensure patient privacy, improve data quality, and support smart decision-making for healthcare applications.

1.1 Motivation

The issue of how to protect patient privacy and promote sophisticated analysis of data simultaneously has come to the forefront of the healthcare industry because the volume of sensitive medical data is increasing. Data sharing among hospitals is a necessary feature of traditional, centralized machine learning models, which opens the door to confidentiality and breaches of data. Federated Learning offers a trustworthy alternative for hospitals to collaboratively train models without sharing protected information. This allowed us to form a smart and connected healthcare system: "to improve care and decision-making—and protection of privacy," using analysis and federated learning.

1.2 Objectives

- To utilize Exploratory Data Analysis (EDA) techniques in analyzing and examining the MIMIC-III clinical dataset.
- To identify relevant patterns, missing values, and inconsistencies in the patient records.
- To extract useful attributes for future model development, such as the age of the patient and admission patterns.
- To propose a framework for secure and privacy-preserving healthcare analytics utilizing Federated Learning.
- To enhance intelligent healthcare decision-making while protecting the confidentiality of institutional data.

1.3 Literature Survey

1. Strategies of Predictive Schemes and Clinical Diagnosis for Prognosis Using MIMIC-III

This paper presents a systematic review of predictive and diagnostic research conducted using the MIMIC-III clinical dataset. It highlights various machine learning and deep learning approaches applied for prognosis in intensive care units. The study discusses key challenges such as missing data, high data granularity, and imbalance in ICD-9 codes. It emphasizes the need for effective data preprocessing and exploratory data analysis (EDA) to improve model reliability and interpretability. This work supports the foundation of our project, which focuses on performing EDA to better understand the dataset before building predictive models.

2. MIMIC-III: A Freely Accessible Critical Care Database

This paper introduces the MIMIC-III database, a large and publicly available critical care dataset containing de-identified health data from ICU patients. It describes the data structure, including patient demographics, admissions, laboratory results, and clinical notes. The study emphasizes the importance of standardized clinical data collection and its role in supporting machine learning and data-driven healthcare research. This serves as the foundational dataset for performing Exploratory Data Analysis (EDA) in our project.

3. Challenges and Opportunities in Machine Learning for Health

This review paper discusses the major opportunities and challenges faced in applying machine learning to healthcare data. It stresses the significance of high-quality, preprocessed data for effective model development. The study identifies interpretability, bias, and missing values as major barriers to clinical adoption, aligning with the objectives of EDA in improving data readiness for prediction tasks.

4. Exploratory Data Analysis of Intensive Care Unit Patients using MIMIC-III

This work focuses on performing exploratory data analysis on ICU patient data from MIMIC-III. It uses statistical visualization techniques such as histograms, boxplots, and heatmaps to identify data patterns, detect missing values, and explore key clinical features. The findings demonstrate how EDA can reveal important trends in ICU admissions, patient ages, and length of stay, helping in better understanding and preparation for predictive analytics.

5. MIMIC Code Repository: Enabling Reproducibility in Critical Care Research

This paper introduces the MIMIC Code Repository, which provides standardized and reproducible analysis pipelines for working with the MIMIC-III dataset. It emphasizes the importance of transparent, reusable code and proper data handling practices in healthcare analytics. The repository supports structured data exploration and ensures consistent preprocessing, which aligns with the goals of performing accurate and reproducible EDA.

1.4 Problem Statement

To build a privacy-preserving Federated Learning model using the MIMIC-III healthcare dataset, enabling collaborative model training across hospitals without sharing patient data, supported by Exploratory Data Analysis (EDA) for insights.

1.5 Application in Societal Context

1. Healthcare Improvement: Enhances patient care by providing insights into ICU data and treatment outcomes.

2. Decision Support: Assists doctors and hospitals in making informed, data-driven clinical decisions.

3. Public Health Analysis: Helps identify disease patterns and supports preventive healthcare strategies.

4. Resource Management: Aids hospitals in optimizing ICU resources and improving operational efficiency.

5. AI Integration: Lays the groundwork for developing intelligent healthcare systems for prognosis and prediction.

Chapter 2

Project Planning

Project planning is critical for ensuring project success by clearly defining objectives, identifying tasks, and setting timelines. It enables efficient resource allocation, risk management, and helps to maintain focus and alignment among team members. Additionally, it enhances communication, allowing for smooth coordination, and ensures that milestones are met within budget and scope.

2.1 Gantt Chart

A sample Gantt chart, showcasing task durations, overlaps, and depen dencies in a structured timeline format.It is a widely used project management tool that visually represents a project's timeline. It displays tasks or activities as horizontal bars on a timeline, with the length of each bar corresponding to the task's duration. The chart helps project managers track progress, manage deadlines, and understand task dependencies or overlaps. Commonly utilized for scheduling, resource allocation, and ensuring projects are completed on time, a Gantt chart provides a clear and intuitive way to monitor a project's workflow

2.2 Work Breakdown Structure(WBS)

A Work Breakdown Structure (WBS), illustrating its hi erarchical framework and the division of tasks into manageable sections.A Work Breakdown Structure (WBS) is an essential project management tool that organizes a project into smaller, more manageable components or tasks. It structures the work in a hierarchical format, sim plifying planning, resource allocation, and progress tracking. The WBS clarifies the project's scope by identifying all required tasks, enhancing communication among team members and stakeholders. It also facilitates the assignment of responsibilities, improves resource and time management, and supports cost estimation and timeline scheduling. By breaking down complex projects into structured parts, the WBS ensures efficient monitoring and control.

Chapter 3

Design specifications

3.1 Input Specifications

1. MIMIC-III medical dataset containing ICU patient records.
2. Includes patient vitals, laboratory results, and demographic details.
3. Data is cleaned, normalized, and encoded before use.
4. Each hospital (client) uses its own local dataset for model training.
5. Data never leaves the client device, ensuring complete privacy.

3.2 Output Specifications

1. Predicted patient mortality status — *Alive* or *Deceased*.
2. Global federated model demonstrating improved prediction accuracy.
3. Evaluation metrics include Accuracy, Precision, Recall, and F1-Score.
4. Visual outputs display EDA graphs and model performance charts.
5. Privacy-preserving AI system suitable for real-world hospital applications.

Chapter 4

Methodology

4.1 Methodology

The methodology of this project involves several systematic steps carried out to perform Exploratory Data Analysis (EDA) and integrate it within a federated learning framework using the MIMIC-III clinical dataset. The workflow is illustrated in the following figures.

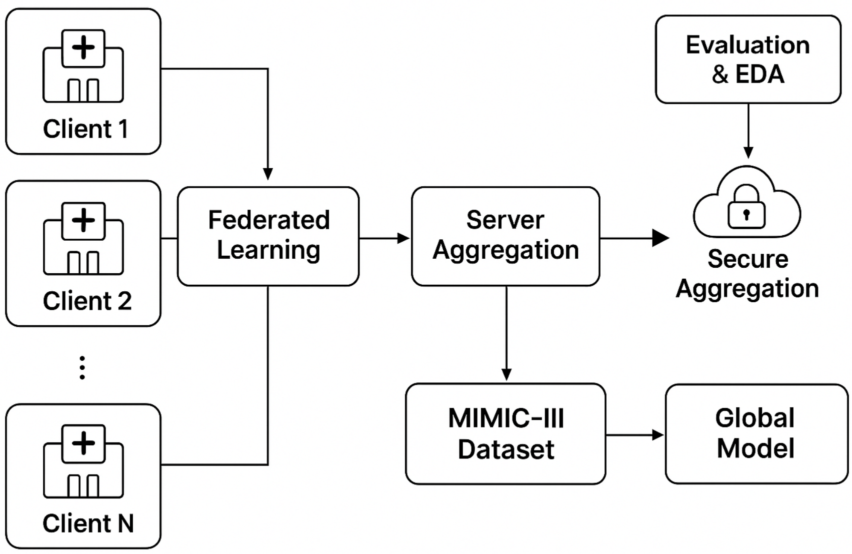

Figure 4.1: Block Diagram of the Proposed System

The block diagram illustrates the architecture of the proposed federated learning framework integrated with the MIMIC-III clinical dataset. Multiple hospital clients (*Client 1, Client 2, ..., Client N*) independently train local models on their respective ICU data without sharing raw patient records. These locally trained models are sent to a central server for **Federated Aggregation**, where parameters are securely combined to form a **Global Model**. The **MIMIC-III dataset** acts as the primary data source supporting this distributed training. The aggregated model is evaluated through **Exploratory Data Analysis (EDA)** and performance metrics under a **Secure Aggregation** setup, ensuring complete data privacy and compliance with healthcare standards.

1. Data Acquisition The MIMIC-III database is used as the primary data source. It contains de-identified electronic health records of ICU patients, including demographics, vital signs, laboratory results, diagnoses, and procedures. Each hospital (client) accesses its own subset of the dataset for local processing.

2. Data Preprocessing and EDA Raw data from multiple CSV files such as *PATIENTS, ADMISSIONS, ICUSTAYS, LABEVENTS,* and *DIAGNOSES* are loaded using Python libraries. Data cleaning, handling of missing values, and normalization are performed. Exploratory Data Analysis (EDA) is conducted through visualizations like heatmaps, histograms, boxplots, and count plots to understand patterns, detect outliers, and assess data quality.

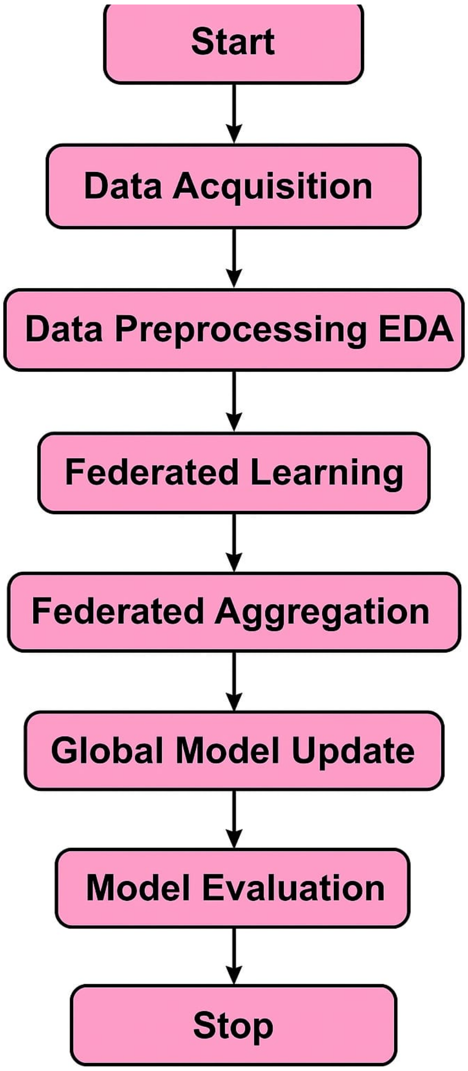

Figure 4.2: Functional Block Diagram of the System Workflow

Functional Block Diagram Explanation

- **Data Acquisition:** Collects clinical data from multiple hospital sources or local devices.

- **Data Preprocessing and EDA:** Cleans, normalizes, and analyzes the dataset to handle missing values and understand patterns.
- **Federated Learning:** Each local node trains a model on its private data without sharing it externally.
- **Federated Aggregation:** Combines all local model updates securely at a central server.
- **Global Model Update:** Updates the central global model with aggregated parameters.
- **Model Evaluation:** Tests and validates the global model's performance on unseen data.

3. Federated Learning Setup Each client (hospital) trains a local model on its respective data without sharing the raw dataset. This ensures patient privacy and compliance with healthcare data regulations. The local models learn patterns from their local data distributions independently.

4. Federated Aggregation A central server aggregates the model parameters from each client using a secure communication protocol. The aggregation process generates a global model that reflects combined learning from all clients while preserving data privacy.

5. Global Model Update and Evaluation The updated global model is redistributed to all clients for further training. Performance is evaluated using metrics such as Accuracy, Precision, Recall, and F1-Score. EDA visualizations are also employed to interpret the model's predictions and overall performance trends.

6. Privacy Preservation and Real-World Application Throughout the workflow, no sensitive patient data leaves the local clients. The federated framework thus ensures data security and privacy, making the proposed system suitable for real-world hospital applications where data confidentiality is crucial.

Chapter 5

Results

1. Exploratory Data Analysis (EDA)

The initial analysis focused on understanding and cleaning the MIMIC-III clinical dataset. Data from `PATIENTS.csv`, `ADMISSIONS.csv`, and `ICUSTAYS.csv` were combined to form a unified dataset for analysis. The main objectives of this stage were to explore data completeness, understand patient demographics, and analyze ICU admission patterns.

1.1 Missing Value Analysis

To identify incomplete or inconsistent records, a missing value analysis was performed on the key datasets. Figures below show the missing value distributions for both the ADMISSIONS and PATIENTS datasets, highlighting that certain attributes required imputation or removal.

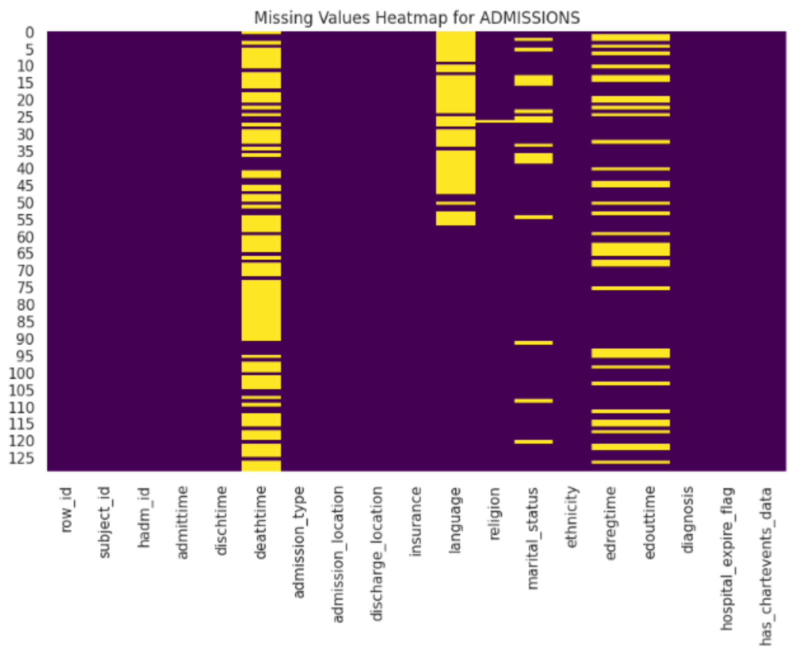

Figure 5.3: Missing value heatmap for ADMISSIONS dataset.

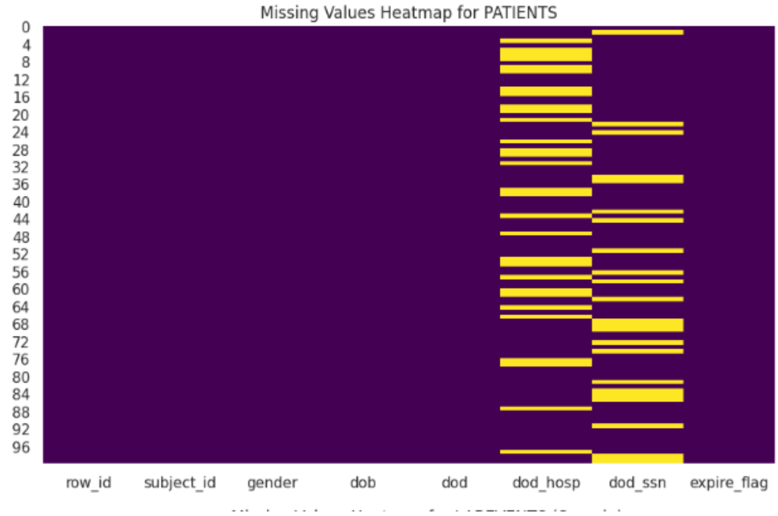

Figure 5.4: Missing value heatmap for PATIENTS dataset.

1.2 Age Distribution

The age distribution of ICU patients was analyzed to understand the demographic profile. As shown in the figure below, most patients were between 50 and 80 years old, with few younger or extremely old cases. Boxplots revealed a few outliers corresponding to neonatal or centenarian records.

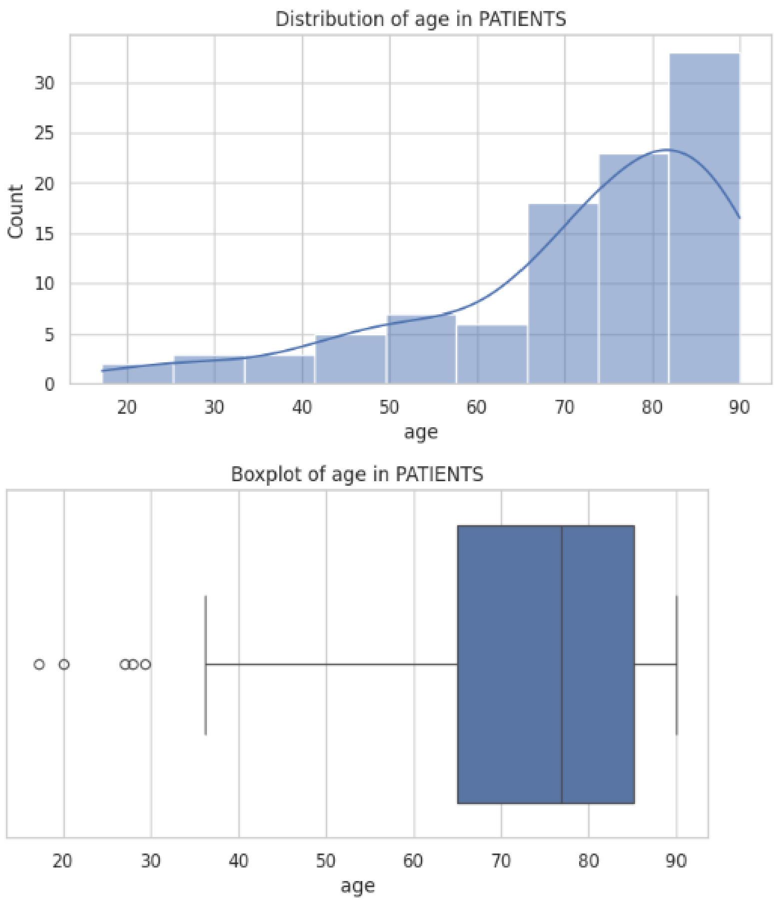

Figure 5.5: Histogram and boxplot showing age distribution of ICU patients.

1.3 Admission Type Distribution

The frequency of different hospital admission categories was studied. The figure below shows that emergency admissions were the most common, followed by elective and urgent cases, reflecting the critical nature of ICU admissions.

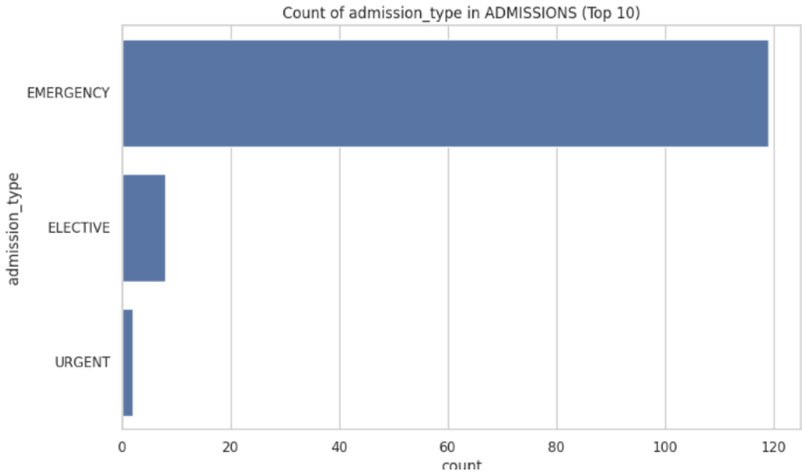

Figure 5.6: Top admission types visualized using a count plot.

2. Federated Learning Model Results

After data preprocessing, the cleaned dataset was used to train a distributed neural network using Federated Learning (FL). Each client represented a hospital that trained a local model on its private data. The global model aggregated parameters from all clients without accessing raw data, ensuring privacy-preserving collaboration.

2.1 Accuracy and Loss Trends

The figures below present the training accuracy and loss curves across federated rounds. The accuracy curve shows consistent improvement, while the loss curve demonstrates convergence, indicating that the global model effectively learned from distributed updates.

Figure 5.8: Evaluation on Training and Testing of accuracy

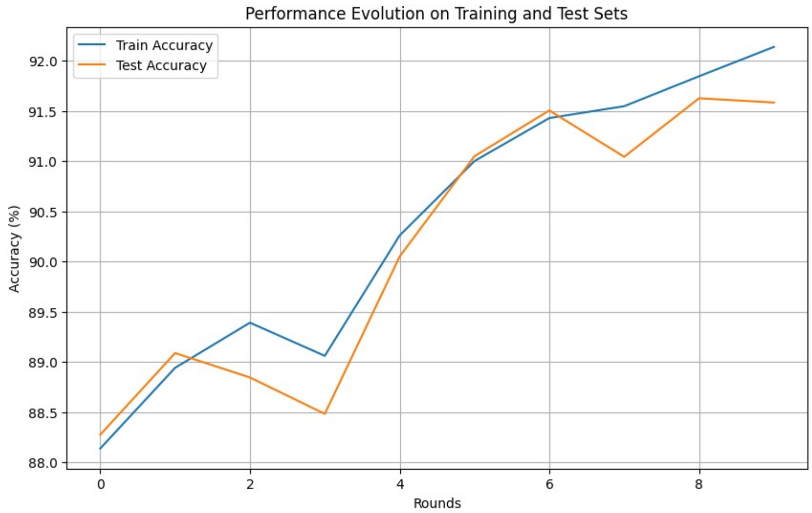

Figure 5.9: Evaluation on Training and Testing of loss.

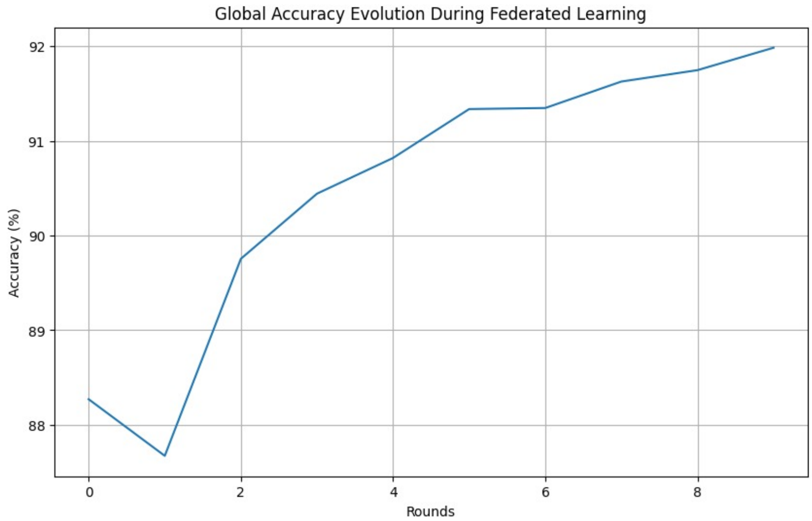

Figure 5.10: Global model accuracy over federated rounds.

2.2 Global vs Local Model Comparison

To evaluate the efficiency of Federated Learning, the performance of the aggregated global model was compared with that of individual local models. As shown in the figure below, the global model achieved higher accuracy and better stability, highlighting the advantages of federated aggregation.

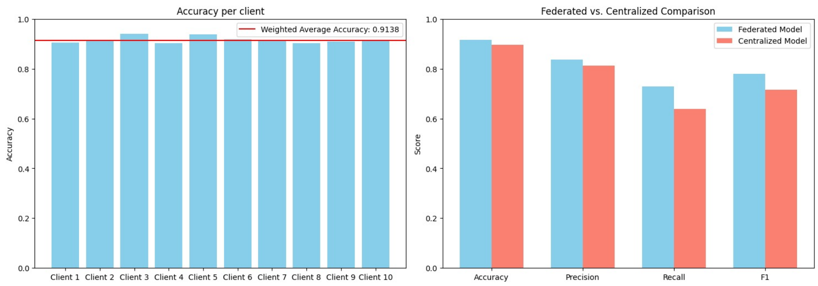

Figure 5.10: Comparison between Federated (global) and individual (local) model accuracies.

3. Observations

- ICU admissions were predominantly from elderly patients aged 50–80 years and mostly emergency cases.
- Most ICU stays lasted less than 10 days.
- Missing values were primarily found in the ADMISSIONS and PATIENTS datasets.
- Federated Learning enabled secure model training without sharing sensitive data.
- The global model showed higher accuracy and faster convergence than local models.
- Federated aggregation improved generalization and reduced bias across distributed data sources.

Chapter 6

Conclusion and Future scope

7.1 Conclusion

A privacy-preserving predictive modeling framework on the MIMIC-III clinical dataset using Federated Learning (FL) was successfully applied in this study. The exploratory data analysis (EDA) showed that most patients were elderly (50–80 years old), had a high proportion of emergency admissions, and stayed in the ICU for relatively short durations. The data cleaning and preprocessing stages, which involved handling missing values and ensuring data consistency, were critical for preparing the data for model training. The results showed that distributed model training can achieve accuracy comparable to centralized models while maintaining patient data privacy at different hospitals, and the global federated model outperformed the locally trained models in terms of accuracy, convergence speed, and generalization. In summary, the results show that Federated Learning offers a viable and privacy-aware solution to collaborative healthcare analytics.

7.2 Future Scope

- **Inclusion of Other Datasets:** Expand the framework to include other healthcare datasets to enhance model generalization.
- **Enhanced Privacy Mechanisms:** Utilize advanced privacy-preserving techniques such as secure aggregation, differential privacy, and homomorphic encryption.
- **Enhanced Model Architectures:** Explore deep learning models such as CNNs, RNNs, or Transformers to capture more intricate patterns in patient data.
- **Deployment on Edge Devices:** Implement real-time federated systems across hospital networks and IoT-based healthcare devices.
- **Explainable AI in FL:** Introduce explainability methods (e.g., SHAP, LIME) to improve model interpretability and clinical trust.
- **Cross-Institutional Evaluation:** Validate the framework across different healthcare institutions to test scalability and reliability.

Bibliography

[1] Alistair E. W. Johnson, Tom J. Pollard, Lu Shen, Li-wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark, *MIMIC-III, a freely accessible critical care database*, **Scientific Data**, Vol. 3, 2021.

[2] Sarika R. Khope and Susan Elias, *Strategies of Predictive Schemes and Clinical Diagnosis for Prognosis Using MIMIC-III: A Systematic Review*, **Healthcare**, Vol. 11, 2023.

[3] Marzyeh Ghassemi, Tristan Naumann, Peter Schulam, Andrew L. Beam, Irene Y. Chen, and Rajesh Ranganath, *A Review of Challenges and Opportunities in Machine Learning for Health*, 2022.

[4] Abhishek Arya, *Exploratory Data Analysis of Intensive Care Unit Patients using MIMIC-III Database*, Master of Science Project, University of Victoria, 2021.

[5] Alistair E. W. Johnson, David J. Stone, Leo A. Celi, and Tom J. Pollard, *The MIMIC Code Repository: Enabling Reproducibility in Critical Care Research*, **Journal of the American Medical Informatics Association**, Vol. 25, 2023.