DECENTRALIZED HORIZONTAL FEDERATED LEARNING USING GRAPH NEURAL NETWORKS FOR ECG SIGNAL ANALYSIS

BY

IBRAHIM PATWARY KHOKAN
ID: 202-15-3802
AND

TANVIR MAHTAB TAPU
ID: 202-15-3800

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

DR. SHEAK RASHED HAIDER NOORI
Professor & Head
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

DR. MD ZAHID HASAN
Associate Professor
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

July 2024

APPROVAL

This Project titled "Decentralized Horizontal Federated Learning Using Graph Neural Networks for ECG Signal Analysis", submitted by Ibrahim Patwary Khokan and Tanvir Mahtab Tapu to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 15-07-2024.

BOARD OF EXAMINERS

Hossain 15.07.2024

Dr. Md. Fokhray Hossain (MFH) — **Board Chairman**
Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Firoz Hasan 15.7.24

Md. Firoz Hasan (FH) — **Internal Examiner 1**
Sr. Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Lamia Rukhsara 15.07.24

Lamia Rukhsara (LR) — **Internal Examiner 2**
Sr. Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

7.24

Dr. S. M. Hasan Mahmud (SMH) — **External Examiner**
Assistant Professor
Department of Computer Science
American International University-Bangladesh

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Dr. Sheak Rashed Haider Noori, Professor & Head, Department of Computer Science and Engineering, Daffodil International University**. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Dr. Sheak Rashed Haider Noori
Professor & Head
Department of CSE
Daffodil International University

Co-Supervised by:

Dr. Md Zahid Hasan
Associate Professor
Department of CSE
Daffodil International University

Submitted by:

Ibrahim

Ibrahim Patwary Khokan
ID: 202-15-3802
Department of CSE
Daffodil International University

Tanvir.

Tanvir Mahtab Tapu
ID: 202-15-3800
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty for His divine blessing making it possible for us to complete the final year project successfully.

We are grateful and wish our profound indebtedness to **Dr. Sheak Rashed Haider Noori**, **Professor & Head**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "Machine learning" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to the **Head of the Department of CSE**, for his kind help in finishing our project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

Data privacy is a major concern in data analysis due to ethical and legal considerations. To overcome this issue, federated learning has transformed data analysis by putting data privacy first throughout the model training process. We proposed a horizontal federated learning system in this study after analyzing ECG data. Our methodology makes use of the decentralized method of federated learning, in which local models are trained on ECG datasets separately without the need for a central server. We collected data for our studies using two publicly available ECG datasets: the PTB-XL and Chapman datasets. Furthermore, we used a variety of preprocessing approaches to denoise the signals, as well as different feature extraction techniques for both datasets to discover important characteristics in the signals. In this method, the graph neural network (GNN) served as the foundation for the local models. As a result, local models are trained separately on the client side to create the global model for our proposed federated learning system. This approach attempts to create an efficient federated learning model with high accuracy, as well as a promising approach to data analysis in a variety of domains. Our experimental findings underscore the usefulness of federated learning in sensitive data analysis by showing that the suggested method achieves excellent accuracy while protecting data privacy. Future research on safe and decentralized machine learning applications in several domains is made possible by this work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
Introduction

1.1 Introduction

The World Health Organization (WHO) reports that cardiovascular disease is the top cause of chronic death globally. Every year, nearly 17.9 million people around the world die from cardiovascular diseases (CVDs) [1]. An electrocardiogram (ECG) is a commonly used tool for monitoring heart health. It's very effective for detecting and keeping track of heart diseases because it provides a lot of important information about how the heart is working [2]. The heart activity data in ECGs is a valuable tool for monitoring and diagnosing cardiovascular diseases, but manual analysis can be time-consuming and laborious, necessitating the implementation of artificial intelligence for improved efficiency [3]. A standard ECG waveform includes several key waves: the P wave, indicating atrial depolarization; the QRS complex, representing ventricular depolarization; and the T wave, corresponding to ventricular repolarization. The PR interval spans the time between the P wave and the QRS complex, while the ST segment is the flat line between the end of the QRS complex and the start of the T wave [4]. In sectors such as healthcare, finance, and telecommunications, ensuring data privacy and security is crucial due to the sensitive nature of the information involved [5]. Combining data from different medical institutions into a single dataset can make more cardiovascular information available while protecting privacy [6].

Our main goal is to develop an automated system that incorporates a federated learning strategy to tackle this challenge, driven by a strong desire to protect data privacy across various sectors. Federated learning is a specialized machine learning model that uses distributed datasets across multiple devices to build a joint model without sharing raw data, thus ensuring privacy [3]. It operates as a secure, decentralized collaborative learning technique, improving global models through local dataset contributions [7]. In this study, extensive preparation of the datasets included wavelet denoising to improve data quality, feature extraction, and a propsed Graph Neural Networks (GNNs) model for the classification. The project's objective is to develop a privacy-preserving federated learning model using ECG datasets. This involves ECG data preprocessing, traditional feature extraction research, graph construction, local model training, and creating a high-accuracy global model. The approach aims to establish a robust privacy-preserving technique that delivers superior results in ECG analysis.

The following list outlines the main contribution of this study:

1. In this study, we combined two ECG-based datasets from distinct sources. Throughout the approach, we faced numerous issues such as dataset mismatch, imbalance, and noise in the datasets.
2. To address the issues faced by the datasets, various preprocessing approaches, such as the Wavelet Transform and others, were employed.
3. To extract important properties from ECG data, three sets of custom feature extraction approaches were used: Set 1: Time domain features; Set 2: Entropy measures; and Set 3: Complexity measures.
4. The study employed a proposed FedGNN (Federated Graph Neural Network) model for efficient classification and model building.
5. Development of a decentralized federated learning model involving two clients, enabling local model training and performance evaluation using federated parameters.

1.2 Motivation

This study is highly motivated by the desire to protect data privacy in a variety of areas. The growing frequency of data breaches, combined with increased awareness of privacy concerns, has highlighted the importance of secure data management techniques. The major purpose of this project is to provide a federated learning framework employing locally learned ECG datasets. This technique tries to reduce the risks associated with centralized data storage by decentralizing data processing and retaining sensitive information in its original location. The datasets were preprocessed using advanced approaches such as the Wavelet denoising method, which reduces noise and improves signal clarity. Furthermore, feature extraction and graph generation approaches are employed to increase data quality and accuracy, guaranteeing that the models developed are both resilient and dependable. By fulfilling these objectives, the research seeks to contribute to the development of a federated learning model that safeguards privacy, clearing the path for its use in a variety of sensitive industries and increasing trust in data security methods.

1.3 Rational of the Study

The purpose of this study is to solve significant data privacy and security concerns in sensitive areas such as healthcare, finance, and telecommunications. Traditional centralized data processing systems carry considerable risks such as data breaches, logistical issues, and single points of failure. This research uses federated learning to ensure that data remains local and secure, dramatically decreasing these risks while enabling rapid and scalable model training. The use of

Graph Neural Networks (GNNs) improves the processing of complicated and structured ECG data, resulting in more accurate and trustworthy diagnostic tools. Advanced preprocessing approaches, such as Wavelet denoising and robust feature extraction algorithms, help to increase data quality and model performance. This revolutionary approach not only establishes a new standard for privacy-preserving ECG analysis, but it also provides a scalable and efficient solution that may be used to other sensitive domains. By doing so, the study hopes to improve data privacy, diagnostic accuracy, and greatly develop machine learning technologies. Furthermore, it tackles the growing worries about data breaches and privacy violations, paving the way for safer data management techniques in a variety of businesses. This research not only addresses current technological needs, but also provides the framework for future advances in safe and decentralized data analysis approaches.

1.4 Research Questions

The study focuses on numerous critical topics under the framework of "Decentralized Horizontal Federated Learning Using Graph Neural Networks for ECG Signal Analysis." These topics are central to the research, which aims to gain a complete knowledge of how federated learning and GNNs might be used in medical diagnostics. The main research questions are as follows:

- **How does decentralized horizontal federated learning improve the accuracy and privacy of electrocardiogram signal analysis?**

  Through the collaborative training of a machine learning model on local data by different institutions without sharing it, decentralized horizontal federated learning improves the accuracy of ECG signal processing. Because only the model parameters are provided and the raw data never leaves the local institutions, strict data protection requirements are followed, protecting patient privacy.
- **How do Graph Neural Networks (GNNs) Enhance Prediction in ECG Signal Analysis within a Federated Learning Framework?"**

  Graph Neural Networks (GNNs) are an important component of the federated learning architecture because they efficiently capture the intricate interactions and dependencies between various components of the ECG signals, which aid in feature extraction and prediction.

- **In what ways does the federated learning strategy address the issues of heterogeneity in data and variability in ECG signals among various institutions?**
  By combining model updates from several institutions, the federated learning strategy reduces the difficulties brought on by data heterogeneity and unpredictability and enables the global model to learn from a variety of data distributions.
- **What benefits and drawbacks come with using GNNs for decentralized horizontal federated learning in ECG signal analysis?**
  The benefits of using GNNs in decentralized horizontal federated learning for ECG signal analysis include better model generalization, more data privacy, and the capacity to use a variety of datasets without requiring data centralization. Nevertheless, there are certain drawbacks, such as the difficulty of establishing and managing the federated learning infrastructure, possible communication overheads, and the requirement for strong procedures to manage asynchronous updates and possible variations in data quality throughout schools.
- **What performance measures are employed to assess the federated learning model's effectiveness?**
  The federated learning model's effectiveness is assessed using accuracy, precision, recall, and F1 score as performance indicators.
- **What areas of study can be followed in the future to improve the use of GNNs and federated learning in medical diagnostics?**
  To further improve data security, future research paths will examine sophisticated privacy-preserving methods including safe multiparty computation and differential privacy. Predictive performance may be enhanced by looking at the usage of more complex GNN architectures and hybrid models that incorporate GNNs with other deep learning techniques.

The study is guided by a set of research questions that together attempt to shed light on the application of federated learning and GNNs for ECG signal interpretation, thereby advancing patient care and medical diagnostics.

1.5 Expected Output

The project's core deliverable is a federated learning model that is effective, privacy-preserving, and accurate. To develop the model, the team used advanced ECG dataset preparation techniques, numerous feature extraction approaches, graph construction, and local model training. The goal is to create an efficient and trustworthy federated learning paradigm. The official expectations include:

- **Development of Federated learning model:**
  The primary outcome of developing a decentralized horizontal federated learning model for ECG signal processing using graph neural networks is the creation of a reliable and accurate global model. This system incorporates advanced deep learning technologies, including Graph Neural Networks. The strategy aims for high accuracy, which will make early intervention measures more effective. This federated learning strategy improves the model's generalizability and reliability by using the unique data distributions of several healthcare facilities while protecting patient privacy. The predicted result is a strong, privacy-preserving federated learning model.

- **Evaluation and Comparative Analysis of Federated Learning Models:**
  A comprehensive evaluation and comparison of the chosen federated learning models will be done. This study will provide insights into each model's performance metrics, accuracy rates, and efficiency, allowing for a thorough knowledge of their usefulness in the context of ECG signal processing. The study's goal is to determine the most dependable and accurate strategy by assessing numerous models, including GNN. This analysis will guide the selection of the best-federated learning model, resulting in the creation of a strong and efficient global model.

- **Integration of Predictive Strategies:**
  Integrating predictive techniques into the federated learning framework will be critical for improving the accuracy of ECG signal analysis. Graph Neural Networks (GNNs) will play an important part in this study because of their capacity to simulate complicated links and interactions within ECG data. GNNs can successfully detect spatial and temporal relationships in signals, allowing for more accurate predictions. By adding GNNs, the predictive model will be able to assess the intricate patterns and linkages in ECG signals, allowing for early and accurate detection.

- **Validation of Results Through Rigorous Testing:**

  An essential part of this research will be the rigorous testing used to validate the outcomes. To guarantee the accuracy and dependability of the created federated learning models in ECG signal analysis, a rigorous testing process will be implemented. In order to evaluate the model's performance under various circumstances, this validation procedure will include cross-validation techniques, benchmarking against established datasets, and real-world scenario testing.

- **Contribution to Academic Discourse:**

  The goal of the study is to further scholarly conversations on the convergence of deep learning and machine learning

1.6 Project Management and Finance

Project Objectives:

A. To develop a decentralized horizontal federated learning platform for ECG signal analysis.

B. Provide a consistent and efficient federated learning model.

C. To provide a platform that allows cardiologists to collaborate on analysis and prediction while ensuring data privacy and security.

Project Timeline:

A. Phase 1: Project Planning and Research (2-4 weeks)

B. Phase 2: Designing Federated Learning workflow and Prototyping (4-6 weeks)

C. Phase 3: Implementing Front-End Interface (6-10 weeks)

D. Phase 4: Developing Back-End Infrastructure for Federated Learning (10-14 weeks)

E. Phase 5: Testing Federated Learning Algorithms (2-4 weeks)

F. Phase 6: Deployment of Federated Learning Platform (4-6 weeks)

G. Phase 7: Post-Launch Evaluation and Marketing (Ongoing)

Resource Planning:

A. Equipment and Tools:

- Ideal Computers for the Development of the Federated Learning Model
- Collaboration Tools (e.g., Jupyter Notebook, or project management software)
- Data storage facilities (e.g., Google Drive)

B. Software and Technologies:

- ● Machine Learning Frameworks (e.g., scikit-learn, TensorFlow)
- ● Front-End Technologies (e.g., HTML, CSS, JavaScript)
- ● Back-end technologies (e.g., Django, Node.js)
- ● Database Management System (e.g., MongoDB)
- ● Server Hosting (e.g., Google Cloud)

C. Data and Content:

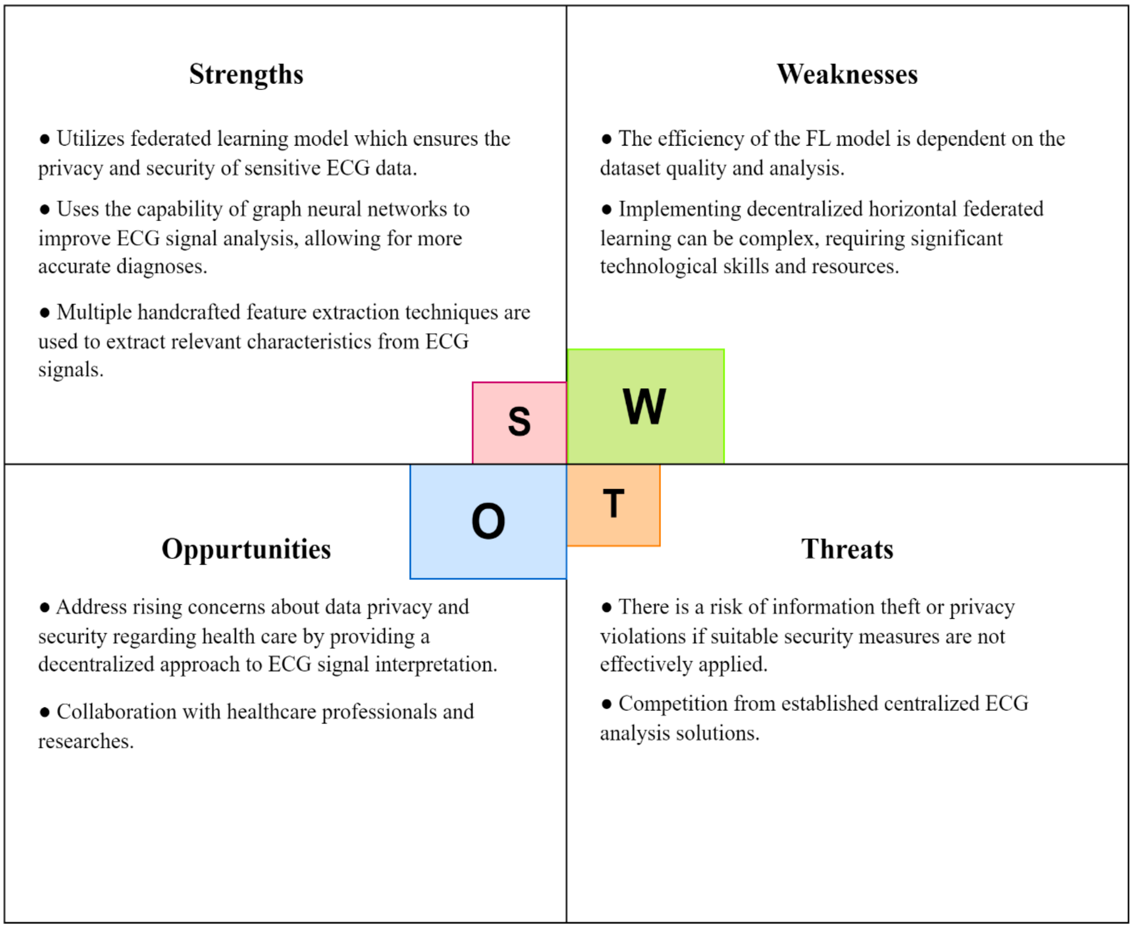
- ● ECG Signal Data: Provided by publicly available healthcare institutions.
- ● User Documentation: Documentation of experimental processes and outcomes for evaluation.

D. Training and Skill Development:

- ● Make sure the development individuals are proficient in dataset analysis, machine learning techniques, federated learning, and web development.
- ● Provide further training and seminars on specific technologies and tools as required for the development team.

Risk Management:

A SWOT analysis assesses a project by offering a full understanding of its internal and external variables, allowing for informed decision-making. Here's a SWOT analysis of the e-commerce platform for the federated learning model:



Communication Plan:

A. Stakeholder Meetings:

Purpose: Meetings with stakeholders to provide updates on progress and project feedback.

Participants: Individuals or Team members and project stakeholders.

Frequency: Twice every week, or as specified in the project plan.

B. Change Control Meetings:

Purpose: Discuss and accept any modifications to the scope or needs of the decentralized horizontal federated learning project for ECG data analysis.

Participants: Individuals or team members with supervisor and co-supervisor.

Frequency: As needed, as new change requests arise. Frequency: As needed when change requests arise.

To build our proposed framework or a federated learning model using ECG datasets into a real-life implementation, the budget depends on the datasets, development tools, UI/UX, and documentation for the project. As we are already using public datasets, the budget to implement the federated learning model implementation will be around 1,88,000 with additional maintenance costs. Table 5 shows all the estimated cost evaluations of the project.

Table 1.6.1: Cost Estimation table

SN	**Components**	**Estimated cost (BDT)**
01	Hardware	80,000
02	Visiting Stakeholders	2000
03	Software and Tools	5000
04	Documentation and Report Writing	1000
05	Website development	1,00,000
Total estimated cost		1,88,000

1.7 Report layout

The proposed report is structured with a comprehensive layout to guide readers through the research process, findings, and implications. Each chapter serves a specific purpose, contributing to the overall coherence and depth of the document.

Chapter 1: Introduction

The introduction will give an outline of the research, emphasizing the importance of creating a decentralized horizontal federated learning model for ECG signal processing utilizing Graph Neural Networks (GNNs). It will emphasize the significance of privacy-preserving approaches and the potential benefits of advanced prediction models. This part will provide the foundation for an extensive review and validation of the proposed system, highlighting its relevance and novel approach.

Chapter 2: Background

The background chapter provides a thorough review of relevant literature and prior research. This section provides a thorough overview of the theoretical foundations, techniques, and technological breakthroughs underlying decentralized horizontal federated learning, graph neural networks, and ECG signal processing. It sets the stage for the current investigation and identifies gaps in existing knowledge, providing a complete overview that emphasizes the importance of the research.

Chapter 3: Research Methodology

The research methodology chapter will go over the systematic approach used to create and evaluate the federated learning model. It will go into the data collection procedure, data preprocessing, feature extraction techniques, architecture of the proposed GNN models, and federated learning framework. This part will also go over the model framework and the training, validation, and testing processes, as well as the metrics used to evaluate model effectiveness.

Chapter 4: Experimental Result

The experimental results chapter will describe the outcomes of thorough testing and validation of the federated learning models. It will feature accuracy and efficiency comparisons of several GNN models. This section will provide a thorough analysis of the results, highlighting each model's strengths and flaws while also illustrating the proposed system's overall effectiveness in accurately diagnosing cardiac irregularities.

Chapter 5: Impact on Society

The chapter on society will address the research's broader implications for healthcare and patient outcomes. It will investigate how the use of a reliable, privacy-preserving ECG analysis tool can improve early detection, improve patient care, and lower healthcare costs. This section will also look at the ethical implications of employing federated learning in medical applications, highlighting the need for data security and patient privacy.

Chapter 6: Summary, Conclusion, Recommendation, and Implications for Future Research

The concluding chapter will present a brief summary of the research findings, emphasizing the importance of the created federated learning model and its contributions to ECG signal processing. The conclusion will summarize the important findings from the investigation, verifying the hypothesis and research objectives. There will be recommendations for practical applications and prospective system improvements, as well as suggestions for future research paths.

This planned organization ensures a logical flow of material, taking the reader through the study process from introduction to broader implications and suggestions. It provides an in-depth understanding of the research process and its possible impact on both the academic and practical components of the federated learning model with GNN.

CHAPTER 2

Background

2.1 Preliminaries

To facilitate a clear understanding of the key concepts and methodologies used in this research on "Decentralized Horizontal Federated Learning Using Graph Neural Networks for ECG Signal Analysis," a foundation must be established through preliminary work and terminological clarification.

Federated learning, a key idea in privacy-preserving machine learning, serves as the study's foundation, allowing the model to use decentralized data from many institutions while maintaining patient privacy.

Graph Neural Networks (GNNs), which are necessary for dealing with complicated and structured data, are used to capture detailed relationships within ECG signals. These networks use nodes and edges to represent data points and their relationships, allowing for a more detailed examination of cardiac abnormalities. The federated learning strategy, when paired with GNNs, promises to improve the accuracy and reliability of ECG signal analysis while protecting data privacy and exploiting varied datasets.

Throughout this study, terms like "federated learning," "graph neural networks," and "ECG signal analysis" are used interchangeably to investigate the synergies between these advanced approaches. Developing a shared knowledge of these preliminary concepts is critical for readers to delve into the complexities of the study and realize the nuances of the strategies used to improve the global model.

2.2 Related Works

Many studies have been published in recent years that use deep learning and machine learning methodologies to create federated learning frameworks using ECG data. In this section [8-17], we reviewed multiple studies based on the Federated Learning Framework using ECG Signals.

Sakib et al. [8] used an asynchronous and synchronous federated learning model on four publicly available arrhythmia detection datasets. They used the MIT-BIH Supraventricular Arrhythmia database for learning/training purposes and the adopted techniques used in the research were extensively tested under three other datasets from the Physionet repositories - MIT-BIH Arrhythmia database, INCART 12-lead Arrhythmia database and the Sudden Cardiac Holter database. The paper introduces an asynchronously updating Federated learning architecture for mobile and deployable UENs to build a decentralized arrhythmia detection system using Federated learning. They experimented with both the model: Sync-FL and Async-FL on three Arrhythmia datasets from the Physionet repositories and experimented results show that the Async-FL model achieved higher accuracy than the synchronous federated learning model of 89%. However, they didn't use any feature extraction methods that could led to better accuracy in this work.

In another study [9], researchers used various ways to fuse data from 12-lead ECG signals to diagnose heart abnormalities. The dataset used in this study was obtained from Chapman University and Shaoxing patients' Hospital and includes eight forms of arrhythmia recorded from 10,646 patients with a frequency greater than 500 Hz. This study suggests combining time-frequency transform and deep convolutional neural network (CNN) in two phases to fuse the leads of ECG signals and detect various cardiac diseases. The proposed technique achieves a mean accuracy of 97.60% on the Chapman ECG dataset. Furthermore, they should have limited the number of features extracted by the algorithms.

Chen et al. [10] used a deep learning model trained on the publicly accessible Chapman 12-lead ECG dataset. The dataset comprises ECG recordings from over 10,000 people. This study applied a deep convolutional neural network on the Chapman dataset to classify AF or atrial fibrillation in critically unwell patients. The experiment results reveal that computational detection of AF is currently feasible and successful, with 84% sensitivity, 89% specificity, 55% PPV, and 97% NPV. Their future study should focus on the classification of NOAF (New-onset atrial fibrillation), which is challenging for diagnosing computationally.

On the other hand, the paper [11] presented novel insights into CHF detection using two publicly available databases at PhysioBank - NSR-RR-interval and CHF-RR-interval database. The NSR-RR database has 54 ECG recordings of subjects with NSRs of 30 men and 24 women aged from 29 to 76. Conversely, beat annotations from 29 CHF patients—eight men, two women, and 19 additional individuals of unknown gender—ranging in age from 34 to 79—are included in the CHF-RR database. All of these patients had a diagnosis of CHF. This work presented a novel deep learning model—a trimmed version of 1-D UNet++ to classify patients with CHF and healthy people. Discriminative features were also derived from the CHF classification by the suggested end-to-end model. The suggested approach produced an accuracy of 89.84% and 87.54% in centralized learning and FL, respectively, as they expanded the CHF detection from centralized learning to federated learning. However, they mentioned the imbalanced division of the dataset which impacts the training process and the accuracy.

Chopannejad, Roshanpoor, & Sadoughi [12] suggested a deep learning strategy for detecting cardiac arrhythmias on the Chapman-Shaoxing dataset. Chapman University and Shaoxing People's Hospital created the ECG dataset, which includes 12 lead recordings from seven different arrhythmia classes. This study describes a hybrid CNN-BILSTM-BiGRU method that uses FL cross entropy and a multi-head self-attention mechanism to tackle the difficult task of identifying diverse cardiac rhythms. The SMOTE-Tomek was also used to address the data imbalance issue in detection. Overall, the results show that the proposed model classified arrhythmias with an accuracy of 98.57% for lead II and 98.34% for lead avF. In the future, they can apply their suggested approach to various datasets and use techniques to improve data privacy.

In a different study [13], the researchers presented a strategy for training ECG classifiers on three datasets: the PTB-XL dataset, which comprises 21799 clinical 12-lead ECGs, the MIT BIH Arrhythmia database, which contains 48 long-term Holter monitor recordings with two leads, and the CinC 2017 dataset, which comprises 8528 short and single-lead ECGs. To replicate other hospitals with various sizes and distinct datasets, these three datasets were divided. In this work, the widely held belief about machine learning is abandoned in favor of training a model that can identify a range of heart diseases. Additionally, a federated learning architecture was employed in this work to safeguard patient privacy. They have used three algorithms to predict the federated classifier: FedAvg, Genetic CFL, and FedCHT on three datasets, partitioned by encoding dimensions. The result shows that, for an encoding dimension of 1000, the federated classifier achieved an accuracy of 73.0%. However, they didn't use any feature extraction techniques to extract meaningful features from the ECG signals that could led to better accuracy.

Raza, Tran, Koehl, & Li [14] also used the MIT BIH Arrhythmia database to classify different arrhythmias. The dataset contains 48 half-hour excerpts of two-channel ambulatory ECG recordings. This widely used dataset has 109,446 samples. In this paper, they proposed a privacy-preserving end-to-end framework in a federated learning setting for ECG signal classification using Explainable AI and deep convolutional neural networks (CNN). Federated learning helps in terms of the privacy of the datasets and the proposed framework classifies different arrhythmias using an autoencoder and a classifier, both based on CNN. This work also used Grad CAM which is an XAI or Explainable AI method to solve the challenges of the black box model. The proposed model achieved an accuracy of 98.9% for Arrhythmia detection using clean data. Nevertheless, the data is collected from multiple devices, not from a single server which could be a better option. The paper [15] provides a novel approach for the detection of cardiac abnormalities based on federated learning. The paper utilized two ECG signal datasets: the MIT BIH Arrhythmia dataset and the PTB database. In this work, the VGG19 model is used to evaluate the proposed method and a sampling technique to better aggregate the parameters to the server. The federated learning model solved the challenges of data privacy. In this way, the proposed approach achieved an accuracy of 87.98% with the robust VGG19 architecture. However, they could use a better aggregation strategy in fewer communication rounds.

Yoo et al. [16] introduced a privacy-preserving framework to analyze the severity of a major depressive disorder (MDD). They have used the HRV dataset that consists of 479 instances from 100 clients that also involved multi-dimensional features extracted from the time and frequency domain. This paper proposed a Personalized Federated Cluster Model (PFCM) to accurately analyze the MDD using the HRV dataset. The proposed method with the CNN based FL analysis method with 2 classes achieved an accuracy of 90.70% which outperformed other machine learning models. The main limitation of this study is the lack of data, which could lead to better accuracy.

Rehman et al. [17] combined a federated learning model and advanced encryption to provide a secure and privacy-preserving healthcare 5.0 system. For the disease prediction analysis, they have utilized the publicly available dataset- Parkinson's disease dataset, and for intrusion detection, they have used the NSL-KDD dataset. For safety and privacy, they have employed the federated learning model and Blockchain technology and combined both of the techniques, the proposed method: the RTS DELM model achieved an accuracy of 93.22% for Parkinson's disease prediction and 96.18% for intrusion detection. However, in the future, they can use multiple models to get more insight into the proposed method.

2.3 Comparative Analysis and Summary

The comparative evaluation undertaken in this study on "Decentralized Horizontal Federated Learning Using Graph Neural Networks for ECG Signal Analysis" is a critical component, shedding light on the subtle differences and synergies between different approaches. The paper thoroughly compares the performance of federated learning and Graph Neural Networks (GNNs) in ECG signal interpretation, taking into account both classic and advanced federated learning approaches. Through rigorous experimentation and analysis of experimental results, the study reveals insights into the strengths and limitations of each strategy, offering a comprehensive understanding of their different contributions to the diagnostic process.

The in-depth look at federated learning and GNN techniques allows for a more nuanced understanding of their roles in improving the accuracy and efficiency of ECG signal analysis. By comparing the results of various approaches, the study adds vital knowledge to the area, assisting practitioners and researchers in picking the most effective strategies for their individual situations. The study does more than just highlight inequalities; it also underscores the possible complementarity between federated learning and GNNs, which contributes to our knowledge of the decentralized and efficient processing of ECG signals.

In conclusion, the comparative analysis conducted in this study provides a critical lens through which the efficacy of federated learning and GNNs in ECG signal interpretation is evaluated. It provides a multidimensional perspective, including both conventional and current methodologies, and emphasizes the importance of a holistic approach in medical diagnostics. The findings not only increase our understanding of ECG signal analysis, but they also provide useful insights for improving existing diagnostic approaches. This comparative analysis serves as the framework for the ensuing chapters, laying the groundwork for the discussion of social impact, recommendations, and future research implications.

Table 1 shows the information about all the related works we have reviewed and we addressed the methods and accuracy of the papers.

Table 2.3.1. Comparative analysis with previous work

Author name	**Dataset**	**Used Algorithm**	**Best Accuracy with Algorithm**
Sakib et al. [8]	1. MIT-BIH Supraventricular Arrhythmia database (DS1) 2. MIT-BIH Arrhythmia database (DS2) 3. INCART 12-lead Arrhythmia database. (DS3) 4. Sudden Cardiac Holter database (DS4)	Async-FL	Async-FL=89%(DS2), 89%(DS3), 76%(DS4)
Meqdad, Abdali-Mohammadi, & Kadry [9]	Chapman dataset	CNN	97.60%
Chen et al. [10]	Chapman dataset	deep convolutional neural network	84% sensitivity, 89% specificity, 55% PPV, and 97% NPV
Zou et al. [11]	1. NSR-RR-interval database. 2. CHF-RR-interval database.	1-D UNet++	89.84% (centralized learning), 87.54% (FL-decentralized learning) (merged dataset)
Chopannejad, Roshanpoor, & Sadoughi [12]	Chapman dataset	CNN-BILSTM-BiGRU	98.57% for lead II and 98.34% for lead avF
Chorney & Wang. [13]	1. PTB-XL dataset	FedAvg, Genetic CFL, and FedCHT	FedCHT= 73.0% (merged

	2. MIT BIH Arrhythmia database 3. CinC 2017		dataset)
Raza, Tran, Koehl, & Li [14]	MIT BIH Arrhythmia database	CNN	CNN = 98.9%
Meqdad, Hussein, Husain, & Jawad. [15]	ECG heartbeat categorization dataset (consisting of MIT BIH Arrhythmia dataset, The PTB database and other two datasets)	VGG19	87.98%
Yoo et al. [16]	HRV dataset	Personalized Federated Cluster Model (PFCM)	90.70%
Rehman et al. [17]	1. Parkinson's disease dataset 2. NSL-KDD dataset	RTS-DELM	93.22% (Parkinson’s disease dataset) 96.18% (NSL-KDD dataset)

2.4 Scope of the Problem

The problem's depth includes several important obstacles and constraints inherent in existing ECG signal analysis techniques. Privacy and data security are primary considerations. Traditional centralized methods demand the collection of sensitive patient information, which raises significant concerns about data breaches and illegal access. Addressing these privacy concerns while facilitating collaborative learning among many healthcare facilities is a critical component of the challenge.

Another major issue is data heterogeneity. ECG data can vary significantly between populations, medical situations, and recording locations. Creating a model that generalizes effectively across these various datasets is critical to guarantee accurate and reliable cardiac abnormality identification. Furthermore, the intricacy of ECG signals necessitates advanced analytic algorithms capable of detecting subtle patterns and linkages in the data.

The current absence of good predictive tools for early diagnosis exacerbates the problem. Many present models struggle to provide timely and accurate predictions, which limits their usefulness in clinical settings. This study intends to address these issues by incorporating sophisticated technologies like Graph Neural Networks (GNNs) into a federated learning framework, thereby improving the accuracy, privacy, and generalizability of ECG signal processing.

CHAPTER 3

Research Methodology

3.1 Research Subject and Instrumentation

Research Subject:

The study "Decentralized Horizontal Federated Learning Using Graph Neural Networks for ECG Signal Analysis" focuses on advanced federated learning methodologies used to analyze electrocardiography (ECG) signals. The study overcomes the limitations of centralized data collecting in medical research by implementing a decentralized federated learning system. Graph Neural Networks (GNNs), a type of neural network that can model complicated relationships in graph-structured data, are used to evaluate and learn from interconnected ECG signals. This method improves the ability to detect subtle patterns and anomalies that indicate cardiac problems across a wide range of patient demographics.

The study investigates the concept of horizontal federated learning, in which local ECG datasets from various medical institutions remain on-premises while preserving data privacy and security. The federated model collects insights from these local datasets while maintaining individual data ownership, allowing for collaborative model training without disclosing sensitive patient information.

The study's goal is to promote medical diagnoses by improving the accuracy, scalability, and privacy of ECG signal analysis. The findings aim not only to advance academic understanding, but also to stimulate practical solutions that could improve diagnostic skills and patient care in cardiology and others.

Instrumentation:

The experimental setup for this study involves operating research on a desktop computer with specific hardware and software specifications. The system features an Intel(R) Core(TM) i5-10400 CPU operating at 2.80GHz with 12 GB of RAM. To enhance computational capabilities, an NVIDIA GeForce GTX 1660 GPU is utilized. All experiments are executed using the Jupyter Notebook environment. This setup is chosen to ensure efficient processing and reliable results of our proposed work.

This project makes use of Python programming language, Python libraries, Pytorch geometric, NetworkX, and Jupyter notebook. Furthermore, many study papers, journals, and online resources aided in decision making and subsequent analysis. Furthermore, to solve the issues, we worked with domain experts and researchers. Overall, these resources aided the project's success over multiple phases.

3.2 Main Workflow

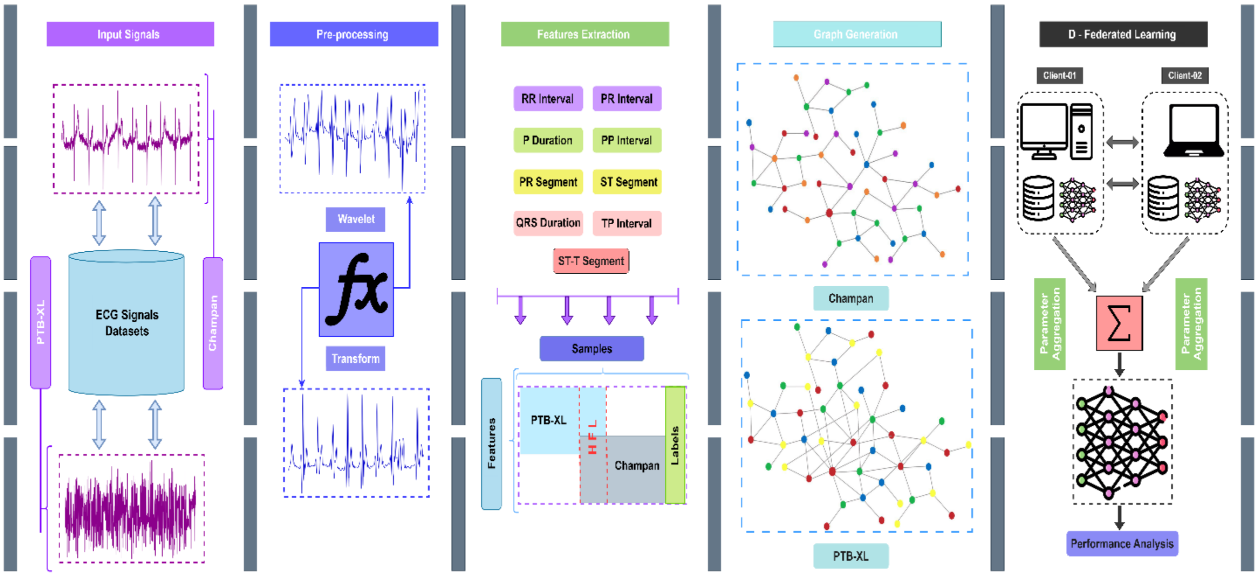


Fig 3.2.1: Main workflow of the proposed work.

Figure 3.2.1 depicts the key steps in developing a privacy-preserving horizontal federated learning model utilizing the ECG datasets. This work is divided into five sections. Section 1 describes the datasets: the PTB-XL and Chapman datasets which are employed to train the local model. Section 2 focuses on preprocessing ECG signals with the Wavelet Transform, which denoises the signals and improves their quality. In phase 3, a variety of feature extraction approaches were applied to extract significant characteristics from both datasets. In addition, phase 4 involves graph construction utilizing the graph neural network (GNN), which serves as the foundation for the local and global models of the Federated learning model for both datasets. Finally, phase 5 involves the creation of a decentralized yet horizontal federated learning model using two clients, from which the local model is trained and the parameter is passed to measure the FL model's performance. Figure 1 is organized into sections that provide a detailed yet understandable overview of this research.

3.3 Dataset Description

In this work, we utilized two large publicly available ECG signal-based datasets: PTB-XL dataset [18], and Chapman dataset [19].

The PTB-XL dataset comprises 21,799 clinical 12-lead ECGs from 18,869 individuals, each lasting ten seconds. The dataset includes 52% males and 48% females, with ages ranging from 0 to 95 years. The signals are captured at 500 Hz and 100 Hz with a 16-bit resolution. The dataset is categorized into five diagnostic categories and 23 diagnostic subcategories, including (CD) for Conduction Disturbance, (HYP) for Hypertrophy, (MI) for Myocardial Infarction, (STTC) for ST/T Change, and (NORM) indicating Normal ECG. Table 1 provides details on the number of ECG signals corresponding to each class within the PTB-XL dataset, along with a description of each category.

The Chapman dataset has 10,646 patient 12-lead ECGs, comprising 5956 men and 4690 females, with a sampling rate of 500 Hz, containing 11 common rhythms and 67 circumstances identified by qualified physicians and integrated into four categories (SB, AFIB, GSVT, and SR). SB only included sinus bradycardia (SB). In contrast, AFIB included atrial fibrillation (AFIB) and atrial flutter (AFL), GSVT included supraventricular tachycardia (SVT), atrial tachycardia (AT), atrioventricular node reentrant tachycardia (AVNRT), atrioventricular reentrant tachycardia (AVRT), and sinus rhythm (SR) and sinus irregularity (SI). Table 2 presents the number of electrocardiogram signals in the ChapmanECG dataset.

Table 3.3.1**:** Number of ECG signals corresponding to the 5 superclasses of the PTB-XL dataset.

Class name	**Number of records**
CD	1,708
HYP	535
MI	2,532
STTC	2,400
NORM	9,069

Table 3.3.2: Number of ECG signals corresponding to the 4 groups of the Chapman dataset.

Class name	**Merged Class name**	**Number of records**
AFIB, AF	AFIB	3,889
SVT, AT, SAAWR, ST, AVNRT, AVRT	GSVT	2,307
SB	SB	2,225
SR, SI	SR	2,225

3.4 Preprocessing of ECG signals

ECG data were contaminated by various noise sources, including power line interference, electrode contact noise, motion artifacts, muscle contraction, baseline wandering, and random noise, hindering statistical analysis. To address this, we implemented a sequential noise reduction approach for preprocessing the raw ECG data. We checked the length of each ECG signal, measured by the number of samples, to ensure uniformity for neural network classification. Signals were organized by sample number, and we selected the group with the most signals of identical length. Signals with fewer or more samples than those in the selected group were eliminated to maintain consistency in signal length for the classification system.

3.4.1 ECG Signals Noise Removing Using a Discrete Wavelet Transform

The wavelet transform is a versatile mathematical tool for signal processing, primarily used for time-scale signal analysis, signal decomposition, and signal compression.

Any digital signal is affected by various types of noise, which distorts its features [20]. To accurately isolate the signal's characteristics, it is essential to remove this noise. An ECG signal corrupted by noise can be represented as follows:

$$W(t) = S(t) + N(t)$$

where $W(t)$ is the ECG signal, $S(t)$ is the ECG signal without noise distortion, $N(t)$ is the noise on the ECG signal [20].

In this study, we utilized the discrete wavelet transform (DWT) from the symlet family, specifically a Daubechies wavelet known for its minimal asymmetry and compact support, to denoise ECG signals [20]. The DWT is grounded in multi-resolution analysis (MRA) improves spatial and spectral localization in image signals by producing a non-redundant representation, surpassing other multi-scale methods like Gaussian and Laplacian pyramids [21]. Conceptually, it decomposes a signal into independent frequency channels aligned spatially.

Using wavelet transform to clean noise from an ECG signal involves three stages. First, obtain noisy wavelet coefficients by performing a discrete wavelet transform (DWT) on the noisy signal. Second, select a thresholding method to determine the presence of noise in the signal. Finally, apply an inverse wavelet transform to reconstruct a cleaned signal [20]. The ECG signal's DWT is:

$$DWT(a,b) \ = \frac{1}{\sqrt{2}}\sum_{j=0}^{N} w_j \int_{j}^{i+1} \Psi\left(\frac{t-b}{a}\right) dt$$

Where N is the number of samples on the ECG signal, W is the ECG signal distorted by noise, Ψ is a symlet, and a and b variables can take the values $a = \ 1 \ldots N$, $b = \ 1 \ldots N-1$.

In order to obtain the DWT, a low-pass analysis filter with g impulse response and a high-pass analysis filter with an h impulse response is used. The filtering process results in the approximation and detail coefficients [22]

$$y_{low}(t) = \sum_{k=-\infty}^{\infty} W(t)g(2t-k)$$

$$y_{high}(t) = \sum_{k=-\infty}^{\infty} W(t)h(2t-k)$$

After cleaning up the ECG signal, the next step is to select a threshold function and apply thresholding to distinguish between signal and noise. If the wavelet coefficient exceeds the threshold value, it is considered part of the signal; otherwise, it is classified as noise. To determine the threshold limit of the ECG signal, the minimax threshold T was utilized [23],

$$T \leq \sqrt{2lnN}, T^2 = 2ln(N+1) - 4ln(In(N+1)) - ln2\pi$$

The wavelet coefficients are transformed using thresholding, employing threshold functions. There are two predominant techniques, namely hard-thresholding and soft-thresholding that are frequently employed [22]. The soft-threshold function was used in this study to identify noise in the ECG signal and where c is the value of the wavelet coefficient [23].

$$0 \leq max\left(1 - \frac{T}{|c|}, 0\right) \leq 1$$

For the reverse DWT that can be used to clean an ECG signal [22]:

$$S(t) = \sum_{k=-\infty}^{\infty} \tilde{g}_k y_{low_k}(t) + \sum_{m=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \tilde{h}_{m_k} y_{high_k}(t)$$

where $S(t)$ denotes the ECG signal without noise distortion, $\tilde{g}_k$ and $\tilde{h}_{m_k}$ represent the approximating and detailing coefficients, respectively, after undergoing processing with the threshold function.

3.5 Implementation Requirements

The research on "Decentralized Horizontal Federated Learning Using Graph Neural Networks for ECG Signal Analysis" has specified requirements that must be met in order for the study to be completed successfully. These implementation requirements include hardware and software components, as well as concerns for data collecting and model evaluation.

Hardware Requirements:

- **Resources for high-performance computing:** Access to computational resources with adequate computing power and memory for handling the complex tasks required for federated learning and ECG signal analysis.

- **GPU:** A GPU is necessary to accelerate the training of deep neural networks and Graph Neural Networks (GNNs), thus significantly reducing model development time.

Software Requirements:

- **Deep learning tools:** TensorFlow and PyTorch are two well-known deep learning frameworks that are used in the building, training, and assessment of federated learning models.
- **Programming language:** We employed the Python programming language for our project because it offers a flexible and popular platform for data analysis and the implementation of deep learning algorithms.
- **Signal processing libraries:** The utilization of signal processing libraries, such as the Discrete Wavelet Transform, for ECG signal preprocessing.

Data Collection:

- **Diverse and labeled dataset:** An large and clearly annotated dataset of ECG signals that is diverse and well-labeled, with a broad variety of representative samples to guarantee reliable and broadly applicable model training and validation.

Model Training and Evaluation:

- **Transfer learning models:** Implementation of transfer learning architectures, leveraging pre-trained models such as those from ImageNet, to capitalize on existing knowledge for pneumonia detection.
- **Metrics for evaluation:** Definition and utilization of appropriate evaluation metrics, such as accuracy, precision, recall, and F1 score, to assess the performance of the models in both detection and segmentation tasks.

Ethical Considerations:

- **Data Confidentiality and patient privacy:** Maintaining the confidentiality and security of patient data throughout the research process by closely following legal and ethical requirements.
- **Informed consent:** Obtaining participants' informed consent to make sure they understand all the possible advantages and hazards associated with the study, as well as how their ECG data will be handled.

Documentation and Reproducibility:

- **Code documentation:** Giving future researchers and collaborators thorough documentation of the source will guarantee clarity, make it easier for them to grasp, and enable reproducibility.

By achieving these implementation requirements, the study can proceed methodically and thoroughly, ensuring the reliability of the findings and propelling advancement in the field of federated learning with Graph Neural Networks for ECG signal processing.

CHAPTER 4

Implementation

4.1 Overview

This implementation introduces an innovative technique to ECG signal processing based on Decentralized Horizontal Federated Learning and Graph Neural Networks (GNN). The proposed methodology uses a federated learning framework to protect data privacy while improving accuracy through the use of large, distributed ECG datasets. The GNN, when paired with feature extraction approaches, handles the complexity and conjunctive nature of ECG data, resulting in efficient clustering and accurate classification. The model's performance is carefully examined using metrics such as accuracy, precision, recall, and F1-score, indicating its ability to considerably improve cardiac diagnostics and contribute to advances in healthcare technology.

4.2 Feature extraction

ECG signal data are complicated and offer a vast amount of information about the heart's electrical activity in individuals. Numerous handcrafted feature extraction techniques have been used in this research to identify and extract meaningful key characteristics from the ECG signals. This research used three sets of handcrafted features on ECG signals: Set-1-Time-domain (13): Mean PR interval [24], Mean PP interval [24], Mean TP interval, Mean P duration, Mean QRS duration [25], Mean QT duration [24], Mean PR segment [27], Mean ST segment [28], Mean R wave amplitude [26], Mean T wave amplitude, Mean P wave amplitude, Mean U-wave Count, Mean RR interval [1], Set-2: Entropy Measures (4): Permutation Entropy [29], Shannon Entropy [29], Wavelet Entropy [29], Kolmogorov Complexity [29], and Set-3: Complexity Measures (1): Largest Lyapunov Exponent [29]. We have employed 18 features in this study, and Table 4 demonstrates the feature description of each set of handcrafted features.

Table 4.2.1: Three sets of handcrafted features with feature list and description.

Category	Feature name	Feature description
Set-1: Time-domain features (13)	Mean PR Interval (ms)	The average duration of PR intervals across numerous heartbeats in an ECG recording is measured in milliseconds. The PR interval is the period between the P wave and the QRS complex.
	Mean PP Interval (ms)	The average duration of PP intervals across numerous

		heartbeats in an ECG recording is expressed in milliseconds (ms). PP interval is the time duration between two successive P waves present in the ECG signal.
	Mean TP Interval (ms)	Average time interval between the subsequent P-wave and the T-wave in miliseconds.
	Mean P Duration (ms)	P-wave duration on average, measured in milliseconds.
	Mean QRS Duration (ms)	The QRS complex's average duration, measured in milliseconds. The QRS-complex consists of the Q-, R-, and S-waves and corresponds to the ventricular systole.
	Mean QT Duration (ms)	The QT duration, also known as the QT interval refers to the time interval between the Q wave and the T wave of the ECG signal, measured in milliseconds.
	Mean PR Segment	The PR segment's average duration. PR segment indicates a delay between atrial and ventricular activation and is found to be associated with QT changes.
	Mean ST Segment	The average duration of the ST section. ST segment can be defined by ending point of S wave and start point of T wave.
	Mean R-wave Amplitude	It is measured as the vertical distance from the peak of the R wave to the baseline (as defined by two consecutive PR intervals).
	Mean T-wave Amplitude	It is measured as the vertical distance between the baseline and the peak of the P wave in an ECG signal.
	Mean P-wave Amplitude	It is defined as the vertical distance between the baseline and the peak of the T wave.
	Mean U-Wave Count	The ECG signal's average count of U-waves.
	Mean RR Interval (ms)	The mean RR interval is the average length of RR intervals over several heartbeats in an ECG recording. It is measured in milliseconds, as are the PR and PP intervals. The RR interval is the time duration of the two successive R waves present in the ECG signal.
Set-2: Entropy measures (4)	Permutation Entropy	This parameter estimates the ECG signal complexity by calculating the couplings between the signals.
	Shannon Entropy	It is one of the spectral entropies and it is used to quantify the

		spectral complexity of the ECG signal.
	Wavelet Entropy	It is the measure of relative energies in the different frequency bands of the ECG signals and is used to determine the degree of disorder.
	Kolmogorov Complexity	It is used to describe the complexity and irregularity present in ECG signal
Set-3: Complexity measures (1)	Largest Lyapunov Exponent	This parameter is used to indicate the chaos present in the ECG signal and Ex LLE is positive, if chaos is present in the signal.

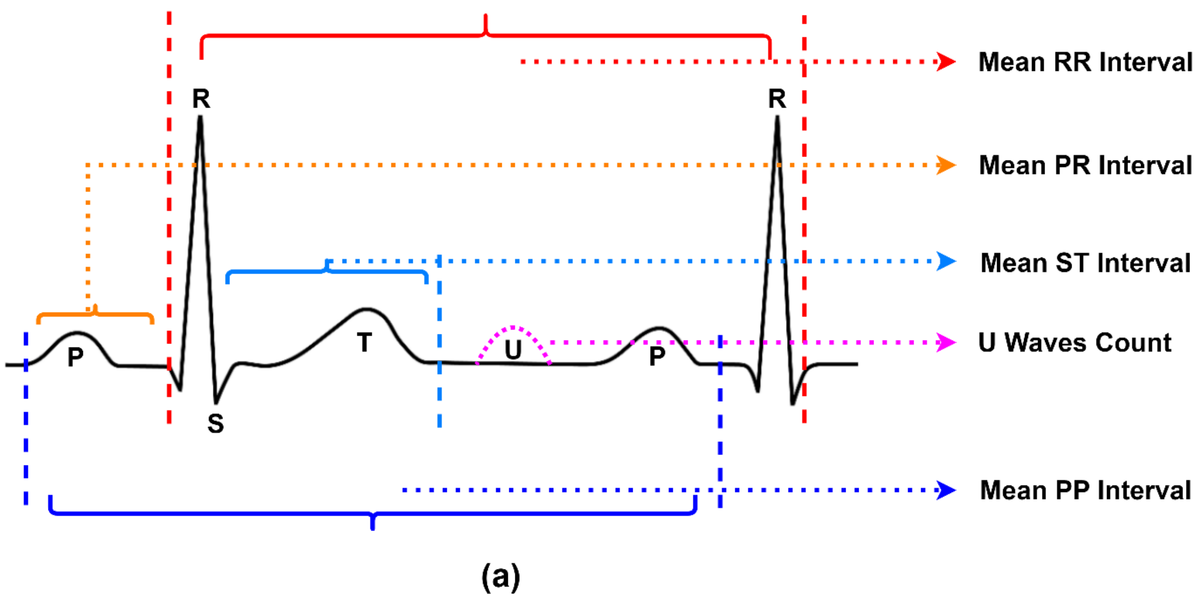


Fig 4.2.1: Illustrates Mean RR Interval, Mean PR Interval, Mean ST Interval, U Waves Count and Mean PP Interval features from ECG signal.

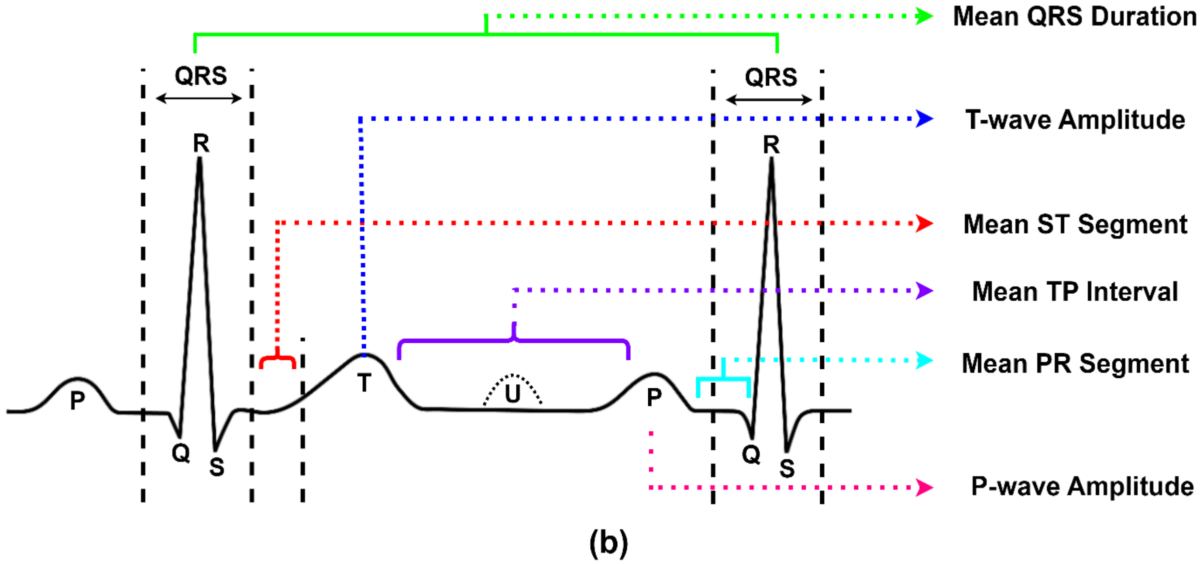


Fig 4.2.2: Illustrates Mean QRS Duration, T-Wave Amplitude, Mean ST Segment, Mean TP Interval, Mean PR Segment and P-Wave Amplitude features from ECG signal.

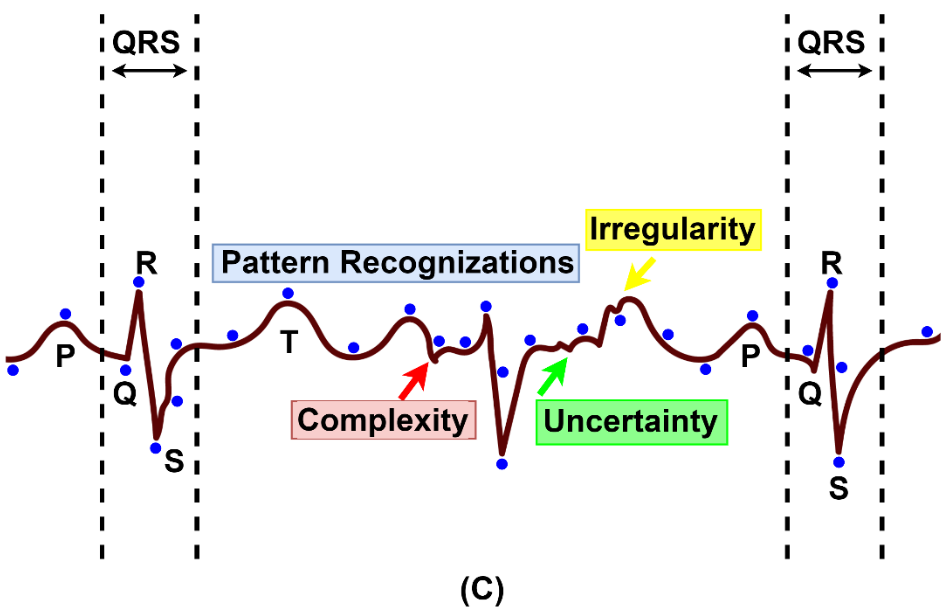


Fig 4.2.3: Illustrates the features, Permutation Entropy, Shannon Entropy, Wavelet Entropy, Kolmogorov Complexity and Largest Lyapunov Exponent extracted from these stances.

4.3 Horizontal-federated learning

In the context of horizontal-federated learning, "horizontal" denotes a method where each participant's training samples are segregated horizontally from the collective dataset. This approach aggregates these horizontally divided samples to amplify the available training data for the model [30]. It optimizes training by expanding the dataset through horizontal segmentation among participants and combining divided samples to enrich the model's training process [31]. Therefore, horizontal-federated learning can be summarized by,

$$T_a = T_b, L_a = L_b, I_a \neq I_b, \forall D_a, D_b \; a = b \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(1)}$$

We utilized horizontal federated learning to enhance our signal feature analysis. After extracting relevant features from our two different datasets, we employed horizontal federated learning to effectively combine similar feature attributes from both datasets. By horizontally integrating the features, we can enhance the overall quality and comprehensiveness of our analysis, leading to more accurate and robust conclusions from the combined ECG signal data.

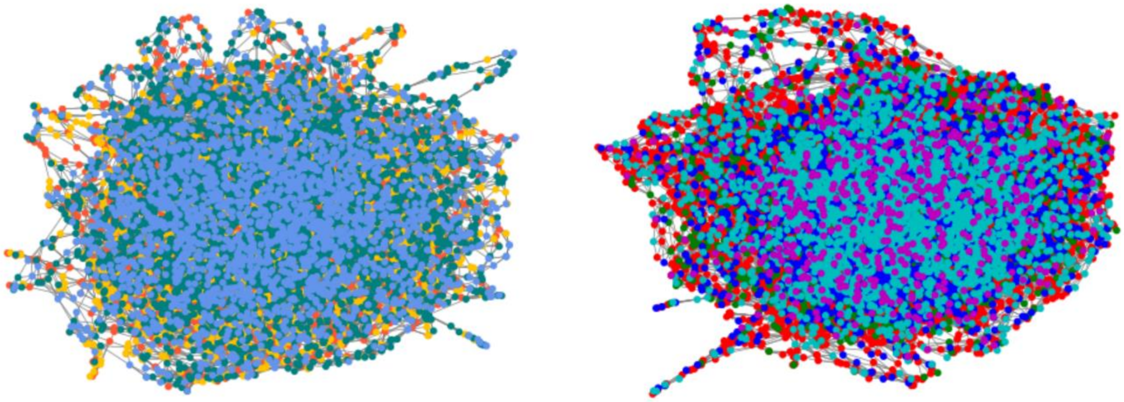
4.4 Graph generation using KNN

The K-Nearest Neighbors (KNN) technique constructs graphs from tabular data by connecting each node to its k-nearest neighbors based on the Euclidean distance. The graph $G_{nn} = V, E$ is defined with vertices V representing data points, edges E linking each vertex v_i to its nearest neighbor v_j, and edge weights $w(i, j)$ as the Euclidean distance $d(v_i, v_j)$ [32]. A custom similarity measure $S(x_i, x_j)$ quantifies feature vector similarity [33]. Edges are weighted inversely to this measure: $w(i, j) = 1/S(x_i, x_j)$ [34]. This KNN graph generation technique is illustrated in Algorithm 1 and fig 3.7.1 illustrates the graphs that generated from the Chapman and PTB-XL datasets.

Algorithm 1 Generate Graph from Tabular Data

Require: Data matrix X, Number of nearest neighbors K
Ensure: Graph G
Initialize $K = N$
Initialize knn with K neighbors using *'distance_metric'*
Fit knn to X
Compute nearest neighbor's indices as $knn\_indices$
Initialize empty graph G
for each data point i ***do***
    Add node i to G
end for
for each data point i and its neighbors in $knn\_indices$ do
    for each neighbor $neighbor$ **do**
    if $i \neq neighbor$ **then**

```
        Add edge between nodes i and neighbor to G
    end if
  end for
end for
```



i) Chapman   ii) PTB-XL

Fig 4.4.1: Graph generation of i) Chapman and ii) PTB-XL datasets

4.5 Decentralized Federated Learning (DFL)

FL is a promising solution to the challenge of distributing computation across several clients (machines) while maintaining data privacy and reducing energy consumption. In FL, a central hub coordinates the learning process, allowing multiple clients to contribute to model training without sharing raw data [35]. DFL allows clients to directly interact without a server to train global models. This method protects privacy by reducing data and model exposure to a central entity. DFL reduces network communication and computation costs and protects against single-point failures by enabling direct client interactions [35]. Our customized FL framework enhances this approach by utilizing a decentralized structure where clients use the same FedGNN architecture. This method distributes the training workload among clients, ensuring robust and personalized model development while preserving patient privacy and optimizing computational resources [36].

In each iteration, clients receive the global model clients (Client-1 and Client-2) and a global model W_i. Each client then either trains using the global model or independently on its local data. After a set number of iterations, the server collects the clients' weights ($W_{i,1}, W_{i,2}, \ldots\ldots W_{i,n}$), averages them, and sends the updated weights W_{i+1} back to all clients. This method, which involves sending only the weights rather than raw data, significantly reduces energy consumption in communication, crucial for battery-powered devices [36].

After taking features from the Chapman and PTB-XL ECG datasets for this project, we use horizontal federated learning to combine the features that are similar between them. Subsequently, DFL is applied, facilitating collaborative model refinement between the Chapman and PTB-XL clients through peer-to-peer communication. Our proposed FedGNN model makes use of the graph structure of the data to increase classification accuracy. After decentralized training, the knowledge from the local models is put together into a global model that uses insights from both datasets to do a better job of classifying data than using just one dataset.

4.6 Proposed GNN Architecture

A graph is a data structure that represents a collection of objects, called nodes, and the connections between them, known as edges. Recently, there has been increasing interest in applying machine learning techniques to graph analysis because of the substantial expressive power of graphs [37]. An embedding of a node in a lower-dimensional space is a way to represent a graph node. At the first step (k=1), the embedding of each node only includes information from its immediate neighbors, which are the nodes directly connected to it. As we increase the number of steps (k = 2, k = 3, etc.), the embeddings learn more about nodes that are farther away in the graph [38]. Subsequently, performing tasks like node classification and graph-level predictions, we gather details about the structure of the graph and the features of its nodes in steps. Each node in the network keeps a feature vector, which it updates by combining vectors from nearby nodes. This merging process creates new vectors, making the node's representation more complete.

In equations 2 and 3, m stands for the message, v for the number of nodes, and k for the number of iterations [38].

$$m_u^{(k+1)} = CONCAT_k\left(m_u^{(k)}, MEAN_k\left(\{m_v^{(k)}, M_v \in N(u)\}\right)\right) \ldots\ldots\ldots\ldots (2)$$

$$m_u^{(k+1)} = CONCAT_k\left(m_u^{(k)}, m_{N(u)}^{(k)}\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

Where $MEAN$ represent aggregates information, while $CONCAT$ updates function. The message, denoted as $m_{N(u)}^{(k)}$, is gathered from the neighborhood $N(u)$ of node in the graph. Superscripts are used to denote embeddings and functions across various iterations of message-passing. The

$MEAN$ function takes as input the embeddings of nodes within the neighborhood $N(u)$ at each iteration k of the GNN, producing a message $m_{N(u)}^{(k)}$ based on this collective neighborhood information. Subsequently, the $CONCAT$ function first updates the message and then combines it $m_{N(u)}^{(k)}$ with the previous message $m_u^{(k-1)}$ of node, u resulting in the updated message $m_u^{(k)}$. Algorithm 1 shows the steps of our proposed FedGNN model.

For our proposed FedGNN model, $x \in \mathbb{R}^{N \times d}$ where N is the number of nides and d is the number of features per node. For the number of edges $E \in \mathbb{R}^{2 \times E}$. The generated graph, constructed from the normalized and log-transformed features of the nodes, is into FedGNN model. The edge index represents the connections between nodes and is crucial for the convolution operations. In the first layer, the GraphSAGE convolution (SAGEConv) aggregates information from each node's neighbors based on the edge index, followed by batch normalization, Rectified Linear Unit (ReLU) activation, and dropout for regularization. This process is repeated through successive layers, using different convolution operations like Graph Attention Convolution (GATConv) and Graph Isomorphism Convolution (GINConv), each followed by appropriate activations, batch normalization, and dropout. This series of operations progressively refines the node embeddings, capturing complex structural and feature information, ultimately outputting the final node embeddings or classifications. Algorithm 3 shows the steps of our proposed FedGNN model.

Algorithm 3 Our proposed FedGNN

Require: $x \in \mathbb{R}^{N \times d}$
Require: $edge\_index \in \mathbb{R}^{2 \times E}$
Require: $dropout\_rate \in [0,1]$
Ensure: $x \in \mathbb{R}^{N \times num\_classes}$
Initialize parameters for:
SAGEConv layers: $W_{sage1}, W_{sage4}, W_{sage7}$
GATConv layers: W_{gat2}, W_{gat5}
GINConv layers: W_{gin3}, W_{gin6}
BatchNorm layers: $bn1, bn2, bn3, bn4, bn5, bn6$
Dropout Layer with $p = dropout\_rate$
$x_i \leftarrow \text{SAGEConv}_1(x, edge\_index)$
$x_i \leftarrow BatchNorm_1(x_i)$
$x_i \leftarrow ReLU\ (x_i)$
$x_i \leftarrow Dropout\ (x_i, p)$
$x_{ii} \leftarrow \text{GATConv}_2\ (x_i, edge\_index)$
$x_{ii} \leftarrow BatchNorm_2\ (x_{ii})$
$x_{ii} \leftarrow ELU\ (x_{ii})$
$x_{ii} \leftarrow Dropout\ (x_{ii}, p)$
$x_{iii} \leftarrow \text{GATConv}_3\ (x_{ii}, edge\_index)$
$x_{iii} \leftarrow BatchNorm_3\ (x_{iii})$
$x_{iii} \leftarrow ReLU\ (x_{iii})$
$x_{iii} \leftarrow Dropout\ (x_{iii}, p)$

$x_{iv} \leftarrow \text{GATConv}_4\ (x_{iii}, edge\_index)$
$x_{iv} \leftarrow BatchNorm_4\ (x_{iv})$
$x_{iv} \leftarrow ReLU\ (x_{iv})$
$x_{iv} \leftarrow Dropout\ (x_{iv}, p)$
$x_v \leftarrow \text{GATConv}_5\ (x_{iv}, edge\_index)$
$x_v \leftarrow BatchNorm_5\ (x_v)$
$x_v \leftarrow ELU\ (x_v)$
$x_v \leftarrow Dropout\ (x_v, p)$
$x_{vi} \leftarrow \text{GATConv}_6\ (x_v, edge\_index)$
$x_{vi} \leftarrow BatchNorm_6\ (x_{vi})$
$x_{vi} \leftarrow ReLU\ (x_{vi})$
$x_{vi} \leftarrow Dropout\ (x_{vi}, p)$
$x_{vii} \leftarrow \text{SAGEConv}_7(x_{vi}, edge\_index)$
Final Output layer: x_{vii}

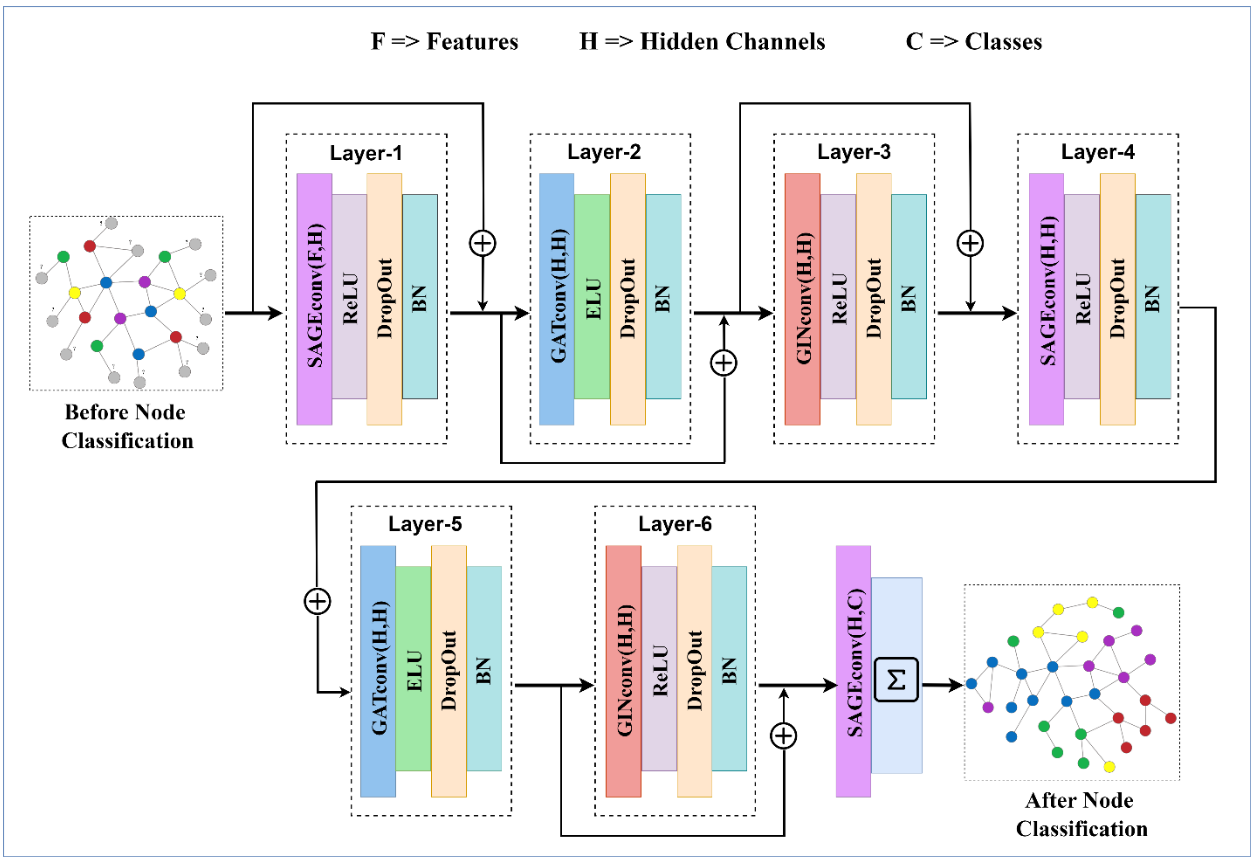


Fig 4.6.1: Proposed GNN Architecture

Our proposed model, FedGNN, is a GNN architecture designed for node classification, as shown in figure 3.9.1. It combines various convolutional layers to leverage their strengths for enhanced feature learning. There are seven convolutional layers in the model: three SAGEConv (GraphSAGE) layers that use MEAN aggregation, two GATConv (Graph Attention) layers that use multi-head attention, and two GINConv (Graph Isomorphism) layers that use sequential neural networks. Each layer processes and transforms the input graph's node features, gradually enhancing their representations. After every convolutional layer, batch normalization is used to make the training more stable, and non-linear activation functions (ReLU or ELU) are added to pick up on complex patterns. Dropout regularization with a probability of 0.3 is employed after each activation function to prevent overfitting. The final SAGEConv layer generates a vector suitable for multi-class classification. The FedGNN model's layered structure, which includes SAGEConv, GATConv, and GINConv, lets it effectively spread information and learn features within graph structures.

CHAPTER 5
Result and Analysis

5.1 Experimental Setup

The experimental setting for the study on "Decentralized Horizontal Federated Learning Using Graph Neural Networks for ECG Signal Analysis" includes a well-organized configuration of hardware, software, and data resources. The computational backbone consists of high-performance central processing unit (CPU) and a powerful graphics processing unit (GPU), which allow for rapid training and deployment of Graph Neural Networks (GNNs). The federated learning models are implemented using deep learning frameworks, specifically TensorFlow or PyTorch, in a Python programming environment. To provide a consistent and representative input for model training and evaluation, the experimental dataset, which consists of a variety of labeled ECG signals, is carefully preprocessed using techniques like denoising and normalization. The federated learning strategy protects data privacy by keeping it locally while sharing model changes, so improving security and compliance with ethical standards. Complete documentation of code particulars model architecture, hyperparameters, and outcomes is maintained throughout the experimentation phase to promote accountability, consistency, and potential collaboration in the advancement of ECG signal processing using decentralized learning frameworks.

5.2 Evaluation metrics

This section evaluates the efficacy of the models using various metrics derived from the confusion matrix, a crucial tool in model assessment. The confusion matrix provides key metrics such as true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) [39]. True positives are correctly identified positive instances, while false positives are incorrect positive predictions. False negatives occur when positive cases are wrongly classified as negative, and true negatives are correctly identified negative instances. From these metrics, precision measures the accuracy of positive predictions, recall indicates the model's ability to identify all relevant positive cases, and the F1-score provides a balanced measure of the model's performance [40]. Accuracy reflects the proportion of correct predictions across all classes. Equations (4–7) were used to evaluate precision, recall, F1-score, and accuracy for chest X-ray disease classification, providing a comprehensive performance assessment.

$$Precision = \frac{TP}{TP + FP} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (5)$$

$$F1 - Score = 2 * \frac{Recall*Precision}{Recall+Precision} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (6)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (7)$$

5.3 Ablation Study

Different feature reconstruction algorithms, extractors, GNN layers, batch sizes, dropout rates, loss functions, learning rates, and K values were tested in the ablation experiments shown in Tables 3 and 4. The goal was to make the proposed model work optimally. In order to achieve optimal performance in a variety of scenarios and configurations, these experiments optimized the model's configurations [41].

Table 5.3.1**:** Ablation study regarding GNN layer, Batch Size and Dropout Rate

Ablation Study 1: Altering the GNN layer							
	Configuration No	**GNN Layer**	**Precision (%)**	**Recall (%)**	**F1-score (%)**	**Test Accuracy (%)**	**Finding**
Chapman	1	1	0.8703	0.8226	0.8268	0.9125	Accuracy dropped
	2	**2**	**0.9063**	**0.9112**	**0.9076**	**0.9302**	**Highest Accuracy**
	3	3	0.8757	0.8889	0.8784	0.9117	Accuracy dropped
PTB-XL	1	1	0.8645	0.8264	0.8356	0.9269	Accuracy dropped
	2	**2**	**0.9534**	**0.8962**	**0.9162**	**0.9706**	**Highest Accuracy**
	3	3	0.8622	0.8438	0.8486	0.9290	Accuracy dropped
Ablation Study 2: Altering the Batch Size							
	Configuration No	**Batch Size**	**Precision (%)**	**Recall (%)**	**F1-score (%)**	**Test Accuracy (%)**	**Finding**
Chapman	1	32	0.9456	0.9452	0.9452	0.9452	Accuracy dropped
	2	**64**	**0.9820**	**0.9820**	**0.9820**	**0.9820**	**Highest Accuracy**
	3	128	0.9919	0.9919	0.9919	0.9919	Accuracy dropped
PTB-XL	1	32	0.6005	0.6172	0.6011	0.6172	Accuracy dropped
	2	64	0.8966	0.8957	0.8949	0.8957	Accuracy dropped
	3	**128**	**0.9524**	**0.9516**	**0.9514**	**0.9516**	**Highest Accuracy**

	Ablation Study 3: Altering the Dropout Rate						
	Configuration No	**Dropout rate**	**Precision (%)**	**Recall (%)**	**F1-score (%)**	**Test Accuracy (%)**	**Finding**
Chapman	**1**	**0.1**	**0.9892**	**0.9893**	**0.9892**	**0.9892**	**Highest Accuracy**
	2	0.2	0.9635	0.9634	0.9634	0.9634	Accuracy dropped
	3	0.4	0.9402	0.9397	0.9397	0.9397	Accuracy dropped
PTB-XL	**1**	**0.1**	**0.9892**	**0.9893**	**0.9992**	**0.9992**	**Highest Accuracy**
	2	0.2	0.9715	0.9715	0.9715	0.9715	Accuracy dropped
	3	0.4	0.9634	0.9634	0.9634	0.9634	Accuracy dropped

Firstly, the study shows that ablation studies that change GNN layers, batch sizes, and dropout rates give important information for making models work better on the Chapman and PTB-XL datasets shown in Table 3. For both datasets, a two-layer GNN architecture yields the highest accuracy, with precision, recall, and F1-scores showing marked improvements over single and three-layer configurations. Batch size significantly impacts performance, with the optimal sizes being 64 for Chapman and 128 for PTB-XL. Additionally, a dropout rate of 0.1 consistently results in the highest accuracy for both datasets, emphasizing the importance of a lower dropout rate to prevent overfitting. These findings highlight the critical role of hyperparameter tuning in enhancing the effectiveness of GNN models.

Table 5.3.2: Ablation study regarding model loss function, Learning rate and K values

	Ablation Study 4: Altering the Loss Functions						
	Configuration No	**Loss Functions**	**Precision (%)**	**Recall (%)**	**F1-score (%)**	**Test Accuracy (%)**	**Finding**
Chapman	1	CrossEntropyLoss	0.98	0.98	0.98	0.99	Accuracy dropped
	2	BCEWithLogitsLoss	0.99	0.98	0.99	0.99	Accuracy dropped
	3	NLL	0.99	0.98	0.99	0.99	Accuracy dropped
	4	**MSE**	**0.99**	**0.98**	**0.99**	**0.99**	**Highest Accuracy**
	5	SmoothL1	0.99	0.98	0.99	0.99	Accuracy dropped
PTB-XL	**1**	CrossEntropyLoss	0.85	0.80	0.82	0.89	Accuracy dropped
	2	BCEWithLogitsLoss	0.85	0.80	0.82	0.89	Accuracy dropped

	3	NLL	0.61	0.62	0.60	0.82	Accuracy dropped
	4	**MSE**	**0.71**	**0.72**	**0.71**	**0.90**	**Highest Accuracy**
	5	SmoothL1	0.65	0.63	0.63	0.82	Accuracy dropped
Ablation study 5: Altering the Learning rate							
	Configuration No	**Learning rate**	**Precision (%)**	**Recall (%)**	**F1-score (%)**	**Test Accuracy (%)**	**Finding**
Chapman	1	0.1	0.8423	0.8332	0.8339	0.8332	Accuracy dropped
	2	0.001	0.9600	0.9661	0.9665	0.9661	Accuracy dropped
	3	**0.006**	**0.9955**	**0.9956**	**0.9955**	**0.9956**	**Highest Accuracy**
PTB-XL	**1**	0.1	0.8516	0.8157	0.8144	0.8157	Accuracy dropped
	2	0.001	0.8913	0.8332	0.8339	0.8332	Accuracy dropped
	3	**0.006**	**0.9800**	**0.9861**	**0.9865**	**0.9861**	**Highest Accuracy**
Ablation Study 6: K values							
	Configuration No	**Values of K**	**Precision (%)**	**Recall (%)**	**F1-score (%)**	**Test Accuracy (%)**	**Finding**
Chapman	1	K=5	0.94	0.94	0.94	0.95	Accuracy dropped
	2	K=10	0.97	0.97	0.97	0.97	Accuracy dropped
	3	K=15	0.97	0.97	0.97	0.97	Accuracy dropped
	4	**K=20**	**0.98**	**0.98**	**0.98**	**0.98**	**Highest Accuracy**
PTB-XL	1	K=5	0.98	0.98	0.98	0.99	Accuracy dropped
	2	**K=10**	**0.99**	**0.99**	**0.99**	**0.99**	**Highest Accuracy**
	3	K=15	0.98	0.98	0.98	0.99	Accuracy dropped
	4	K=20	0.87	0.82	0.83	0.90	Accuracy dropped

Furthermore, the additional ablation studies highlight the effects of altering loss functions, learning rates, and K values on GNN performance for the Chapman and PTB-XL datasets shown in Table 4. For loss functions, MSE achieved the highest accuracy for both datasets, though precision, recall, and F1-scores were generally high across all tested functions. In terms of learning rates, a rate of 0.006 yielded the highest accuracy, emphasizing its critical role in model training. Regarding K values, K = 20 was optimal for Chapman, while K = 10 was best for PTB-XL, demonstrating the importance of neighborhood size in k-nearest neighbors graph construction. These results underscore the need for careful selection of these hyperparameters to maximize model performance.

5.4 Comparison of different ML models

The line plot compares the accuracies of various machine learning models on the Chapman and PTB-XL datasets and does not give satisfactory accuracy, with LightGBM achieving the highest accuracy on Chapman at 58% and Random Forest showing the highest on PTB-XL at 66%. Conversely, K-Nearest Neighbors has the lowest accuracy on Chapman at 37%, while Support Vector Classifier records the lowest on PTB-XL at 31%. These results highlight that LightGBM is the best performer for Chapman and Random Forest for PTB-XL, whereas K-Nearest Neighbors and Support Vector Classifier are the least effective for Chapman and PTB-XL, respectively as shown in figure 4.

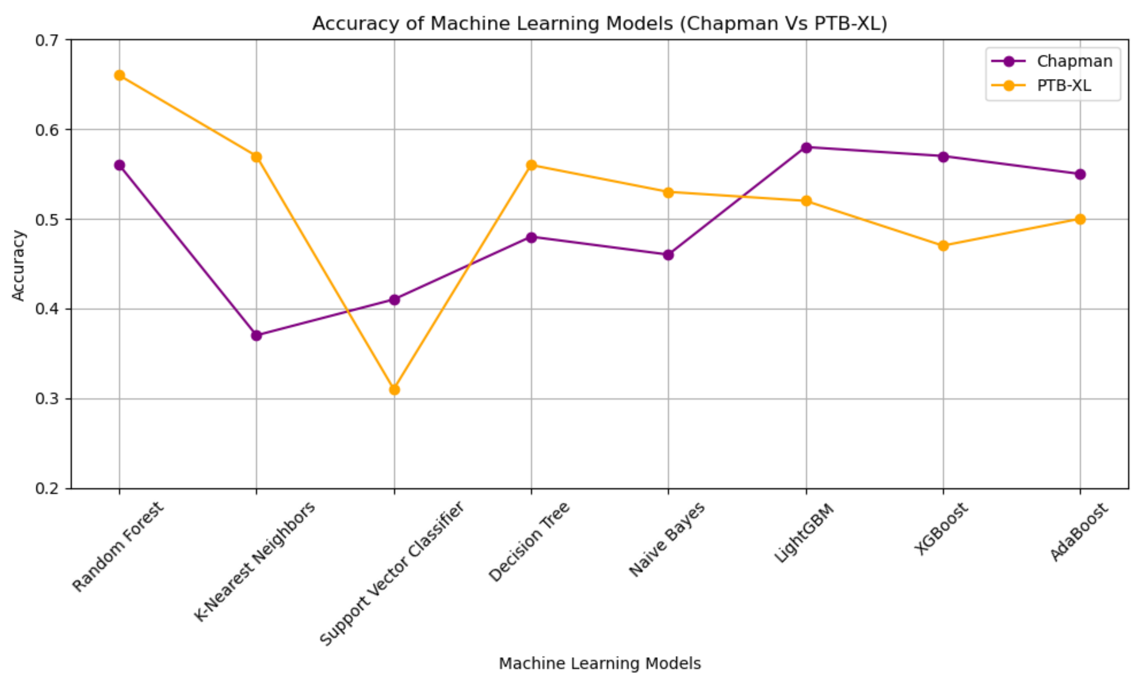


Figure 5.4.1: Comparison of different ML models accuracies on two datasets.

5.5 Comparison of different GNN models with Chapman and PTB-XL datasets

Our proposed FedGNN model outperforms all other compared GNN models on the Chapman dataset, achieving the highest accuracy score of 97.49%. This exceptional performance underscores FedGNN's proficiency in handling graph-based data. GCN closely follows with an accuracy of 81.00%, and GAT with 80.00%. However, there is a noticeable performance gap between FedGNN and the other models. The baseline model ranks higher than GIN, with accuracies of 72.00% and 69.00%, respectively. The superiority of our FedGNN model demonstrates the method optimization in GNN can significantly enhance the accuracy of ECG signal classification, surpassing other models as shown in Table 8.

Table 5.5.1: Comparison between different GNN models and our proposed model (FedGNN) for the ChapmanECG dataset

Comparison for the Chapman dataset								
Model	**F1-Score**	**Sensitivity**	**Specificity**	**Miss Classification Rate**	**Precision**	**False Positive Rate**	**False Negative Rate**	**Accuracy**
GIN	69.00%	69.00%	88.44%	31.00%	69.00%	11.56%	30.70%	69.00%
Base Line	72.00%	70.83%	90.26%	28.00%	72.00%	9.74%	29.17%	72.00%
GAT	78.00%	79.00%	91.15%	20.00%	77.00%	8.85%	21.72%	80.00%
GCN	81.00%	81.00%	94.00%	19.00%	83.00%	5.98%	18.99%	81.00%
FedGNN (Our Proposed model)	97.50%	97.82%	97.96%	2.51%	97.19%	1.04%	1.18%	97.49%

Our proposed FedGNN model outperforms all other compared GNN models on the ChapmanECG dataset, achieving the highest accuracy score of 97.49% as shown in table 9. This exceptional performance underscores FedGNN's proficiency in handling graph-based data. GCN follows with an accuracy of 81.00%, while GAT and GIN achieve 80.00% and 69.00%, respectively. The baseline model ranks higher than GIN, with an accuracy of 72.00%. The superiority of our FedGNN model is also evident in its sensitivity of 97.82% and specificity of 97.96%, which are significantly higher than those of the other models. Additionally, the FedGNN model has the lowest miss classification rate of 2.51%, false positive rate of 1.04%, and false negative rate of 1.18%, alongside the highest precision of 97.19% and F1 score of 97.50%. This highlights how optimization in GNN can significantly enhance the accuracy of ECG signal classification, surpassing previous models.

Table 5.5.2: Comparison between different GNN models and our proposed model (FedGNN) for the PTB-XL dataset

Comparison for the PTB-XL dataset								
Model	**F1-Score**	**Sensitivity**	**Specificity**	**Miss Classification Rate**	**Precision**	**False Positive Rate**	**False Negative Rate**	**Accuracy**
Base Line	70.80%	70.80%	91.33%	30.00%	70.80%	8.67%	29.20%	70.00%
GIN	74.00%	74.64%	86.41%	26.00%	73.80%	13.59%	25.36%	74.00%
GCN	73.00%	75.50%	75.60%	24.00%	70.00%	24.36%	24.45%	76.00%
GAT		79.00%	91.15%	20.00%	77.00%	8.85%	21.72%	80.00%
FedGNN (Our Proposed model)	99.00%	99.80%	99.96%	10.00%	99.00%	0.55%	0.01%	99.00%

5.6 Comparison Performance Analysis of ECG Signal Dataset Distributed Across Two Different Client

Table 5.6.1 describes the performance metrics for two individual clients (Client-1 and Client-2) and a global model in a DFL framework. Client-1 demonstrates impressive performance with a sensitivity of 97.06%, specificity of 99.15%, accuracy of 98.63%, and precision of 97.49%, with a false positive and false negative rate of 0.85% and 2.94%, respectively. Client-2 exhibits better results compared to Client-1 in certain metrics, achieving a sensitivity of 97.17%, a specificity of 99.28%, and an accuracy of 98.85%. The precision is 97.00%, with a false positive rate of 0.72% and a false negative rate of 2.83%. On the other hand, the global model combines the strengths of both clients, achieving a sensitivity of 97.12%, a specificity of 99.22%, and an accuracy of 98.74%. Precision is 97.25%, with a false positive rate of 0.79% and a false negative rate of 2.89%. Overall, the global model provides a balanced and robust performance by leveraging the DFL approach.

Table 5.6.1: Comparison of clients (Client-1 and Client-2) with global model

	Sensitivity (%)	**Specificity (%)**	**Miss Classification Rate (%)**	**Precision (%)**	**False Positive Rate (%)**	**False Negative Rate (%)**	**F1 Score (%)**	**Accuracy (%)**
Client-1	97.06	99.15	1.37	97.49	0.85	2.94	97.28	98.63
Client-2	97.17	99.28	1.15	97.00	0.72	2.83	97.17	98.85
Global	97.12	99.22	1.26	97.25	0.79	2.89	97.23	98.74

5.7 Comparison of global accuracies of different GNN models in DFL

The accuracy of various GNN models is significantly varying in the bar chart in figure 5. FedGNN achieves the highest accuracy at 98.74%, while GIN has the lowest at 57.19%. BaselineGNN, GCN, and GAT also demonstrate satisfactory performance, with accuracies of 67.82%, 69.90%, and 73.25%, respectively. This comparison underscores our proposed model FedGNN's exceptional performance, which nearly reaches 99%, while GIN is the most underperforming at just over 57%.

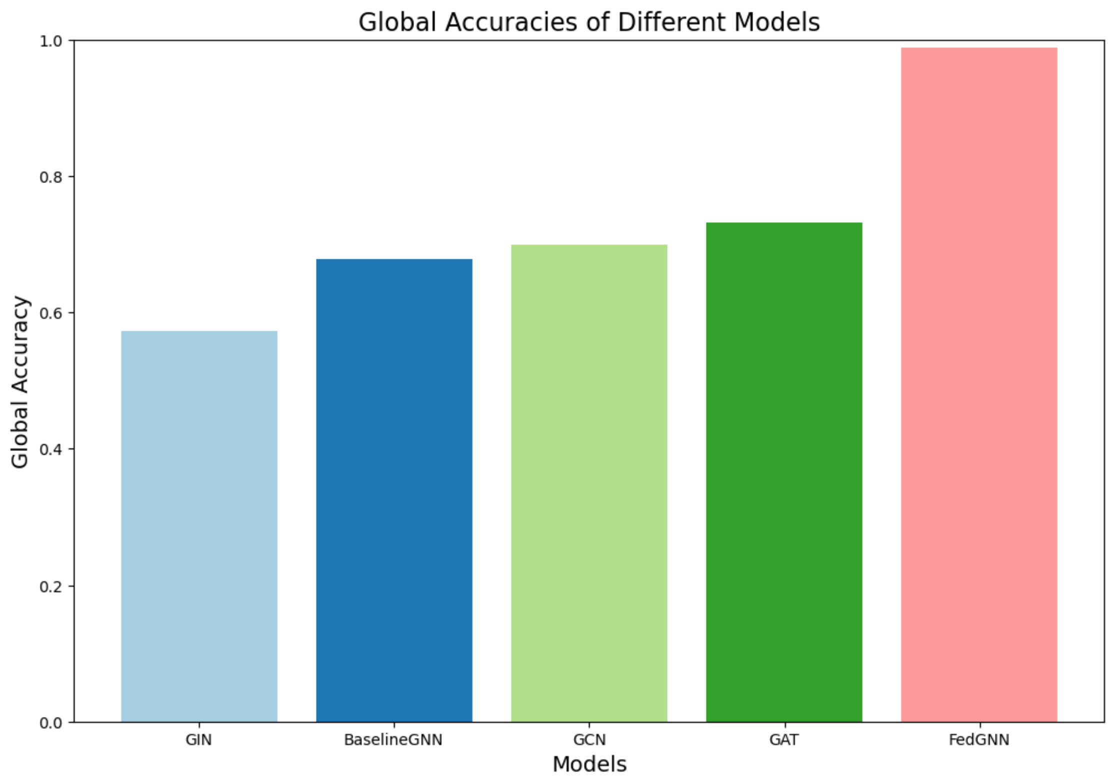


Fig 5.7.1: Global accuracies of different GNN models and FedGNN(Our proposed) model.

5.8 t-distributed stochastic neighbor embedding (t-SNE) node embedding

The t-SNE method visually identifies patterns and clusters in complex datasets while preserving their structures [42]. Its visualization shows clear differences in how things are distributed across different categories, which shows how well it works for data analysis [43]. Figure 7 illustrates t-SNE maps of feature maps from different models. Using four colors for the Chapman Dataset [19] categories and five colors for the PTB-XL Dataset [18] categories, the visualization initially showed classes. After applying the models, the GIN model performed exceptionally well, followed closely by the GAT model, and the GCN model improved clustering slightly. However, our proposed FedGNN model significantly enhanced class clustering, demonstrating its effectiveness in analyzing large and complex datasets.

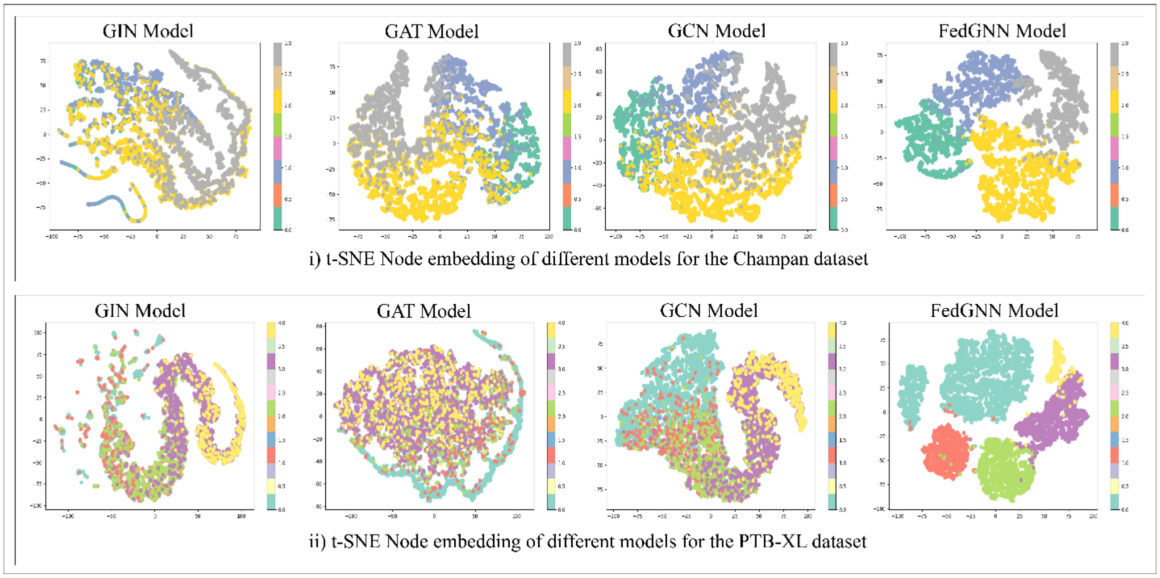


Fig.5.8.1: t-SNE node embedding visualization with different GNN models.

5.9 Performance analysis of the FedGNN model

Our proposed model, FedGNN, demonstrates exceptional performance in accurately classifying ECG signals, achieving an impressive accuracy of 98.63% for Chapman dataset (Client-1) and 98.85% for PTB-XL dataset (Client-2) [40]. The confusion matrices for both clients demonstrate the high effectiveness of FedGNN in accurately predicting all classes, with very few incorrect predictions. This demonstrates the absence of discrimination against any particular category, showcasing its efficacy across the entire range [44]. In addition, the FedGNN model consistently achieved optimal accuracy levels that reached 98.63% for all categories. The receiver operating characteristic (ROC) curves highlight the model's reliability and consistency in predicting various classes of ECG signals.

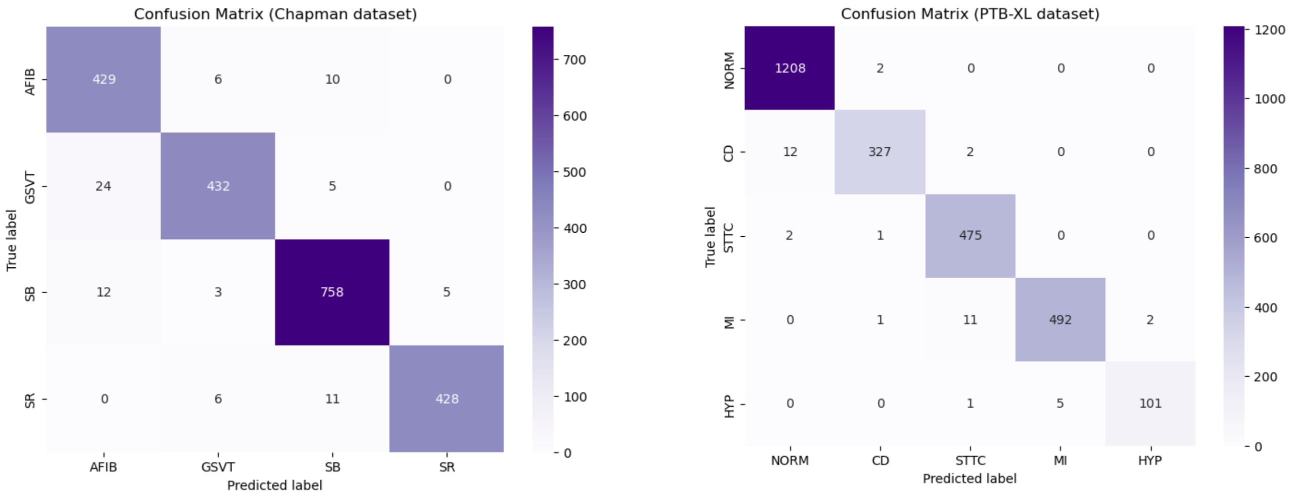


Fig.5.9.1: Confusion matrix of FedGNN model.

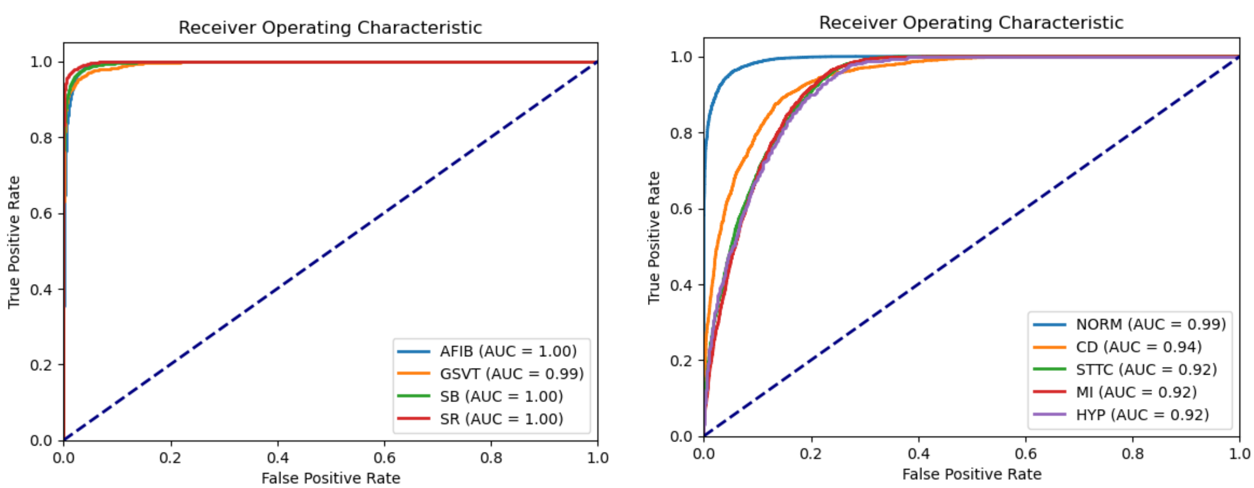


Fig.5.9.2: ROC curve of FedGNN model.

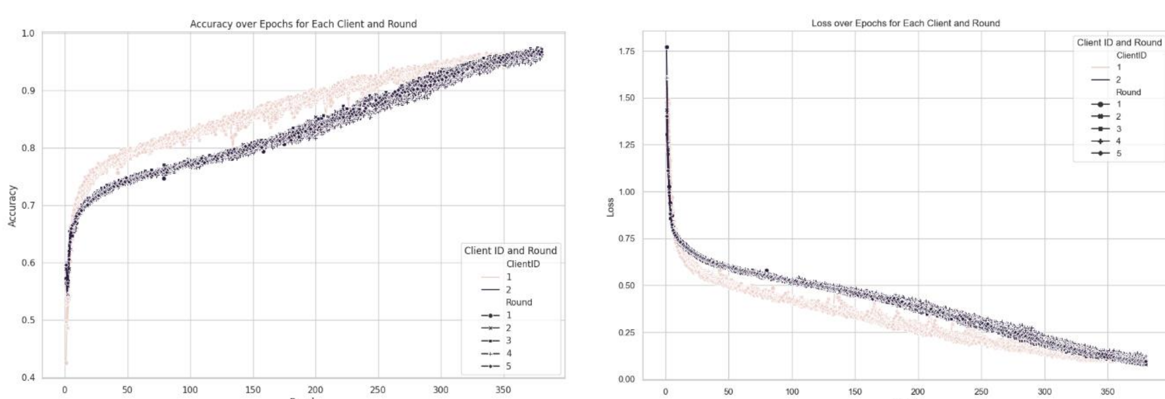


Fig.5.9.3: Illustration of the training and validation accuracy and loss curves of the FedGNN model

The loss and accuracy curves for each epoch show a generally constant pattern with some variations [43]. Our model's accuracy and loss curves show a consistent pattern with minor variations over 350 epochs. Both training and validation accuracies steadily improve with minimal fluctuation, indicating balanced learning without significant overfitting. The plot highlights the robust performance of the FL model, with accuracy consistently rising for both Client 1 (light pink) and Client 2 (dark blue) across all rounds. There were no noticeable changes in the loss and accuracy plots during the training time, which means there were no problems with overfitting. The FedGNN model's loss and accuracy curves can be seen in Figure 5.9.3.

CHAPTER 6

Impact on Society, Environment and Sustainability

6.1 Impact on Society

This study's findings on "Decentralized Horizontal Federated Learning Using Graph Neural Networks for ECG Signal Analysis" have far-reaching consequences for society and healthcare. Accurate and early detection using the federated learning global model has the potential to transform medical diagnostics, resulting in better patient outcomes and wider social benefits. This study intends to improve the quality and efficiency of ECG signal processing using federated learning and Graph Neural Networks (GNNs), allowing healthcare practitioners to make more informed and confident judgments.

This research has a significant social impact, particularly in terms of public health. Numerous deadly diseases impose significant stress on global healthcare systems. The use of strong and accurate diagnostic technologies has the potential to improve patient care, eliminate misdiagnoses, and speed up therapeutic interventions. This, in turn, can help to reduce illness complications, hospitalizations, and related healthcare expenses.

Furthermore, an upgraded and powerful federated learning model has the potential to reach underprivileged populations and regions with inadequate healthcare resources because it is both accessible and affordable. The technology developed in this study has the potential to cross boundaries of geography, bringing advanced diagnostic services to areas where traditional medical services may be unavailable and assisting in the overarching goal of reducing disparities in health and improving overall equality in healthcare.

In summary, the social impact of this discovery derives from its potential for bringing in a new era of precision healthcare, notably in the detection and treatment of various diseases. This research, which uses cutting-edge technologies, not only enhances scientific understanding of diagnosis but also holds the key to practical gains in public health outcomes, healthcare accessibility, and the overall well-being of various populations around the world.

6.2 Impact on Environment

While the research on "Decentralized Horizontal Federated Learning Using Graph Neural Networks for ECG Signal Analysis" is primarily concerned with healthcare and technology breakthroughs, its environmental impact is complex but significant. While federated learning and Graph Neural Networks (GNNs) can lead to advancements in medical diagnoses, they are

fundamentally computationally costly. These operations, particularly the training and optimization of sophisticated neural networks, require significant computer power and frequently rely on energy-intensive technology such as graphics processing units (GPUs).
The environmental impact is caused by the energy consumption required to train and deploy these advanced models. The high computing needs, particularly in scenarios with huge datasets and complicated model designs, contribute to greater energy usage.
However, it is important to remember that improvements in hardware and cloud computing, and sustainable computing practices are evolving. Researchers and developers are becoming more aware of the environmental risks involving AI applications, pushing efforts to improve algorithms and hardware for energy efficiency. However, the utilization of renewable energy sources in servers can help to reduce environmental effects while also connecting technology advancement with eco-friendly considerations.

6.3 Ethical Aspects

The research on "Decentralized Horizontal Federated Learning Using Graph Neural Networks for ECG Signal Analysis" is supported by a strong dedication to ethical considerations throughout the technique. The use of private health information, such as ECG signals, implies a significant responsibility for protecting patient privacy and confidentiality, with a focus on strict adherence to moral standards and regulatory regulations. Federated learning meets these ethical imperatives by keeping data decentralized and local, ensuring that patient information is never exchanged directly across institutions.
All study participants provide informed consent, which is an essential component of ethical research techniques. This ensures that patients understand how their data will be utilized, as well as the potential advantages and hazards of the research. The potential social influence of healthcare research highlights the significance of ethical deployment, which ensures that advances in diagnostic capacities lead to better patient care without harming individual rights. Furthermore, the use of this technology in real-world clinical settings is driven by ethical considerations to guarantee that it improves healthcare delivery while avoiding biases or injustices.
Overall, ethical considerations are fundamental to this study's dedication to improving society and the proper use of modern technology in the healthcare area. This project aims to advance diagnostic technology while safeguarding and upholding individual rights.

6.4 Sustainability Plan

The sustainability plan for this research is intended to reduce environmental impact while maintaining positive societal contributions. To reduce overall energy consumption related with model training and deployment, efforts will be focused on optimizing computational efficiency and implementing energy-efficient techniques.

Furthermore, the research team is dedicated to developing and implementing sustainable strategies. The strategy also calls for the implementation of sustainable methods in data management and storage, such as using energy-efficient data centers and improving data processing workflows to reduce resource consumption.

Furthermore, the sustainability strategy calls for frequent assessments of the environmental impact of research operations, as well as the implementation of necessary adjustments. This project aims to benefit both the field of medical AI and the larger goal of environmental sustainability by including a dedication to ethical and eco-conscious scientific innovation.

CHAPTER 7

Conclusion and Future work

7.1 Summary of the Study

The fundamental goal of this project is to create a privacy-preserving, decentralized horizontal federated learning model based on ECG signal data. The study employs two publicly available electrocardiography datasets to extract useful information from ECG signals. The research uses Wavelet preprocessing techniques and advanced feature extraction methods to ensure high-quality data for analysis. Graph Neural Networks (GNNs) are used to create graphs that represent the complicated correlations found in ECG data.

The study entails training local models at specific healthcare institutions and pooling their parameters to create a strong global model without disclosing sensitive patient information. This technique respects patient privacy and complies with data protection requirements. The global federated learning model is thoroughly evaluated for correctness and dependability.

Overall, this study makes major contributions by introducing a novel approach that blends federated learning with GNNs. The study emphasizes the use of modern AI algorithms to improve the accuracy and privacy of ECG signal processing, resulting in better patient outcomes and healthcare efficiency. The findings of this study open the path for future advances in medical AI, underlining the relevance of ethical issues and the appropriate use of AI technologies in healthcare.

7.2 Conclusions

The primary purpose of this project was to develop a privacy-preserving horizontal federated learning model based on ECG signal data. The study succeeded in extracting essential insights from ECG signals by training local models on two publically available electrocardiography datasets, offering valuable information. The use of Graph Neural Networks (GNNs) assisted the development of graphs that encapsulate the intricate relationships found in ECG data.

This study aimed to construct a global model in a decentralized manner, with local models trained individually and their parameters aggregated into a global federated learning model. The findings demonstrate the effectiveness of integrating federated learning and GNNs to improve accuracy.

This study adds considerably to the field of healthcare by proposing a novel approach that protects data privacy while maintaining analytical performance.

Future study should look into integrating additional preprocessing approaches, as well as investigating other advanced GNN architectures to improve model performance.

7.3 Implication for Further Study

The outcomes of this study on constructing a privacy-preserving horizontal federated learning model for ECG signal interpretation using Graph Neural Networks (GNNs) to provide useful insights and open up new options for future research. One important implication for future research is the examination of other large databases. In addition, using alternative preprocessing approaches could increase the model's accuracy and robustness.

Furthermore, more advanced feature extraction techniques could be used to improve the model's ability to collect essential characteristics of ECG data.

Given the computational burden of the federated learning strategy, future study could look into alternative and more efficient federated learning techniques. This suggests the need to create a more efficient and scalable federated learning paradigm.

Future research could focus on unique GNN architectures, various datasets, and real-world clinical applications to gain a more comprehensive understanding of this key area in medical diagnosis.

In conclusion, the implications for future research include improving preprocessing procedures, investigating feature extraction techniques, federated learning frameworks, and broadening the scope of research in ECG signal analysis and beyond. These activities will broaden the scope of medical AI and promote improvements in diagnostic accuracy, efficiency, and patient outcomes.

REFERENCES

[1] "Cardiovascular diseases," Who.int. [Online]. Available: https://www.who.int/health-topics/cardiovascular-diseases/. [Accessed: 27-Jun-2024].

[2] S. K. Berkaya, A. K. Uysal, E. S. Gunal, S. Ergin, S. Gunal, and M. B. Gulmezoglu, "A survey on ECG analysis," Biomedical Signal Processing and Control, vol. 43, pp. 216–235, 2018.

[3] R. Tang, J. Luo, J. Qian, and J. Jin, "Personalized federated learning for ECG classification based on feature alignment," Secur. Commun. Netw., vol. 2021, pp. 1–9, 2021.

[4] G. X. Yan and C. Antzelevitch, "Cellular basis for the normal T wave and the electrocardiographic manifestations of the long-QT syndrome," Circulation, vol. 98, no. 18, pp. 1928–1936, 1998.

[5] M. Paul, L. Maglaras, M. A. Ferrag, and I. Almomani, "Digitization of healthcare sector: A study on privacy and security concerns," ICT Express, vol. 9, no. 4, pp. 571–588, 2023.

[6] J. Konečný, H. B. McMahan, D. Ramage, and P. Richtárik, "Federated Optimization: Distributed machine learning for on-device intelligence," arXiv [cs.LG], 2016.

[7] L. Yuan, Z. Wang, L. Sun, P. S. Yu, and C. G. Brinton, "Decentralized federated learning: A survey and perspective," arXiv [cs.LG], 2023.

[8] S. Sakib, M. M. Fouda, Z. Md Fadlullah, K. Abualsaud, E. Yaacoub, and M. Guizani, "Asynchronous federated learning-based ECG analysis for arrhythmia detection," in 2021 IEEE International Mediterranean Conference on Communications and Networking (MeditCom), 2021, pp. 277–282.

[9] S. Baghersalimi, T. Teijeiro, D. Atienza, and A. Aminifar, "Personalized real-time federated learning for epileptic seizure detection," IEEE J. Biomed. Health Inform., vol. 26, no. 2, pp. 898–909, 2022.

[10] T.-H. Hwang, J. Shi, and K. Lee, "Enhancing privacy-preserving personal identification through federated learning with multimodal vital signs data," IEEE Access, vol. 11, pp. 121556–121566, 2023.

[11] L. Zou, Z. Huang, X. Yu, J. Zheng, A. Liu, and M. Lei, "Automatic detection of congestive heart failure based on multiscale residual UNet++: From centralized learning to federated learning," IEEE Trans. Instrum. Meas., vol. 72, pp. 1–13, 2023.

[12] R. N. Asif et al., "Detecting electrocardiogram arrhythmia empowered with weighted federated learning," IEEE Access, vol. 12, pp. 1909–1926, 2024.

[13] W. Chorney and H. Wang, "Towards federated transfer learning in electrocardiogram signal analysis," Comput. Biol. Med., vol. 170, no. 107984, p. 107984, 2024.

[14] A. Raza, K. P. Tran, L. Koehl, and S. Li, "Designing ECG monitoring healthcare system with federated transfer learning and explainable AI," Knowl. Based Syst., vol. 236, no. 107763, p. 107763, 2022.

[15] M. N. Meqdad, A. H. Hussein, S. O. Husain, and A. M. Jawad, "Classification of electrocardiogram signals based on federated learning and a gaussian multivariate aggregation module," Indones. J. Electr. Eng. Comput. Sci., vol. 30, no. 2, p. 936, 2023.

[16] J. H. Yoo et al., "Personalized federated learning with clustering: Non-IID heart rate variability data application," in 2021 International Conference on Information and Communication Technology Convergence (ICTC), 2021, pp. 1046–1051.

[17] A. Rehman, S. Abbas, M. A. Khan, T. M. Ghazal, K. M. Adnan, and A. Mosavi, "A secure healthcare 5.0 system based on blockchain technology entangled with federated learning technique," arXiv [cs.LG], 2022.

[18] P. Wagner, N. Strodthoff, R.-D. Bousseljot, W. Samek, and T. Schaeffter, "PTB-XL, a large publicly available electrocardiography dataset." PhysioNet, 09-Nov-2022.

[19] Zheng, Jianwei (2019). ChapmanECG. figshare. Collection. https://doi.org/10.6084/m9.figshare.c.4560497.v1

[20] P. Lyakhov, M. Kiladze, and U. Lyakhova, "System for neural network determination of atrial fibrillation on ECG signals with wavelet-based preprocessing," Appl. Sci. (Basel), vol. 11, no. 16, p. 7213, 2021.

[21] K. Bnou, S. Raghay, and A. Hakim, "A wavelet denoising approach based on unsupervised learning

model," EURASIP J. Adv. Signal Process., vol. 2020, no. 1, 2020.
[22] G. Luo and D. Zhang, "Wavelet denoising," in Advances in Wavelet Theory and Their Applications in Engineering, Physics and Technology, D. Baleanu, Ed. London, England: InTech, 2012.
[23] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," Biometrika, vol. 81, no. 3, pp. 425–455, 1994.
[24] V. Bhagyalakshmi, R. V. Pujeri, and G. D. Devanagavi, "GB-SVNN: Genetic BAT assisted support vector neural network for arrhythmia classification using ECG signals," J. King Saud Univ. - Comput. Inf. Sci., vol. 33, no. 1, pp. 54–67, 2021.
[25] T. Bender et al., "Analysis of a deep learning model for 12-lead ECG classification reveals learned features similar to diagnostic criteria," IEEE J. Biomed. Health Inform., vol. 28, no. 4, pp. 1848–1859, 2024.
[26] X. Sun et al., "Decreases in electrocardiographic R-wave amplitude and QT interval predict myocardial ischemic infarction in rhesus monkeys with left anterior descending artery ligation," PLoS One, vol. 8, no. 8, p. e71876, 2013.
[27] S. Senthamil Selvan, S. P. Arjunan, R. Swaminathan, and D. K. Kumar, "Complexity analysis in the PR, QT, RR and ST segments of ECG for early assessment of severity in cardiac autonomic neuropathy," Appl. Sci. (Basel), vol. 12, no. 11, p. 5746, 2022.
[28] I. Campero Jurado, A. Fedjajevs, J. Vanschoren, and A. Brombacher, "Interpretable assessment of ST-segment deviation in ECG time series," Sensors (Basel), vol. 22, no. 13, p. 4919, 2022.
[29] U. R. Acharya et al., "Automated detection and localization of myocardial infarction using electrocardiogram: a comparative study of different leads," Knowl. Based Syst., vol. 99, pp. 146–156, 2016.
[30] J. Wang, L. Tian, G. Zhu, C. Liu, and F. Long, "Indoor positioning privacy protection method based on federated learning in MEC environment," Mob. Inf. Syst., vol. 2022, pp. 1–10, 2022.
[31] X. Zhou et al., "Hierarchical federated learning with social context clustering-based participant selection for internet of medical things applications," IEEE Trans. Comput. Soc. Syst., vol. 10, no. 4, pp. 1742–1751, 2023.
[32] J.-H. Kim, J.-H. Choi, Y.-H. Park, C. K.-S. Leung, and A. Nasridinov, "KNN-SC: Novel spectral clustering algorithm using k-nearest neighbors," IEEE Access, vol. 9, pp. 152616–152627, 2021.
[33] S. Kang, "K-nearest neighbor learning with graph neural networks," Mathematics, vol. 9, no. 8, p. 830, 2021.
[34] P. K. Rao, S. Chatterjee, K. Nagaraju, S. B. Khan, A. Almusharraf, and A. I. Alharbi, "Fusion of graph and tabular deep learning models for predicting chronic Kidney Disease," Diagnostics (Basel), vol. 13, no. 12, p. 1981, 2023.
[35] S. Baghersalimi, T. Teijeiro, A. Aminifar, and D. Atienza, "Decentralized federated learning for epileptic seizures detection in low-power wearable systems," IEEE Trans. Mob. Comput., pp. 1–16, 2024.
[36] S. Baghersalimi, T. Teijeiro, D. Atienza, and A. Aminifar, "Personalized real-time federated learning for epileptic seizure detection," IEEE J. Biomed. Health Inform., vol. 26, no. 2, pp. 898–909, 2022.
[37] Z. Li, Z. Cui, S. Wu, X. Zhang, and L. Wang, "Fi-GNN: Modeling feature interactions via graph neural networks for CTR prediction," in Proceedings of the 28th ACM International Conference on Information and Knowledge Management, 2019.
[38] B. Khemani, S. Patil, K. Kotecha, and S. Tanwar, "A review of graph neural networks: concepts, architectures, techniques, challenges, datasets, applications, and future directions," J. Big Data, vol. 11, no. 1, 2024.
[39] A Lightweight Robust Deep Learning Model Gained High Accuracy in Classifying a Wide Range of Diabetic Retinopathy Images. .
[40] K. Fatema et al., "Development of an automated optimal distance feature-based decision system for diagnosing knee osteoarthritis using segmented X-ray images," Heliyon, vol. 9, no. 11, p. e21703, 2023.
[41] S. Rana et al., "DeepChestGNN: A comprehensive framework for enhanced lung disease identification through advanced graphical deep features," Sensors (Basel), vol. 24, no. 9, p. 2830, 2024.

[42] S. Duan et al., “Semi-supervised classification of fundus images combined with CNN and GCN,” J. Appl. Clin. Med. Phys., vol. 23, no. 12, 2022.
[43] D. M. Chan, R. Rao, F. Huang, and J. F. Canny, “T-SNE-CUDA: GPU-Accelerated T-SNE and its Applications to Modern Data,” in 2018 30th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD), 2018.

Appendix A

Course Outcomes, Complex Engineering Problems (EP) and Complex Engineering Activities (EA) Addressing

Title:
DECENTRALIZED HORIZONTAL FEDERATED LEARNING USING GRAPH NEURAL NETWORKS FOR ECG SIGNAL

Student ID: [202-15-3800, 202-15-3802]

CO Description for FYDP

CO	CO Descriptions	PO
Phase -I		
CO1	Integrate recently gained and previously acquired knowledge to identify a problem for the Final Year Design Project (FYDP)	**PO1**
CO2	Analyze different aspects of the goals in designing a solution for this FYDP	**PO2**
CO3	Explore diverse problem domains through a literature review, delineate the issues, and establish this goals for the FYDP	**PO4**
CO4	Perform economic evaluation and cost estimation and employ suitable project management procedures throughout the development life cycle of the FYDP	**PO11**
Phase -II		
CO5	Design and develop technical solutions and system components or processes that meet specified requirements, ensuring compliance with public health and safety standards, as well as considering cultural, socioeconomic, and environmental factors in this FYDP	**PO3**
CO6	Choose and apply appropriate methodologies, resources, and contemporary engineering and IT technologies to address complex engineering processes, encompassing prediction and modeling, while adhering to relevant constraints in this FYDP	**PO5**
CO7	Analyze societal, health, safety, legal, and cultural considerations, along with associated responsibilities, in the context of professional engineering practice and the resolution of this problem, employing logical reasoning guided by contextual understanding.	**PO6**
CO8	Comprehend and evaluate the enduring sustainability and impact of professional engineering endeavors in addressing intricate engineering challenges within social and environmental frameworks.	**PO7**
CO9	Implement ethical principles and adhere to professional standards and norms in this FYDP	**PO8**
CO10	Capable of operating proficiently both individually and as a team member or leader across diverse teams and interdisciplinary settings in this FYDP.	**PO9**

CO11	Proficiently communicate with the engineering community and broader society regarding complex engineering endeavors, including the ability to comprehend and generate comprehensive reports and design documentation, as well as provide and receive clear instructions throughout this FYDP.	**PO10**
CO12	Acknowledge the importance of self-directed and life-long learning within the evolving landscape of technology, and possess the readiness and capability to engage in lifelong learning endeavors.	**PO12**

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP), and Attainment of Complex Engineering Activities (EA)

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP):

SN	EP Definition	Attainment	CO	Justification (with Knowledge Profile)	References
1.	**EP1: Depth of Knowledge required**	Yes	CO1, CO2, CO3, CO5, CO6, CO7 and CO8	In our project, we utilized **the natural science morals (K1)** to understand the electrocardiogram or ECG signals to get insights into the abnormalities of signals, waveforms, and patterns, while applying **mathematical concepts (K2)** for data analysis and model optimization to improve the effectiveness and efficiency of our methodology.	**Page no:** [31-32] **Section:** [3.3] **Page no:** [32-34] **Section:** [3.4]
				This project illustrates **fundamental engineering (K3)** principles by using graph neural networks, data preprocessing techniques, and a variety of GNN structures for signal processing purposes. The study exhibits **specialist knowledge (K4)** by combining federated learning with advanced GNN architectures to improve ECG signal processing.	**Page no:** [43-46] **Section:** [4.6] **Page no:** [41-46] **Section:** [4.3, 4.4, 4.5, 4.6]

				By applying the principles of **engineering practice & design (K5),** we have designed our proposed approach and implemented various techniques to build the federated learning framework. and the project deals with **engineering practice & technology (K6)** by employing federated learning model.	**Page no:** [31] **Section:** [3.2] **Page no:** [41-43] **Section:** [4.3, 4.5]
				Incorporating **K8 (Research Literature)**, we have reviewed several studies related to our approach to get insights into the models, approaches, and gaps that improved the vision of our project.	**Page no:** [22-28] **Section:** [2.2, 2.3]
2.	**EP2: Range of Conflicting Requirements**	Yes	CO2, and CO7	This project deals with **EP-2** by addressing issues in data preparation for both datasets and extracting various features from the signals, which were handled in our project using data preprocessing techniques and handcrafted feature extraction methods.	**Page no:** [37-40] **Section:** [4.2]
3.	**EP3: Depth of analysis required**	Yes	CO2, and CO6	Obtaining an efficient and transparent answer for our planned work is a difficult task. To create the proposed federated learning model using ECG datasets, this research targets **EP-3** by assessing several methodologies and selecting deep learning as the specific model.	**Page no:** [43-46] **Section:** [4.6]

4.	**EP4: Familiarity of Issues**	Yes	CO8	This project addressed **EP-4** by studying different types and manners of federated learning models and various feature extraction techniques based on ECG signals that helped us to figure out the type of federated learning we utilized and the types of features we have extracted from the signals.	**Page no:** [22-28] **Section:** [2.2, 2.3]
5.	**EP6: Extends of stakeholders involved and conflicting**	Yes	CO8	Throughout our project, we have visited several stakeholders such as researchers, and cardiologists to get an idea of our project and the ECG signals we have used to build the model, that indicates **EP6.**	**Page no:** [17-20] **Section:** [1.6]
6.	**EP7: Interdependence**	Yes	CO5	Our project builds from different modules that depend on each other, including data collection and preprocessing, feature extraction, graph generation, training the local models, developing the global model by passing the parameters, and a real-life implementation, which ensures **EP-7**.	**Page no:** [37-46] **Section:** [4.2, 4.3, 4.4, 4.5, 4.6]

Addressing CO11 with Complex Engineering Activities (EA) [Some or all of the following]:

SN	EA Definition	Attainment	CO	Justification	References
1.	**EA1: Range of resources**	Yes	CO11	Our project makes use of a variety of resources, including high-performance computing infrastructure, GPUs, deep learning frameworks, ECG datasets, and ethical considerations, to guarantee planned research and contribute to improvements in decentralized horizontal federated learning with Graph Neural Networks for ECG signal analysis.	**Page no:** [34-36] **Section:** [3.5]

2.	**EA4: Consequences for society and the environment**	Yes		This initiative benefits society by improving diagnosis with advanced ECG signal analysis methods, as well as encouraging environmental sustainability by using effective computing power and complying to ethical rules for patient data protection.	**Page no:** [58-60] **Section:** [6.1, 6.2, 6.3, 6.4]
3.	**EA-5: Familiarity**	Yes		By analyzing a novel approach to ECG signal analysis using Graph Neural Networks (GNNs) and decentralized horizontal federated learning, thorough comparative analysis, this project builds on previous research and provides new insights into the field.	**Page no:** [22-28] **Section:** [2.2, 2.3]

Addressing CO (4, 9, 10, and 12):

SN	COs	Attainment	Justification	References
1	CO4	Yes	In the study of decentralized horizontal federated learning using Graph Neural Networks (GNNs) for ECG signal analysis, this project addresses the **CO4** by integrating efficient project management and financial management by guaranteeing careful scheduling, allocating resources, and estimation of budget for the best use all through the research lifecycle.	**Page no:** [17-20] **Section:** [1.6]
2	CO9	Yes	The initiative follows ethical standards by prioritizing patient privacy, gaining informed consent, and fully documenting the research procedure. This promotes responsible information transmission and societal well-being by ethically applying modern technologies in decentralized horizontal federated learning with Graph Neural Networks (GNNs) for ECG signal processing, in accordance with **CO9.**	**Page no:** [59] **Section:** [6.3]
3	CO12	Yes	The project's dedication to continuous learning **(CO12)** highlights the significance of self-directed and lifelong learning in today's rapidly changing technological context. Recognizing that improvements in federated learning and ECG signal analysis are always evolving, we highlight the	**Page no:** [37-46] **Section:**

			importance of continuing education and skill development. Our project exemplifies this approach by encouraging continual learning and adaptability, preparing individuals to navigate and contribute effectively in the dynamic world of technology.	[4.1, 4.2, 4.3, 4.4, 4.5, 4.6]

Federated learning

ORIGINALITY REPORT

22%	16%	11%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	7%
2	Submitted to Daffodil International University Student Paper	2%
3	www.mdpi.com Internet Source	2%
4	Shakil Rana, Md Jabed Hosen, Tasnim Jahan Tonni, Md. Awlad Hossen Rony et al. "DeepChestGNN: A Comprehensive Framework for Enhanced Lung Disease Identification through Advanced Graphical Deep Features", Sensors, 2024 Publication	1%
5	www.ncbi.nlm.nih.gov Internet Source	1%
6	www.tandfonline.com Internet Source	1%
7	U. Rajendra Acharya, Hamido Fujita, Vidya K. Sudarshan, Shu Lih Oh et al. "Automated detection and localization of myocardial	1%