International Journal of Mathematical, Engineering and Management Sciences
Vol. 10, No. 1, 218-248, 2025
https://doi.org/10.33889/IJMEMS.2025.10.1.013

Privacy-Preserving and Collaborative Federated Learning Model for the Detection of Ocular Diseases

Seema Gulati
Chitkara University Institute of Engineering and Technology,
Chitkara University, Punjab, India.
E-mail: seemag.gulati@gmail.com; seema.g@chitkara.edu.in

Kalpna Guleria
Chitkara University Institute of Engineering and Technology,
Chitkara University, Punjab, India.
Corresponding author: guleria.kalpna@gmail.com; kalpna@chitkara.edu.in

Nitin Goyal
Department of Computer Science and Engineering,
School of Engineering and Technology,
Central University of Haryana, Mahendragarh, Haryana, India.
E-mail: dr.nitingoyal30@gmail.com; dr.nitin@cuh.ac.in

(Received May 27, 2024; Revised on July 18, 2024; Accepted on September 3, 2024)

Abstract
Ocular diseases significantly impact the health of the public globally. According to the World Health Organization (WHO) reports, at least 1 billion people suffer from near or distance vision impairment that could have been prevented or has yet to be addressed. These conditions cause difficulty in living a healthy lifestyle and impair individual quality of life. The article explores the application of federated learning in detecting two vision-threatening ocular diseases- diabetic retinopathy and diabetic macular edema. A federated learning framework enhances the technological capabilities of artificial intelligence that leverages decentralised data sources without creating data banks to maintain privacy. The methodology implements a federated learning environment with 2, 3, and 4 clients, using MobileNetV2 as the backbone deep learning model. The model is trained on a composite of 2 datasets procured from the Kaggle repository, comprising coloured fundus images labelled for diabetic retinopathy, diabetic macular edema, and normal cases. The federated learning process involves training at the client end to build client models called local models. The clients in a federated learning system only share updates regarding their local models. The original data is never shared with a central server. The server integrates these local models into the central global models using aggregation strategies such as FedAvg, FedProx, etc. Performance metrics, including prediction accuracy, class-wise accuracy, precision, recall, and F1 score, are calculated across 30 communication rounds. The results demonstrate that the federated learning model achieves an average prediction accuracy of 96%, and a class-wise accuracy of 100% in detecting diabetic macular edema and diabetic retinopathy. The high performance of the federated learning system highlights the significance of federated learning as a viable solution for ocular disease detection while ensuring data privacy.

Keywords- Machine learning, Preserving privacy, Federated deep learning, Ocular disease, MobileNetV2.

1. Introduction
Permanent loss of vision and impairment might occur from various causes, such as cataracts, macular edema, diabetic retinopathy, glaucoma, and other infectious diseases of the eye, etc. Although, both children and adults of all ages can be impacted, but most of these conditions affect individuals over the age of 50. Early occurrence of severe vision impairment in young children can lead to lower educational achievements and reduced career opportunities throughout life (Al-Fahdawi et al., 2024). Preventative strategies are crucial in reducing the impact of ocular diseases. Preventive strategies include modifying a sedentary lifestyle with increased outdoor activities, to reduce the risk of myopia in children. Another example could

be, managing systemic diseases like diabetes and hypertension which further cause conditions such as diabetic retinopathy. Even though cost-effective solutions such as spectacles for vision correction and surgery for cataracts are available, there remains a significant gap in meeting the demand for eye-care services globally. For critical eye conditions, such as diabetic retinopathy and glaucoma, continuous monitoring and corrective treatment are necessary to control disease progression and prevent permanent vision loss (Tang et al., 2024).

The challenges associated with such eye conditions are amplified by the shift in lifestyle habits with more indoor activity and deskbound work culture, leading to a new range of diseases. Thus, the situation demands a prompt adjustment of responses and strategies to counter these challenges effectively. Many efforts are required to reduce the occurrence and further progression of eye diseases by studying the disease patterns and their causes. Regular eye checkups, medical care and treatment, and effective public health policies are vital in accommodating the needs of the citizens. A holistic approach is imperative in addressing the threats posed by ocular diseases (Davidson et al., 2007). It should include preventive measures, timely prognosis, and effective treatment, as well as rehabilitation if the situation demands support for those with irreversible vision impairment. Such efforts will prove to be fruitful not just for improving the quality of life of millions of people worldwide but also for alleviating the broader public health. Thus, decreasing the negative economic impacts of vision impairment and blindness (Nguyen et al., 2022).

Ocular diseases such as diabetic retinopathy and diabetic macular edema are a few of the leading causes of permanent vision globally. Early detection through screening is imperative for the prevention of irreversible vision loss. The development of ML in ocular disease detection has come a long way and holds a lot of potential for enhancement of diagnostic accuracy and efficiency for various conditions (Gulati et al., 2022a; Sanghavi and Kurhekar, 2024). Machine learning (ML), a branch of artificial intelligence, focuses on creating algorithms capable of learning from data to make predictions and decisions without explicit programming. This is particularly beneficial in ophthalmology, where diagnosing diseases relies heavily on interpreting complex imaging data. (Gulati et al., 2023b; Thotad et al., 2023) Traditional ML methods show great potential in detecting ocular disease. However, there is a challenge to data privacy and accessibility.

One of the main applications of ML in this field is analysing retinal images. Many ML models have been developed to detect diabetic retinopathy that utilise large datasets of retinal images and build models which are capable of accurately identifying different levels of the disease severity. These systems comprise sub-networks that are designed for special tasks such as image quality assessment, lesion detection, and disease grading (Gulati et al., 2023a, 2023c). Glaucoma detection is another area where ML has shown good performance. The models process retinal images to identify glaucoma indicators, such as variation in the optic disc and cup, using image segmentation techniques to focus on important diagnostic areas. These ML methods not only support early detection but also reduce dependence on manual feature extraction by clinicians, thereby streamlining the diagnostic process (Bajwa et al., 2019; Safi et al., 2018).

The advantage of ML in ocular disease detection lies in its ability to efficiently process huge sets of data and recognize patterns not easily visible to humans. This proves fruitful in improving diagnostic accuracy and speed, leading to earlier medical interventions and better patient outcomes. Additionally, the models continuously improve as they are exposed to more data. Thus, revealing new insights into disease detection mechanisms and indicators (Alharbi, 2024; Gulati et al., 2022b; Sharma et al., 2023). However, deploying ML in clinical settings faces challenges, including the need for large, annotated datasets for model training and validation, and ensuring models perform well across different patient populations and imaging equipment. Despite these challenges, the ongoing advancement of ML in ophthalmic disease detection is promising for more personalized, efficient, and accurate eye care (Aamir et al., 2020; Kaur et al., 2022).

Federated learning is a decentralized approach to ML that offers a solution by training algorithms across different client devices that have their local subsets. These clients train on these local datasets without exchanging their data samples. Federated learning has emerged as a revolutionary learning approach that facilitates the creation of models on various distributed devices. Each of these devices has their dataset and these devices work independently of each other. There is no exchange of the source data between the clients and the server, only model updates are communicated between them. This approach is particularly suited to the healthcare sector for many reasons, as given below (Hartmann, 2018; Srivastav et al., 2023).

***Privacy preservation*:** Countries like the United States and Europe control the data generated by the healthcare industry with their strict laws like the Health Insurance Portability and Accountability Act and the General Data Protection Regulation, respectively. These laws help in the protection of sensitive patient data, at the same time making it a tough job to utilize the data for research. In particular, for the automation of disease diagnosis ML and DL models require lots of data. Although the data is available in many medical organizations but it cannot be shared due to privacy concerns (Sharma et al., 2021). A practical solution to this problem emerged in the form of a collaborative framework popularly known as federated learning (FL). FL enables the decentralization of the learning process by transporting the learning algorithm to the source site containing the data, instead of creating a pool of data at a central location. This implies that the sensitive patient information is not transported and remains in the originating location. In simpler words, the sharing of patient data does not occur, but the model updates are shared to form a collaboratively learned efficient model. The model updates are the weight gradients that are sent to the central location for the aggregation of the client models to upgrade the global model. The process reduces the potential risk involved in sharing the personal health information of the patients and aligns with the local privacy laws to ensure that the confidentiality of the patients is maintained (Peyvandi et al., 2022).

***Data silos*:** The data in the healthcare domain is generally fragmented into silos, and each of the institutions holds its dataset. The data silos hamper data sharing among institutions that can collaborate to gain a more comprehensive and cross-institutional data analysis. FL helps in breaking these barriers by facilitating a collaborative approach to learning. The participating organizations contribute to the learning process without needing to share any of their localized data directly with other institutions. The collaborative framework fuels the learning and leverages diverse datasets to amplify the quality and variability of the available data used in model training. FL, with its collaborative framework, has been fruitful in improving the scope of research and opening the door for more inclusive and comprehensive healthcare solutions.

***Improvement in model generalisation*:** The data available in various medical organisations is diverse due to variations in the statistics of the patients, disease presentations, and treatment outcomes. FL takes advantage of the variability of data for training efficient models which are robust and can generalise on different populations. This capability of generalising is significant in applications related to the healthcare industry since the effectiveness of diagnostic tools and treatments can significantly differ among different groups. Thus, the models trained with the FL approach are exposed to a broader spectrum of data characteristics, leading to improved accuracy and reliability in real-world applications. Thus, ensuring that the diagnostic and the predictive models are more inclusive and reflective of the global population.

***Reduction in the data transfer costs and latency*:** In a traditional learning approach all the data which is usually large, is integrated at a central location to form a data bank. This process is quite expensive in terms of communication costs and takes a long period to transfer. The size of medical data for disease identification is essentially growing with the advancements in technology (Sharma et al., 2022). FL mitigates this issue by executing the learning process locally on the machines holding the data. The exchange is only of the model weights, parameters and gradients which is much smaller when compared with the huge amounts of data. The federated framework decreases the bandwidth requirements and significantly reduces the communication cost incurred by transferring raw data. Furthermore, by eliminating the need to transmit large datasets, FL reduces latency in training and updating. Thus, facilitating more efficient and scalable ML projects (Adnan et al., 2022).

Real-time learning**:** The medical environment is continuously advancing, with new disease strains and treatment options emerging regularly. FL approach supports the real-time updating of models with new data as and when it becomes available. This means as and when new patient records are added they can start contributing to improving the prediction accuracy of models. This characteristic feature of FL is particularly beneficial while responding to fast-moving health crises, such as the COVID-19 pandemic. In such situations, timely data analysis and model adaptation are important in improving public health responses. FL is thus a dynamic and responsive approach to healthcare modeling. Thus, ensuring that medical professionals utilise the most recent and relevant tools for diagnosis and treatment planning (Kaur et al., 2023).

The key difference between a traditional centralized ML approach and a collaborative federated learning framework is the location of the data samples used in the learning process. If the data samples are pooled at a central location, then the learning scenario is the traditional centralized scenario and if the data samples are not pooled and they reside on the originating site then it is the collaborative federated learning framework (Pfitzner et al., 2021; Saini et al., 2023). This difference is highlighted in **Figures 1** and **2** given below with the diagrammatic representation of a centralized traditional ML scenario in **Figure 1** and the pictorial representation of a collaborative decentralized federated learning framework in **Figure 2**.

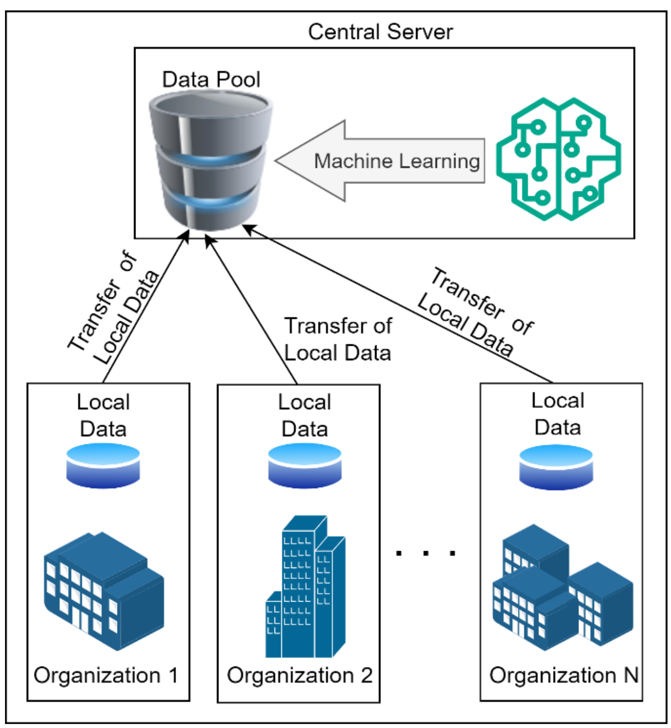

Figure 1. The diagrammatic representation of a traditional centralized machine learning scenario.

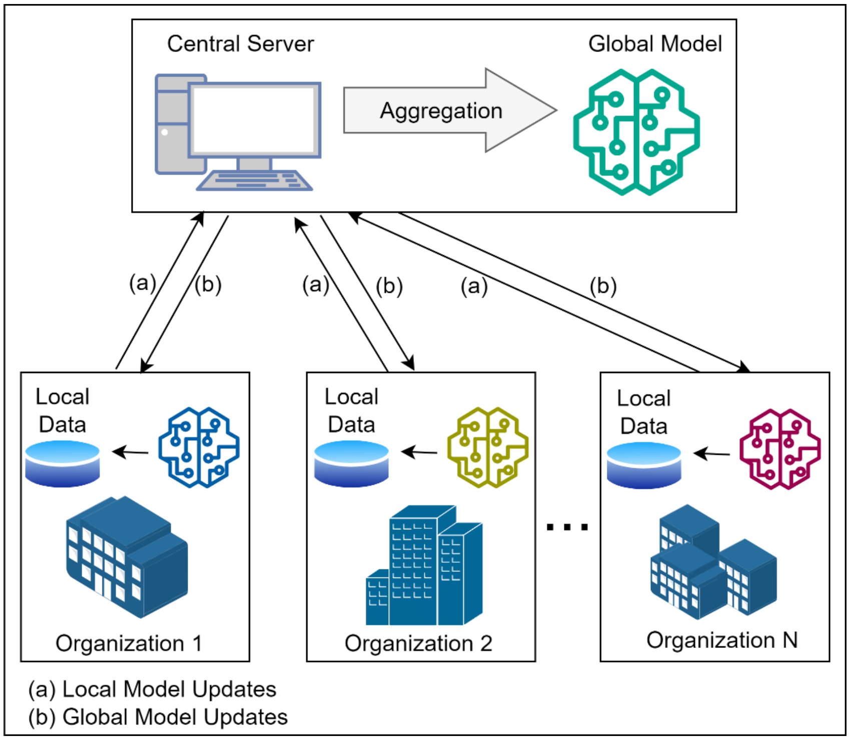

Figure 2. The pictorial representation of a collaborative decentralized federated learning framework.

This article serves as a reference in understanding the privacy-preserving and collaborative machine learning framework called federated learning. Apart from this, the other noteworthy contributions of the article are as follows:

- The article emphasizes the importance of federated learning in maintaining the privacy of healthcare data, particularly in relation to vision-threatening ocular diseases.
- The article proposes a privacy-preserving collaborative federated learning model for detecting ocular diseases: diabetic macular edema and diabetic retinopathy using MobileNetV2-based architecture, FedAvg aggregation strategy.
- The proposed federated learning model has been implemented for 2, 3, and 4 client architectures and the results exhibit that the proposed privacy-preserved collaborative federated learning model shows an average prediction accuracy of 96% and class-wise accuracy of 100% in detecting diabetic macular edema and diabetic retinopathy.

The rest of the article structure includes the following sections: the second section discusses recent research on ocular disease detection using federated learning. Next, the third section describes the materials used in the proposed work, that is the dataset used in the study along with all the parameters, the methodology and the proposed methodology. The fourth section analyses the results obtained by implementing federated learning in detecting diabetic retinopathy and diabetic macular edema.

2. Related Work

This section details the recent research in ocular disease detection using a federated learning framework for protecting the privacy of sensitive user data and collaborative training for improved results.

The paper (Hossain et al., 2023) introduces a federated learning framework designed for predicting diabetic retinopathy without sharing the patient data among various healthcare organisations. It utilises differentially private federated learning with machine learning models like AlexNet, ResNet50, SqueezeNet1.1, and VGG16 applied at the back end. These are tested on a dataset comprising 35,120 retinal images classified into five categories. This approach incorporates a noise addition strategy to ensure privacy and checkpoint techniques to reduce communication overhead. ResNet50 emerged as the most effective model for the application, with an accuracy of 83.05% without noise application and 79.35% with a noise multiplier of 8.0.

The authors (Matta et al., 2023) explore Federated Learning (FL) for diabetic retinopathy detection within a multi-center fundus screening network. They have utilised the OPHDIAT fundus photograph dataset, which includes nearly 700,000 images. Two FL algorithms, cross-center and cross-grader, were developed and compared with centralized learning. Both the algorithms used EfficientNetB5 as the underlying DL model. Conclusively, the FL models demonstrated similar performance to traditional centralised learning, with Area Under the Curve (AUC) scores of 0.9317 for the first cross-center algorithm and 0.9522 for the cross-grade algorithm, versus 0.9482 for the centralised deep learning model. Thus, showcasing FL's potential in medical imaging to maintain data privacy while still achieving competitive accuracy.

In another research work (Nasajpour et al., 2022) the authors have focused on federated transfer learning for diabetic retinopathy detection using CNN architectures. In particular, the backbone CNN-based model used in the work is AlexNet. The authors have assessed three models- the standard transfer deep learning model, the model with federated averaging (FedAvg) aggregation scheme, and the model with federated proximal (FedProx) aggregation technique. The models have been implemented across five publicly available diabetic retinopathy datasets. These models achieved accuracies of 92.19%, 90.07%, and 85.81%, respectively, showcasing the efficacy of federated transfer learning algorithms in diabetic retinopathy detection alongside the advantage of preserving data privacy. Thus, elaborating the need for federated learning in practical applications of medical data.

An IoT-based Federated Learning model for classifying Hypertensive Retinopathy lesions through regional feature fusion was discussed by Soni et al. (2023). It employs two neural networks for classifying arterial and venous nicking (AVN) and hypertensive retinopathy, using pre-processed fundus images to enhance classification accuracy. Various DL models used foundationally in the FL system are IterNet, U-Net and SeqNet. The model was tested on an experimental dataset initially containing 48 IoT-FHR fundus images, expanded to 1276 images through data augmentation. They achieved an accuracy of 95.14%, a sensitivity of 74.98%, and a specificity of 97.54%. The authors have demonstrated the potential of combining local lesion characteristics with global classification models to improve diagnostic accuracy.

A federated learning framework for retinal microvasculature segmentation and diabetic retinopathy classification using optical tomographic images was proposed in the article (Lo et al., 2021). The approach involves training models across multiple clients without sharing sensitive data. Also, comparing the performance of federated learning with centralised learning methods. The dataset comprised 153 OCTA images for segmentation and 700 eyes for diabetic retinopathy classification. The architecture utilised for microvasculature segmentation is the residual U-Net model. The key performance metrics included accuracy, dice similarity coefficient (DSC), and area under the receiver operating characteristic curve (AUROC). They have demonstrated that federated learning shows performance comparable to internal models. Thus, proving its value in medical image analysis while safeguarding data privacy which is at the epicentre of a federated learning framework. The article emphasises federated learning's potential, coupled with domain adaptation, for enhancing both privacy and accuracy in medical image analysis, particularly in ocular disease recognition.

The articles relating to ocular disease detection using federated learning have been summarised in tabular form in **Table 1**.

Table 1. Summary of the articles related to the current work.

Reference	Methodology/Models/ Summary	Dataset	Performance outcomes	Future scope and research gaps
Hossain et al. (2023)	A differentially private federated learning system was introduced for grading Diabetic Retinopathy into various levels. Four DL models: ResNet50, AlexNet, SqueezeNet1.1 and VGG16 were used and compared for performance. ResNet50 exhibited superior performance.	The dataset consists of 35126 retinal images from the Kaggle repository.	Best accuracy was achieved by ResNet50. The accuracy score was 83.05% with noise and 79.35% accuracy without noise.	Additional deep learning models can be explored to further improve the prediction accuracy for diabetic retinopathy diagnosis. Even more advanced differential privacy techniques could be implemented to enhance the privacy-preserving capabilities of the federated learning framework.
Matta et al. (2023)	Two FL-based algorithms were developed and evaluated. The cross-center FL and cross-grader FL for distributing data among screening centres and graders respectively. EfficientNet-B5 is used as the backbone DL model for all experiments in the article.	The dataset was collected from the OPHDIAT, a French multi-center screening network and there was a total of 697,275 fundus photographs.	Performance indicators include the area under the ROC curve (AUC) with the outcomes revealing an AUC of 0.9482 for centralized learning (CL), 0.9317 for cross-center FL, and 0.9522 for cross-grader FL, showcasing comparable performance.	The impact of varying the hyperparameters is not given in the article. A discussion on the computational resources required for implementing FL on such a large-scale dataset has not been given.
Nasajpour et al. (2022)	The authors proposed an FL-based algorithm for detecting DR from coloured fundus images. These implementations of centralised learning, FL with the FedAvg algorithm and FL with the FedProx Algorithm have been compared. A CNN-based AlexNet architecture is used for implementing FL.	The datasets used for Diabetic Retinopathy detection include EyePACS, Messidor, IDRID, APTOS, and the University of Auckland (UoA). A total of 2787 images were used in the experiment.	Accuracy, Sensitivity, Specificity, and Precision were used as performance indicators. The comparison based on accuracy reveals that the performance of FL-based models is comparable to that of centralised transfer learning models. The scores are 92.19%, 90.07%, and 85.81% respectively for centralised learning, FL with FedAvg and FL with FedProx.	Augmentation techniques can be applied for further improvement of the model and to increase privacy differential privacy can also be explored.
Soni et al. (2023)	The focus is on classifying the hypertensive retinopathy lesions using DL models in the federated framework. Various DL models used foundationally in the FL system are IterNet, U-Net and SeqNet. These were used to propose a new method for classifying lesions of hypertensive retinopathy.	The dataset is procured locally from a hospital and contains a set of 1276 fundus images for classification purposes.	Performance indicators such as accuracy metrics were utilised to assess the model's effectiveness, showcasing improved classification accuracy and addressing overfitting concerns. The fusion of global and local components in the model resulted in an accuracy of 95.14%, sensitivity of 74.98%, and specificity of 97.54%, showcasing the enhanced performance and classification capabilities of the proposed method.	A small and noisy dataset was used in the study, and larger and more diverse datasets should be explored for model generalizability. The work used only ResNet50 in the experiment. So, the other pre-trained models should also be used for the classification.

Table 1 continued…

Lo et al. (2021)	The study aimed to assess a federated learning framework for retinal microvasculature segmentation and Referable Diabetic Retinopathy (RDR) classification using OCT and OCTA imaging. The architecture utilized for microvasculature segmentation is the residual U-Net model.	The dataset comprises 153 OCTA (Optical Coherence Tomography Angiography) for microvasculature segmentation and 700 eyes with OCTA en-face images and structural OCT projections acquired from 2 commercial OCT systems for Referable Diabetic Retinopathy (RDR) classification. So, a total of 853 images were accumulated. Up-sampling and down-sampling techniques were also applied for better generalisability.	For microvasculature segmentation, the federated learning model achieved an average Dice similarity coefficient (DSC) of 0.793 across all test sets, which is comparable to the centralized models that had an average DSC of 0.807. In the classification of Referable Diabetic Retinopathy (RDR), the federated learning approach achieved mean AUROCs (Area Under the Receiver Operating Characteristic Curve) of 0.954 and 0.960, while the internal models reached mean AUROCs of 0.956 and 0.973, indicating similar performance.	Further studies can investigate how data distribution and imbalanced datasets affect the performance of federated learning, especially in scenarios with a large number of images or diverse data sources. The dataset is quite small, which might lead to issues in the generalizability of the model.

3. Materials and Methodology

This section describes the federated learning framework used to detect diabetic retinopathy and diabetic macular edema from a set of images of normal, diabetic retinopathy or diabetic macular edema coloured fundus images. The section also includes the data selection methodology, preprocessing methods, the architecture of the neural network model, and training procedures.

3.1 Dataset Description

The dataset used for detecting diabetic retinopathy and diabetic macular edema has been procured and assembled from two datasets. These datasets are available publicly on the Kaggle repository for machine learning enthusiasts. The Kaggle repository contains many datasets on ocular diseases from which two have been taken. The first one is the Indian Diabetic Retinopathy dataset which contains coloured fundus images of patients' left and right eyes. These images are labelled into three types: diabetic retinopathy, diabetic macular edema and normal. The second is the Aptos blindness detection 2019 dataset from which samples of diabetic retinopathy and normal images of ocular diseases have been taken and pooled. The dataset thus formed contains a total of 838 images, as shown in **Table 2** below. 278 images are of the diabetic macular edema class and 280 images each of diabetic retinopathy and normal class. The images shown in **Figure 3** are coloured fundus images from the dataset. Referring to **Figure 3** the first image shows diabetic macular edema and the second image is of an eye with diabetic retinopathy. The last image in **Figure 3** is of a normal eye.

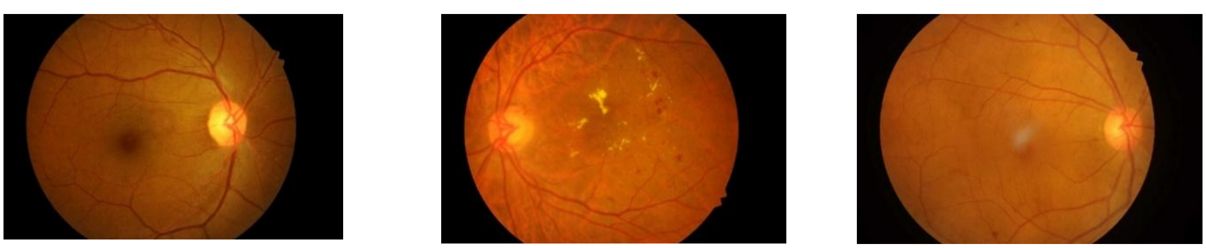

Figure 3. The example images from the dataset used and the classes from which the image belongs are Diabetic Macular Edema, Diabetic Retinopathy and Normal respectively.

Table 2. The dataset used in the implementation of the federated learning model.

Class number	Class name	Images
Class 0	Diabetic Macular Edema	278
Class 1	Diabetic Retinopathy	280
Class 2	Normal	280

3.2 Methodology

The collaborative federated learning framework implemented is an iterative process for building a robust global model (central model). The collaboration is handled by the central server and the clients contribute to the central model by building local models. The participating clients hold the local data securely and do not share any of their data. The server sends the initial weights and parameters of the global model to the participating clients. The clients then with the global weights and parameters apply training on their data locally to update the weights at their location. The models thus formed at the client end are known as the local client models. After several epochs, the client models stabilise and after stabilising the client communicates the local model updates to the central server. At the server end when the client models in the form of gradients are received from clients, the local model weights are incorporated in the global model weights to improve the global model. The server aggregates the weights and gradients according to the aggregation strategy implemented, such as FedAvg, FedProx, MOON, etc. After the aggregation step, the updated global model weights are again sent to the clients to build new client models with updated weights. The process is iterated several times until the model converges or stops showing any improvements. These iterations are known as the communication rounds. The flowchart in **Figure 4** depicts this iterative process of federated learning.

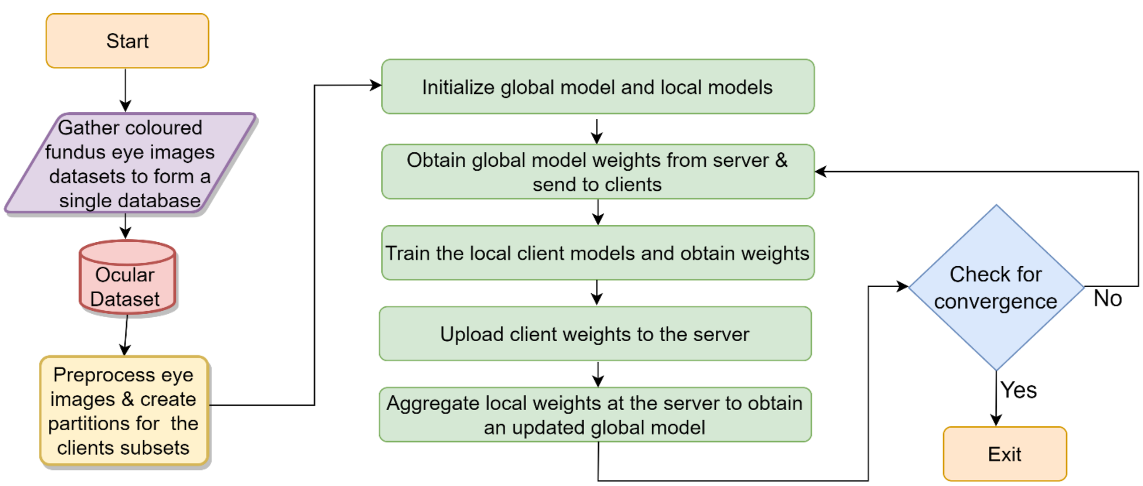

Figure 4. Flowchart for the federated learning methodology.

The complete process of a federated learning scenario involves basic 4 steps:

a) ***Dataset preparation*****:** After procuring the dataset from the open-source repositories the data needs to be processed before the actual learning takes place. In this step, various transformations are applied to the raw data. For detecting diabetic macular edema and diabetic retinopathy, the dataset is subjected to augmentation techniques to improve the learning process. The augmentation techniques applied are cropping, flipping and rotation. The basic purpose is to enhance the data in terms of quantity and diversity of the data to improve model generalisation. After the dataset is augmented the data needs to be divided into training and testing subsets for each client, according to the number of clients (Ng et al., 2021).

b) *Server learning*: The next step is to initialise the central model at the server and obtain weights. In the proposed work MobileNetV2 architecture is used as the backbone model of the FL system. The server uses pre-trained MobileNetV2 weights to initialise the global model and sends them to the clients for further steps. The server is responsible for the practical implementation of the distributed FL framework. The roles and responsibilities of the server include selecting the aggregation technique, the communication strategy, and ensuring communication privacy.

c) *Client learning*: The term client refers to the local sites which participate in the collaborative learning process. The devices at the client end store the actual local data on which the learning is applied. In practical scenarios, these are real-world datasets of different organisations where machine learning is applied to design automated systems which assist medical practitioners. The clients, after receiving the weights and parameters of the central model initialise their local models by using the same parameters and weights. The local models are then trained over the local dataset and update the gradients according to their datasets. This model updates or gradients are then transported to the server (Banabilah et al., 2022; ur Rehman and Gaber, 2021).

d) *Aggregation*: Aggregation is the process of integrating the client models into the central model. The server is responsible for the selection of clients and the aggregation technique implemented. The server shall aggregate the weights received from the various clients based on the aggregation strategy chosen and over a series of communication rounds and the server aggregation central model converges. At the server end, it is also decided which clients will participate in the next communication round and which client model should be aggregated in the current round, again for obtaining the optimal central model. At the server policies for selecting the participating clients and the aggregations are framed and implemented for improving the efficacy of the central model (Qi et al., 2024).

In this research work, a federated learning framework with 2, 3 and 4 clients, has been implemented for the detection of two ocular diseases diabetic macular edema and diabetic retinopathy. The root cause of both these diseases is diabetic mellitus commonly known as diabetes. These diseases affect about one-fourth of the people suffering from diabetes for a prolonged period. Both these diseases can cause complete vision loss but they have some key differentiating factors. Diabetic retinopathy is a complication caused in the retina of the eye and diabetic macular edema affects the macula of the eyes. The detection system presented here is based on a federated learning environment with horizontal federated learning.

Federated learning is classified into three types of learning based on the feature space and the sample space. If the feature space is different and the sample space is the same, it is called horizontal federated learning. At times, horizontal federated learning is also referred to as homogeneous federated learning and sample-based federated learning. In horizontal federated learning, the dataset and the models are distributed on different nodes while keeping the privacy of the data in mind. The steps involved in a federated learning framework are given in **Figure 5** below (Li et al., 2020; Zhang et al., 2020).

There are numerous benefits of using a horizontal federated learning framework. First and foremost is enhanced data privacy due to minimal exposure to sensitive information, which is a characteristic feature of any federated learning framework. Thus, reducing the risk of data breaches. In addition, it adheres to local privacy regulations of various participating organisations since the pooling of data is not required.

Next, is the improvement in model proficiency and model generalizability since the training is done on diverse datasets while preserving the privacy of the dataset. Apart from this, it saves bandwidth since data transfer is limited to the gradients and not the complete dataset. It also helps in greater scalability by supporting personalization on large datasets. FL framework lowers the communication cost and the computational overhead compared to traditional centralised learning. Lastly, it supports training on non-identical and independently distributed datasets across multiple parties, applicable in real-life scenarios (Kairouz et al., 2021; Shanmugarasa et al., 2023).

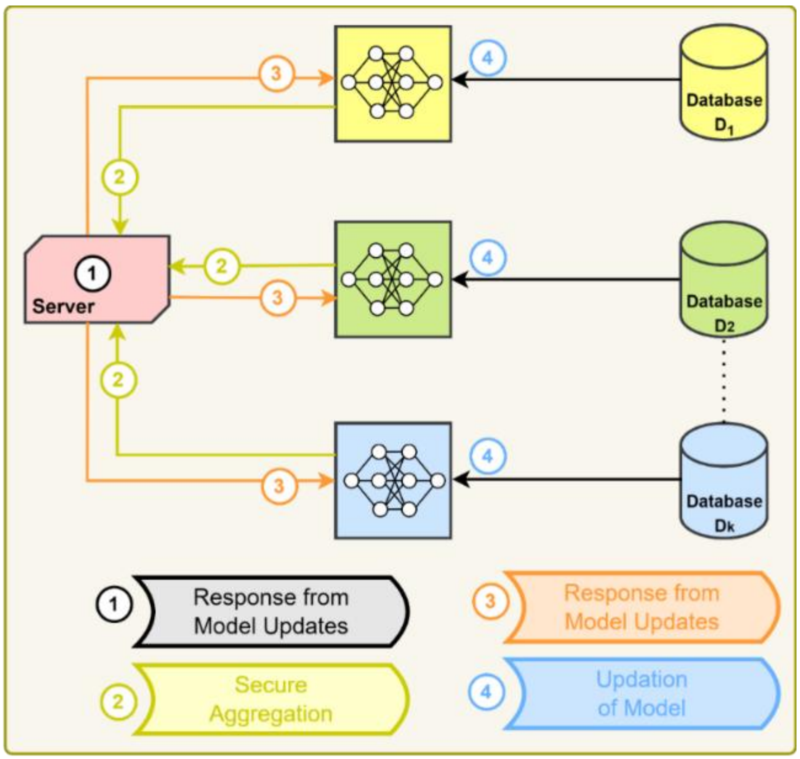

Figure 5. The steps involved in a horizontal federated learning framework.

3.3 Proposed Methodology

The work presented in this article is based on the implementation of horizontal federated learning with MobileNetV2 as the backbone deep learning model. The model uses Adam optimizer for the deep learning model at the foundation since it adapts the learning rate according to the learning parameters and is suitable for large datasets as well. The parameters used in the simulation of the framework are given in a tabular form in **Table 3** below. Image augmentation techniques are applied on the dataset to enhance the dataset and improve model efficiency and generalizability. The augmentation is done using random rotations, horizontal flipping, and vertical flipping.

The aggregation strategy used in the work is a simple averaging scheme called the FedAvg. The performance metrics used to assess the model are prediction accuracy, class-wise accuracy, precision, recall and F1-score. These metrics are calculated after every communication round till the completion of 30 rounds. The test-train ratio of the data is taken as 80:20 for the implementation. The batch size is carefully selected to be 32 and the initial learning rate is kept at 0.001.

Table 3. Hyper-parameters used in the implementation of the federated learning model.

Parameter	Value
Model architecture	MobileNet V2
Clients	2, 3, 4
Learning rate	0.001
Optimizer	Adam
Batch size	32
Epochs	5 (per round)
Communication rounds	30
Aggregation strategy	FedAvg
Augmentation techniques	Yes, Random rotation, Horizontal flipping, Vertical flipping and Normalisation
Train test ratio	80:20
Performance metrics	Accuracy, Class wise - Accuracy, Precision, Recall and F1-score

The diagram shown in **Figure 6**, illustrates the process of a federated learning system that trains deep learning models across multiple clients without sharing any data directly. The flowchart begins by working on the dataset used for detecting ocular diseases. The data which consists of coloured fundus images are first collected from different sources and then good-quality images in terms of colour contrast and brightness, are manually selected for the federated learning process. The next step is to preprocess the selected data, augmentation is done by applying random rotations, horizontal flipping, and vertical flipping. The image dataset thus formed is now ready for the federated deep learning process. Next, the data is split into training and testing subsets and then divided for various clients. The test is to train ratio chosen in the proposed work is 80:20. In the proposed work, the clients are varied with three different implementations of 2, 3 and 4 clients. The client data is assigned using the data loader function with 2, 3 and 4 clients respectively. Thereafter, the main process of federated deep learning starts. The server is initialised with the pre-trained model weights of MobileNet V2. MobileNetV2 is a lightweight deep neural network architecture. The pre-trained model weights are represented by W_g^0 and global model weights are represented by W_g^t as shown in Algorithm 1. The clients then train on the local dataset D_i, to obtain the updated weights $W_{l,i}^{t+1}$ and then clients send their optimised local model weights $W_{l,i}^{t+1}$ and the size of their dataset n_i back to the central server. After receiving these updated client weights, the central server uses a FedAvg aggregator with a function $\frac{1}{N}\Sigma_{\mathrm{i}=1}^{K} n_i W_{l,\mathrm{i}}^{t+1}$ to integrate the client weights. The total number of data points across all clients is N=$\Sigma_{\mathrm{i}=1}^{K} n_i$ where K is the number of clients participating in the training round. The updated global weights after the communication round are denoted as W_g^{t+1}. These updated global model weights are then communicated to the clients for local training. The clients again train on local data D_i form local models which are again sent to the server for aggregation. The process is repeated for many communication rounds until the global model stabilises.

The architectural diagram of the federated learning framework with 2, 3 and 4 clients has been depicted in **Figure 7**. The diagram shows that the clients have their own datasets and computing resources. They perform all computations locally and train models independently of each other without sharing any data or weights. The server acts as a central coordinator, aggregates the updates received from the clients and computes a new global model in each communication round. The server also shoulders the responsibility for initiating the training process by distributing the global model. The aggregator component of the server is responsible for applying the aggregation strategy, in the proposed work FedAvg algorithm has been chosen to average the model updates received from the clients. The server sends the initial global model parameters W_g^t to all clients which is shown with black lines in the diagram. Each client receives the same model parameters at the beginning of training and modifies these weights by training on their local dataset D_i to update the weights $W_{l,i}^{t+1}$. These updated weights are sent by the clients to the server as shown with red lines. The aggregator module at the server with the FedAvg algorithm then averages these weights to form an updated global model with weights $W_{l,i}^{t+1}$. These updated weights are again sent to clients, depicted by blue lines. These steps keep on iterating until the exit condition becomes true.

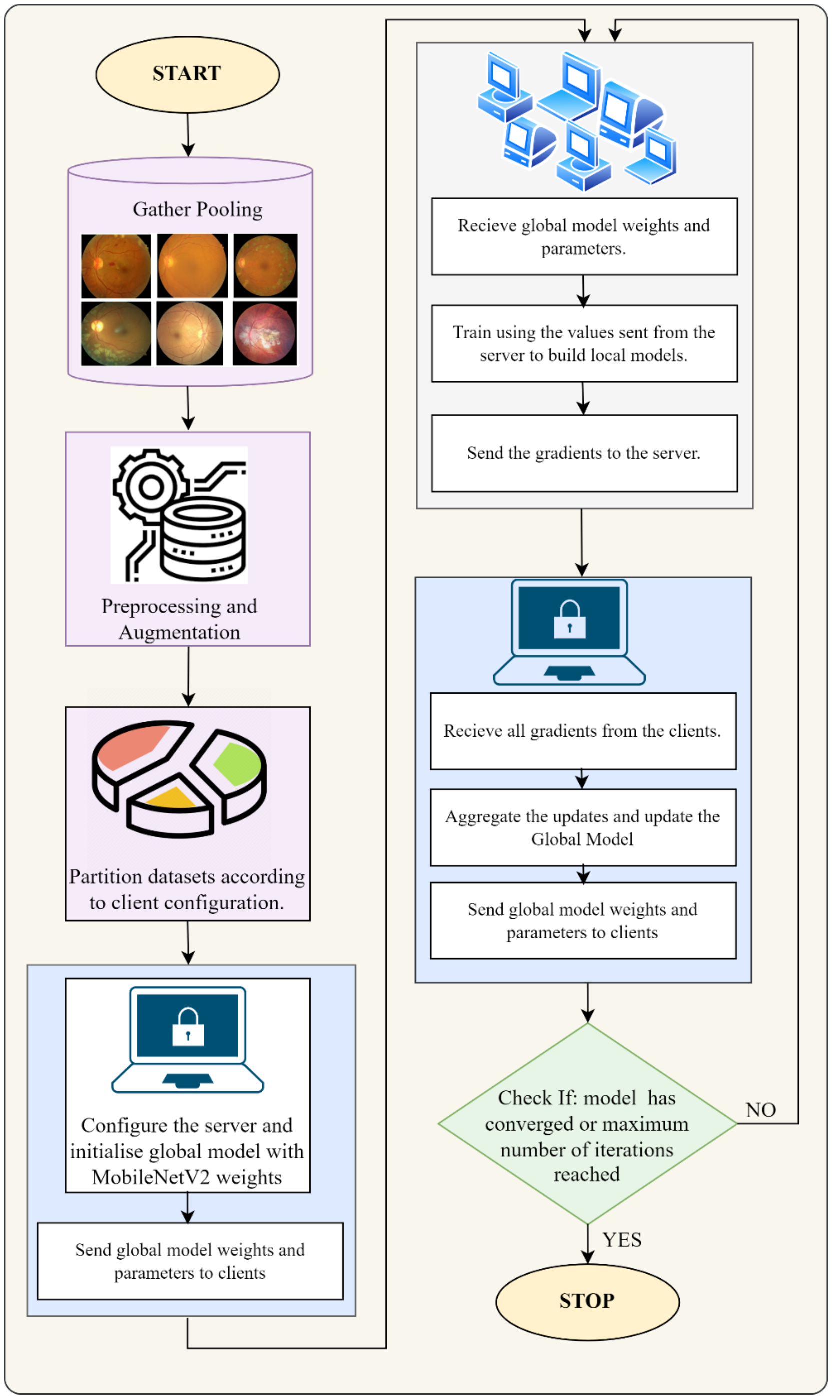

Figure 6. Work flow of horizontal federated learning framework used in the proposed work for detecting ocular diseases.

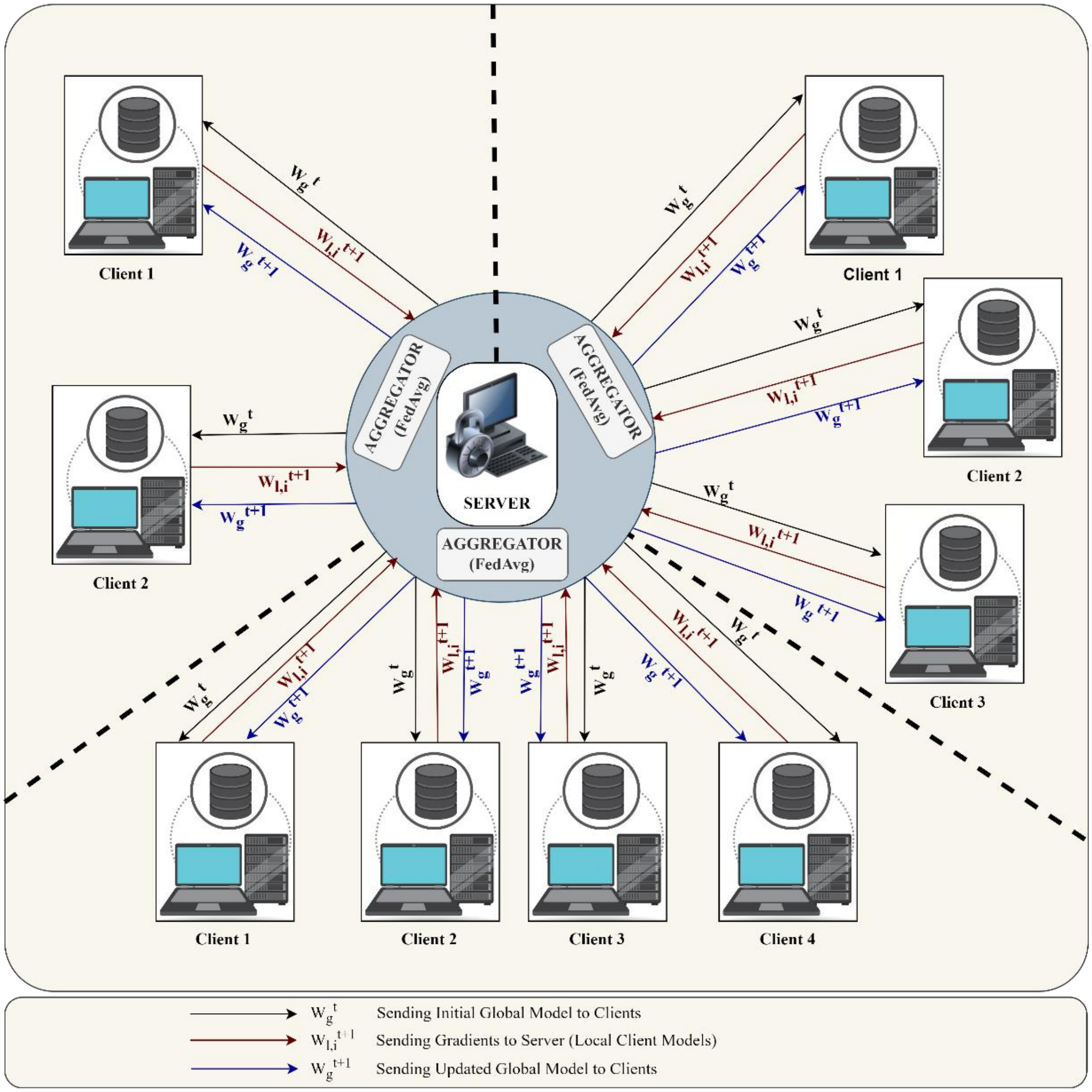

Figure 7. The architectural design of proposed horizontal federated learning framework for ocular disease detection.

The process of horizontal federated learning involves various steps which are iterative and these are shown in algorithm 1. The number of communication rounds is set to 30 for the current implementation.

***Algorithm* 1:** Federated Learning Process for ocular disease-detection
(a) ***Initialization*:** The central server initialises the Global Model with pre-trained weights W_g^0.

$W_g^0 \leftarrow$ Initialize

(b) ***Distribution Phase*:** The central server distributes the Global Model's weights W_g^t to selected client devices. Distribute $(W_g^t) \rightarrow$ Clients
(c) ***Local Training Phase*:** For each client device *i* in parallel, do:

(i) Receive the Global Model's weights W_g^t.

(ii)Perform local training using the client's local dataset D_i with the received weights to obtain the updated weights $W_{l,i}^{t+1}$. $W_{l,i}^{t+1}1 \leftarrow$ Local Train (D_i,W_g^t)

(d) ***Aggregation Phase*** **(FedAvg):**

(i) Each client sends their optimised local model weights $W_{l,i}^{t+1}$ and the size of their dataset n_i back to the central server. Send ($W_{l,i}^{t+1}$, n_i) $\rightarrow$ Server

(ii)The central server aggregates these updates using the Federated Averaging algorithm, where the updates are weighted by the number of data points from each client.

$$W_g^{t+1} \leftarrow \frac{1}{N} \Sigma_{\mathrm{i}=1}^{K} n_i W_{l,\mathrm{i}}^{t+1}$$

where, N=$\Sigma_{\mathrm{i}=1}^{K} n_i$ is the total number of data points across all clients, and K is the number of clients participating in the training round.

i. ***Update global model*****:** The central server updates the Global Model with the aggregated weights. $W_g^{t+1}$is already updated using the FedAvg aggregation formula.

ii. ***Iteration*****:** Repeat steps 2 to 5 iteratively until the Global Model achieves the desired levels of prediction accuracy and precision, and shows no further improvement.

iii. ***Conclusion*****:** Finalize the Global Model W_g^{final} as the output of the federated learning process.

iv. Exit.

4. Results and Discussion

This section analyses the results obtained by following the above-mentioned methodology for implementing a federated learning framework. The results have been arranged into 2 subsections: clientwise results and comparative results. In the clientwise results, the graphs of metrics have been organised first for 2 clients then for 3 clients and at the end for 4 clients. In comparative results, the graphs have been shown to compare the metrics for all three scenarios of 2, 3 and 4 clients.

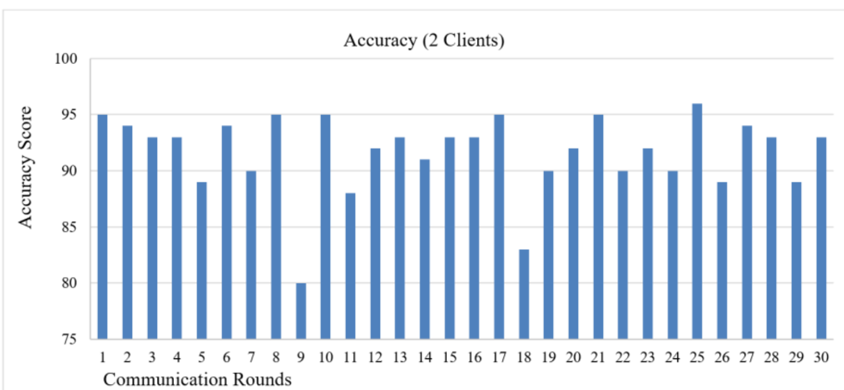

Figure 8. Accuracy score for 2 clients in 30 communication rounds.

The **Figure 8** shows the accuracy score achieved in 30 communication rounds for a 2-client scenario of the federated learning framework implemented in the work. The accuracy scores of the federated learning model show that the model performs well, the lowest accuracy is in the 9th communication round with 80% accuracy and the highest accuracy of 96% is achieved in the 25th communication round. In several communication rounds peaks with 95% accuracy which is the second-highest, are seen, 8, 17 and 21st. These are suggestive of the optimal performance of the model. The graph plotted in **Figure 9** is quite like the accuracy graph. The lowest performance in terms of the precision score of the model is at the 9th communication round with a 0.84 precision score. The best score of 0.96 is in the 8th, 10th and 25th communication rounds, followed by a precision score of .95 in the 6th, 17th and 21st rounds. Referring to **Figure 10** the values plotted in the graph are of the recall score obtained in the 30 communication rounds. The performance in case of recall score is lowest at 0.80 in the 9th communication round and the best recall score is at the 25th communication round with 0.96 magnitude. The second-highest recall score of 0.95 is in 8th, 10th and 21st communication round. The peaks at these rounds are indicative of a consistently good performance of the model.

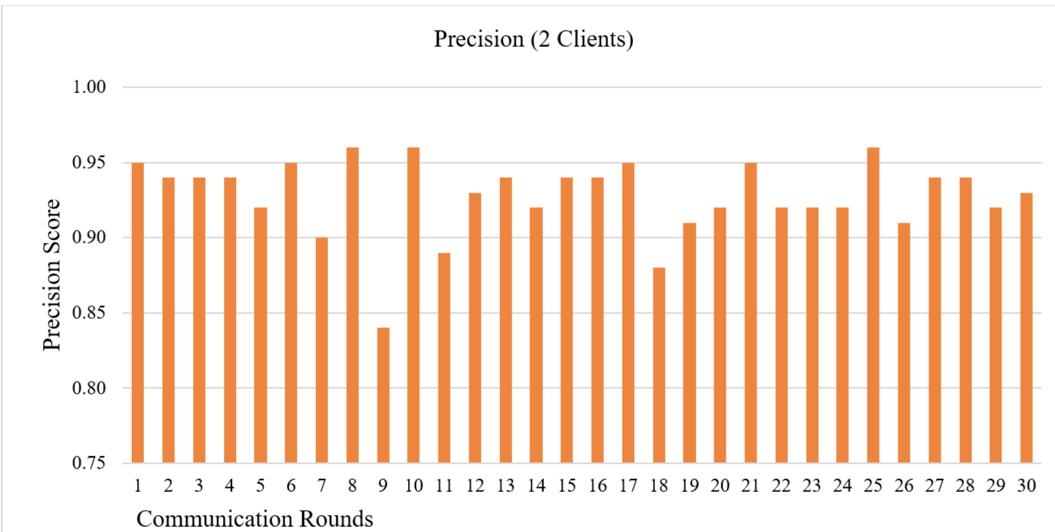

Figure 9. Precision score for 2 clients in 30 communication rounds.

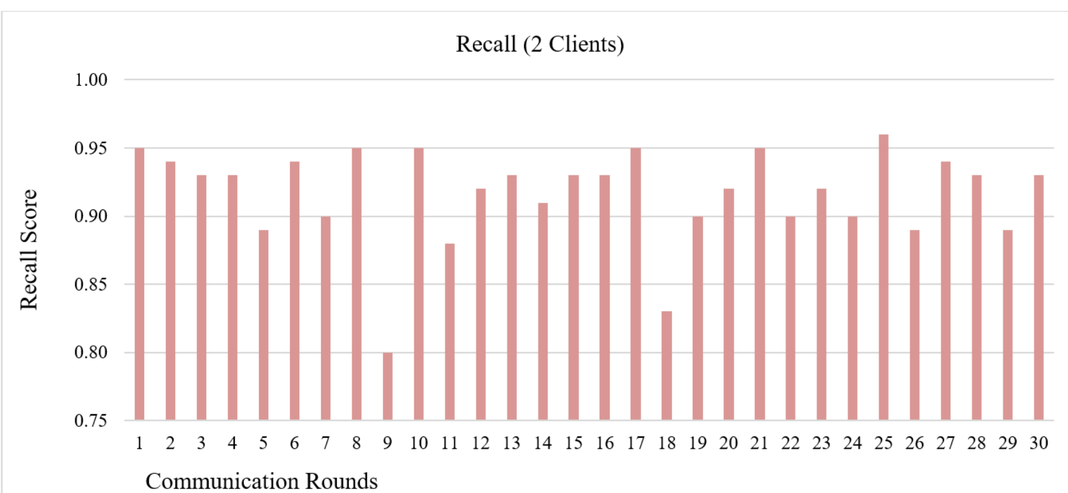

Figure 10. Recall score for 2 clients in 30 communication rounds.

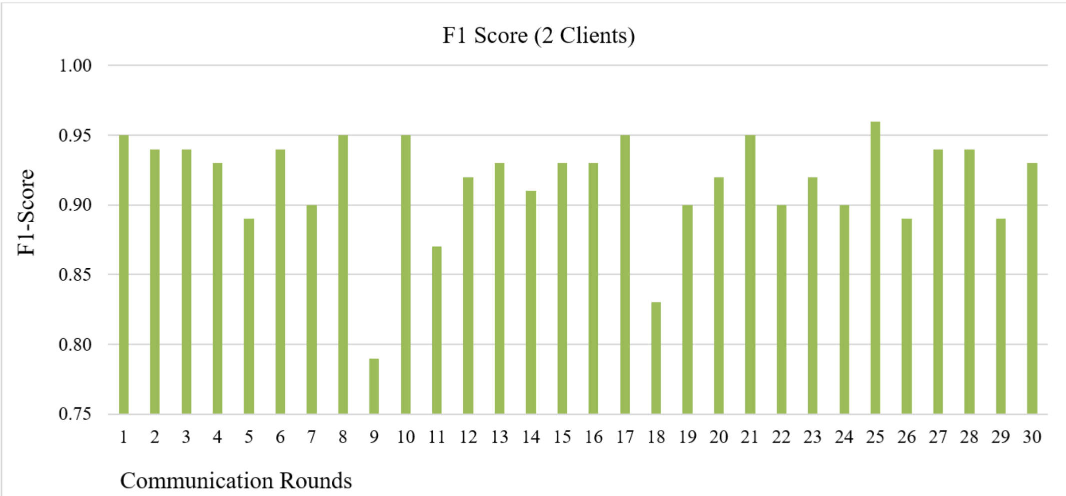

Figure 11. F1-Score for 2 clients in 30 communication rounds.

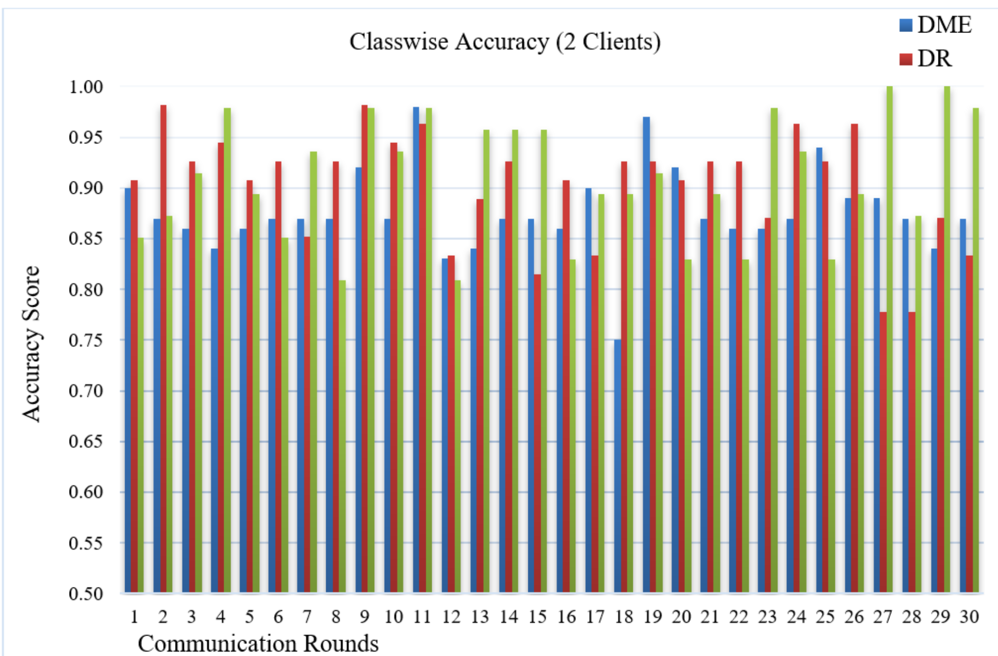

Figure 12. Class-wise accuracy for 2 clients in 30 communication rounds.

The F1-score metric depicted in **Figure 11** for the 2-client privacy-preserving framework; follows the same trend as in the case of the accuracy and recall score. The model performance in terms of F1-score is the lowest in the 9th communication round with a magnitude of 0.79 and the highest is at the 25th communication round with a 0.96 value. The second highest values of 0.95 for the f1-score are obtained in the 8th, 10th, 17th and 21st communication rounds. The graph for class-wise accuracy plotted in **Figure 12** shows that the model performs well for both the DME and DR classes. The best performance in the case of DME is in the 11th round with 98% accuracy followed by 97% in the 19th round and the lowest performance is in the 18th round with an accuracy of 75%. While the model performs excellently for the DR class with 100% accuracy in the 2, 4, 6, 7, 8, 9, 10, 12, 14, 23, 25, 30 communication rounds, the worst performance of 73% is shown in the 18th communication round.

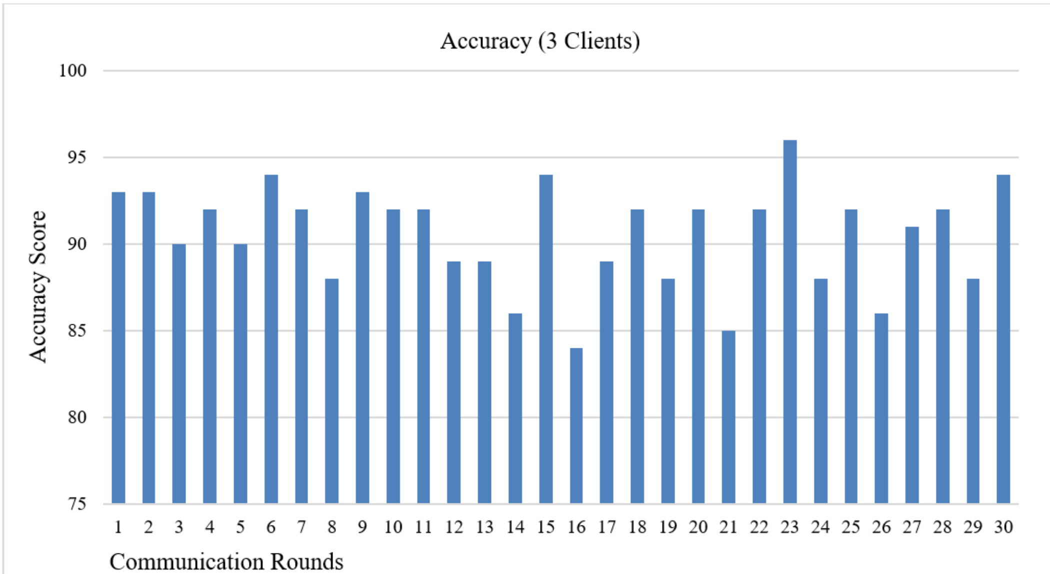

Figure 13. Accuracy score for 3 clients in 30 communication rounds.

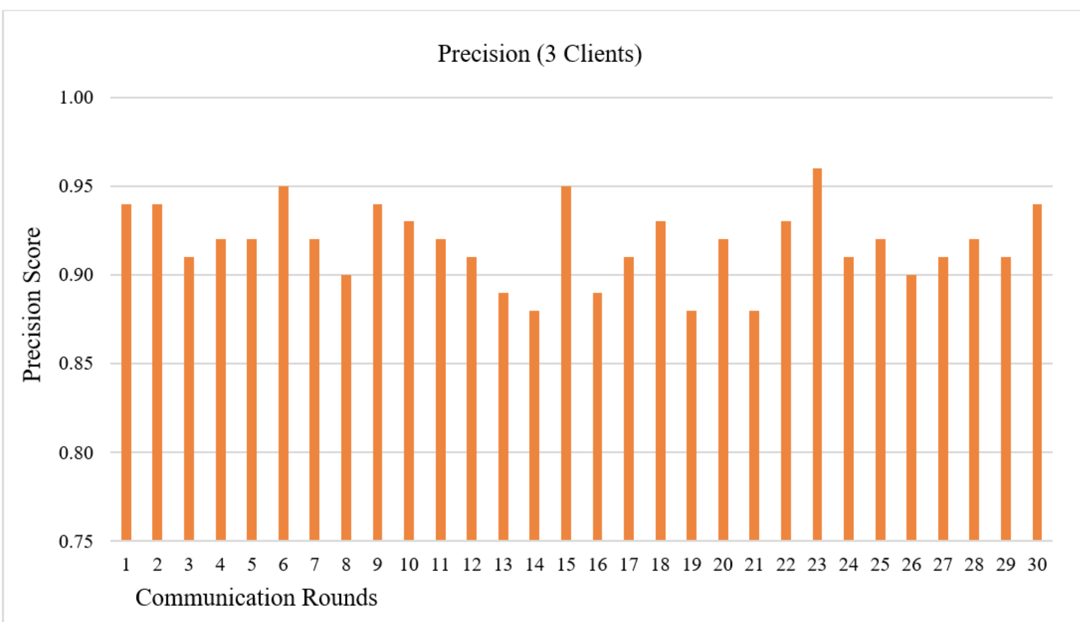

Figure 14. Precision score for 3 clients in 30 communication rounds.

The performance metrics of the federated framework with three clients are shown in the subsequent graphs. The graph for accuracy in **Figure 13** shows that the overall efficiency of the model is quite good with an accuracy score of 96% achieved in the 23rd communication round, which is followed by 94% in the 6th communication round. After the 23rd round, the accuracy starts to decrease. The lowest efficacy of the model is at the 16th round with an accuracy of 84%. The graph for the precision score is given in **Figure 14**, the results are much like the accuracy score. The precision score is at the highest peak in the 23rd communication round with a magnitude of 0.96. The second highest is 0.95 in the 6th and 15th rounds. The lowest score of 0.88 is attained in the 14th, 19th and 21st rounds. Considering, the precision score range is between 0.88 to 0.96 which is a good range. In the next graph recall score obtained over the 30 communication rounds is given as shown in **Figure 15**. The recall score value of 0.96 is highest in the 23rd communication round, which is preceded by 0.94 value in the 6th, 15th and 30th round. The recall score value is the lowest in the 16th communication round with a magnitude of 0.84.

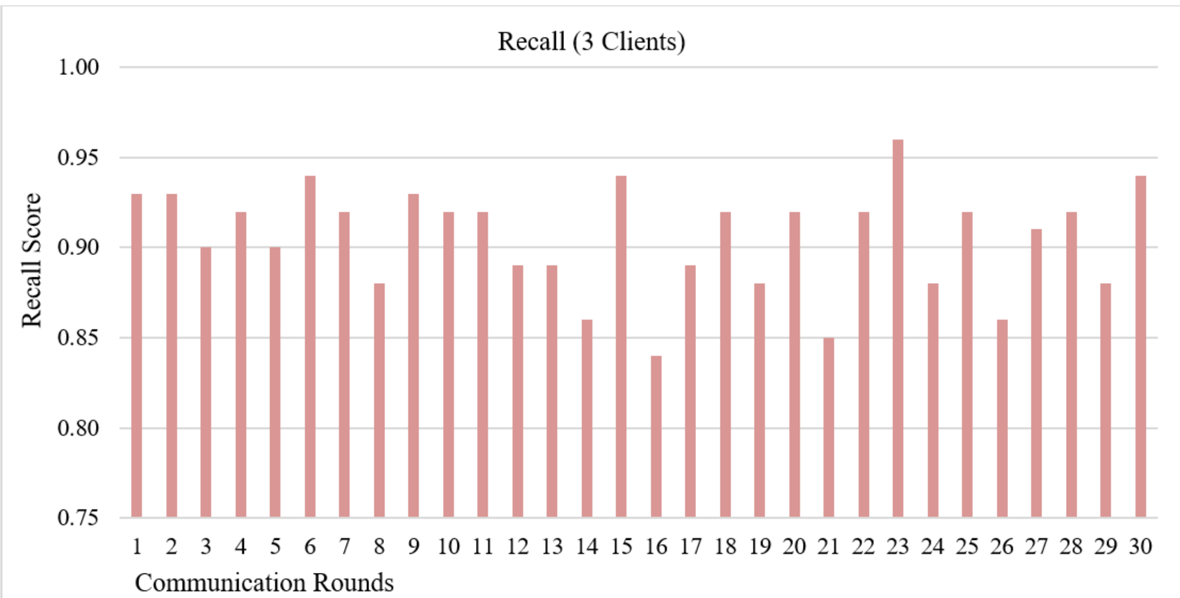

Figure 15. Recall score for 3 clients in 30 communication rounds.

Moving to the next performance metric for the 3-client federated learning framework, referring to the graph of the f1-score given in **Figure 16**, the highest score is obtained again at the 23rd round with 0.96 magnitudes and the second highest value of 0.94 is in the 2, 6, 9, 15, 30 communication rounds. The lowest magnitude is obtained in the 16th communication round. The f1-score shows not much deviation and obtains almost similar throughout. The last metric used for the analysis of the model performance is the class-wise accuracy, depicted in **Figure 17**. There is some variation in the accuracy scored for the DME class with the lowest score of 78% in the 27th and 28th communication round and the highest of 98% in the 2nd and 9th communication round. The second highest accuracy of 96% is scored in the 11th and 24th communication rounds.

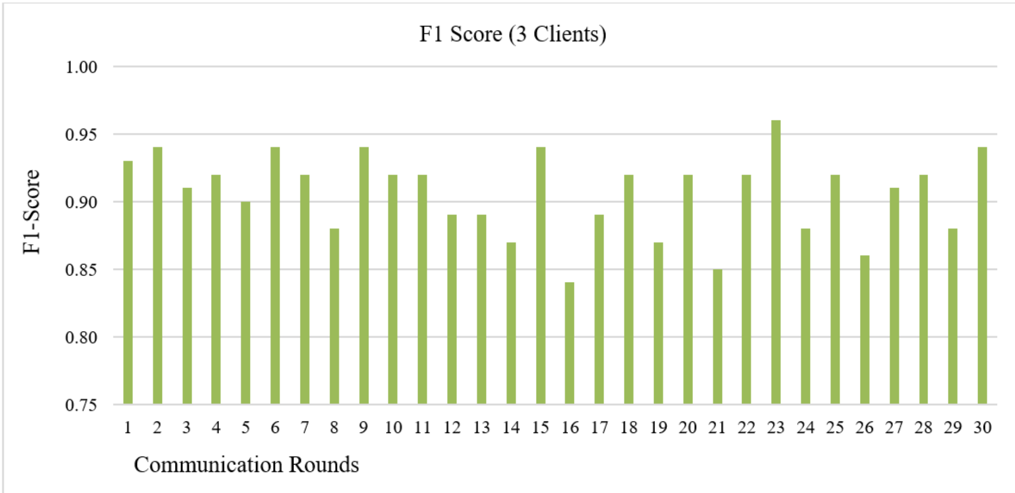

Figure 16. F1-Score for 3 clients in 30 communication rounds.

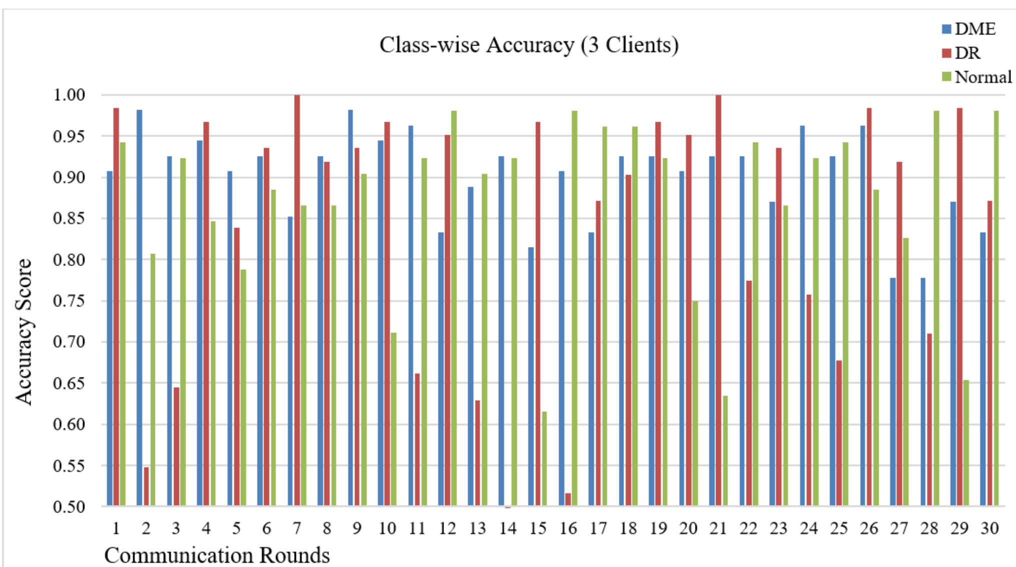

Figure 17. Class-wise accuracy for 3 clients in 30 communication rounds.

Analysing the graphs for the key performance indicators for 4 clients in the privacy-preserving collaborative framework the first indicator is the average accuracy. The average accuracy shown in **Figure 18**, reveals that the model efficacy is the highest in communication rounds 18, 25 and 29 with 93%; the second highest is 92% in communication rounds 3, 4 and 10. The lowest value is in the communication round number 21 with a magnitude of approximately 80%. The precision score values plotted in the shown in **Figure 19** depict similar values obtained. The highest value was 0.94 in the 18th and 25th communication rounds and the second highest value of 0.93 in the 4th and the 10th communication rounds. The lowest value is 0.85 in the 21st and the 30th communication round.

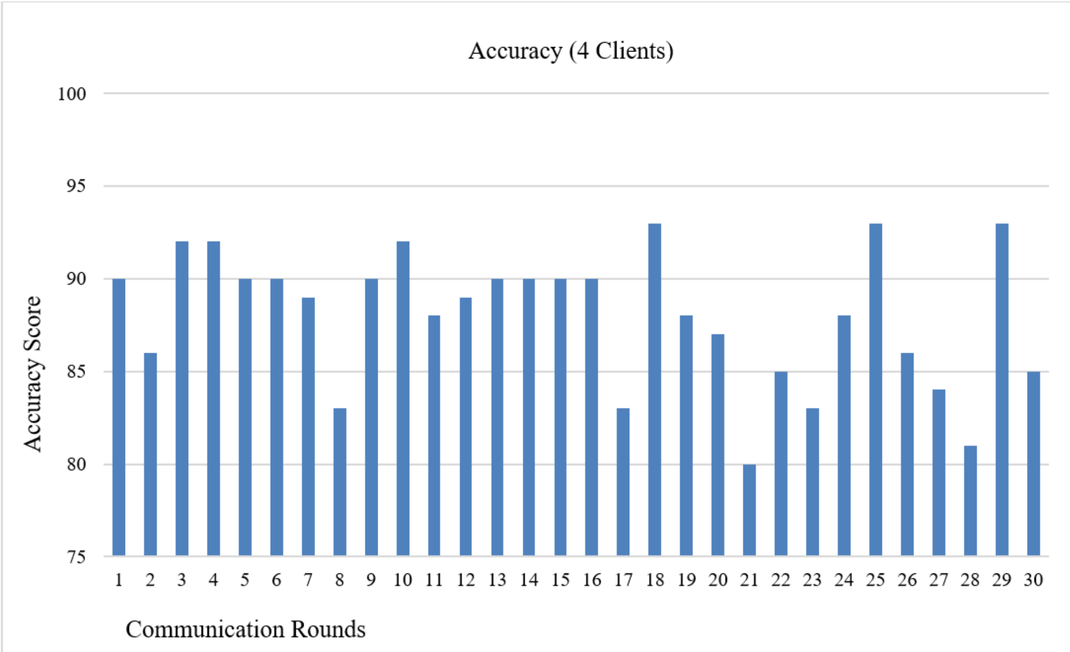

Figure 18. Accuracy score for 4 clients in 30 communication rounds.

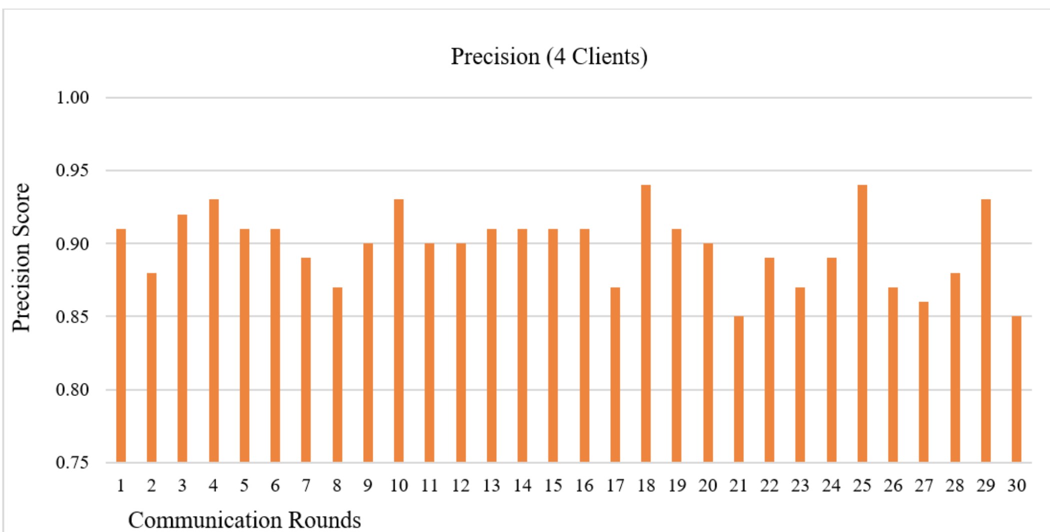

Figure 19. Precision score for 4 clients in 30 communication rounds.

The graph in **Figure 20** (shown below) is for the recall score performance indicator for 4-clients in a collaborative federated learning scenario. It reveals that again the lowest score which is 0.80 is in the 21st communication round and the highest score of 0.93 is in communication rounds 18, 25, 29. The second highest performance of 0.92 for recall score is in the communication rounds 3, 4, 10.

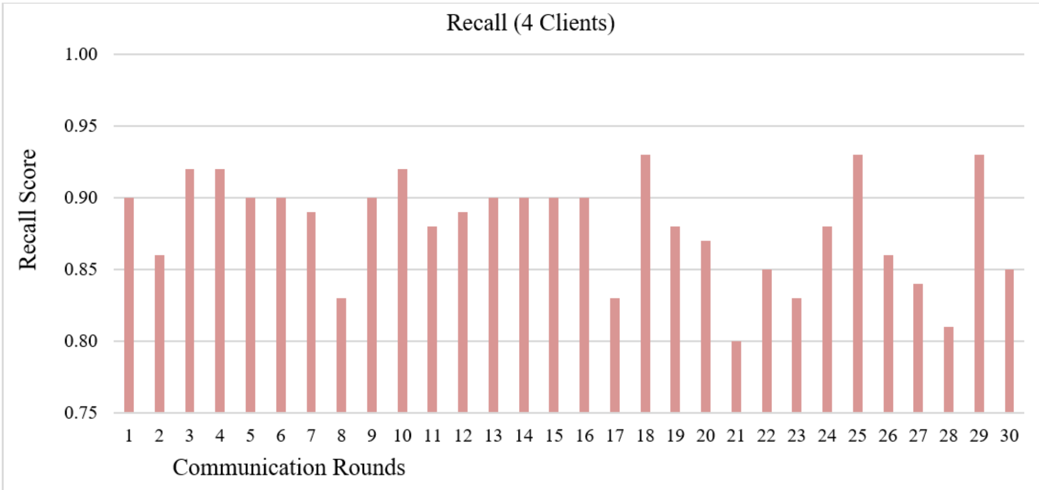

Figure 20. Recall score for 4 clients in 30 communication rounds.

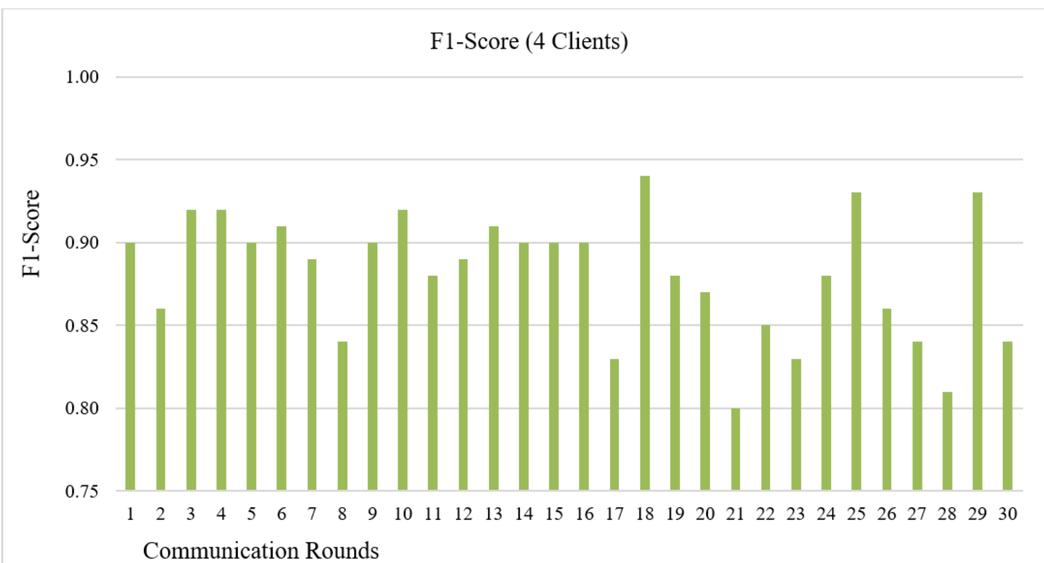

Figure 21. F1-Score for 4 clients in 30 communication rounds.

The F1-score metric values for 4 clients in the collaborative federated framework, over the 30 communication rounds are shown in **Figure 21**. Similarly, in the other metrics for the 4-client scenario the lowest performance for the F1-score is in the 21st communication round and the highest value of 0.94 for the f1-score is in the communication round number 18. However, the second highest f1-score value of 0.93 is in rounds number 25 and 29 which were in the highest-performing rounds earlier. The last metric is class-wise accuracy which is shown in **Figure 22**. The graph infers that the efficacy of identifying the DME class images is better in comparison to the DR class. The best performance of 100% is shown in the 27th and 29th communication rounds, followed by 97.9% (98% approx.) shown in the 9, 4 and 23 communication rounds. The lowest performance is 81% for the DME class and is shown in 8 and 12 communication rounds. On the contrary, the lowest performance of 57% for the DR class is shown in communication round 28 and the highest performance of 97% is shown in the 5th and 14th communication round. The second-highest accuracy score of 95% for the DR class is shown in the 7, 9, 25, and 26 rounds.

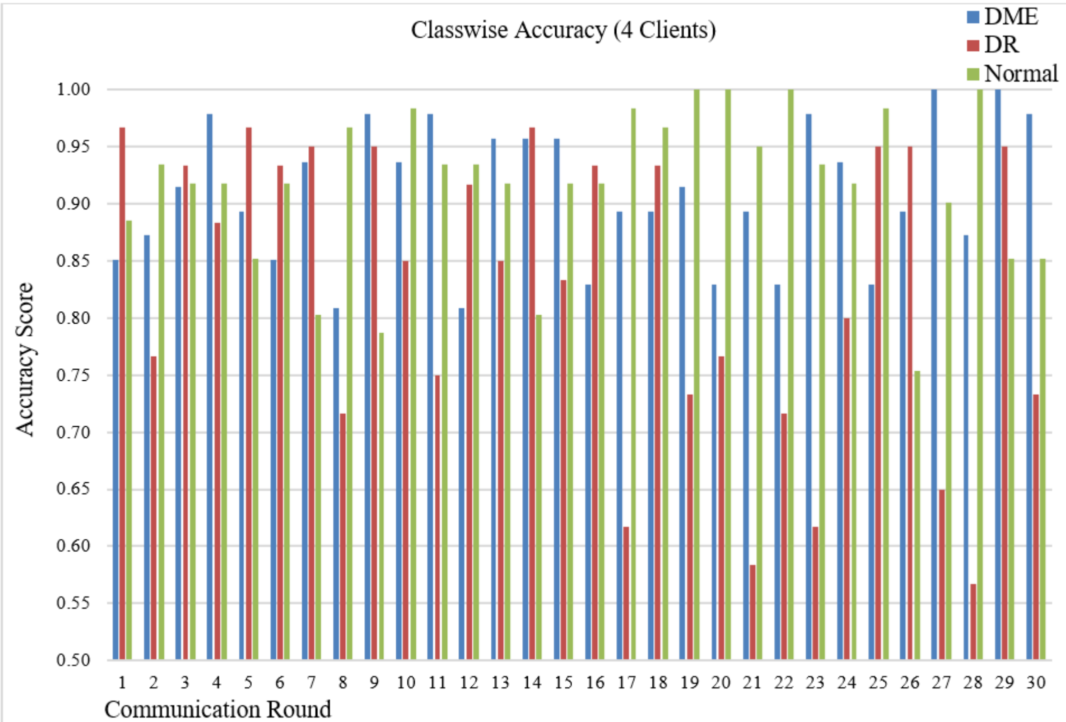

Figure 22. Class-wise accuracy for 2 clients in 30 communication rounds.

The second part of the result analysis is the clientwise comparison of the performance metrics average accuracy, precision, recall, f1-score, and class-wise accuracy. The first graph in **Figure 23** compares the average accuracy scores for the federated learning scenario with 2, 3 and 4 clients. The best scores of prediction accuracies are present in the 2 and 3 client scenarios in comparison to the scenario with 4 clients. Even though the score is lower for 4 clients yet if we see the curve there is more stability in the range of accuracies, this is indicative of better training that the model is quite stable. On the contrary, there is a lot of variation in the curves for 2 clients. In the 3-client implementation, the values lie in between that of 2-client and 4-client implementations.

The precision score graphs shown in **Figure 24** depict a comparison between three federated learning scenarios where the number of clients is 2, 3 and 4. The values are shown across 30 communication rounds. It is quite evident from the graphs that the scores of 2 clients show better peaks but also reach the minimum scores in all three scenarios, which is indicative of the variability and less stability of the model in terms of precision score. The 4-client training is better in terms of stability since the scores do not vary as much in comparison to the 2-client scenario. However, if see individual scores then the scores improve in 3 client scenarios and the convergence is faster with higher scores reaching in early communication rounds. Also considering the variability factor, the scores don't vary as much in 3 clients as in 2 clients.

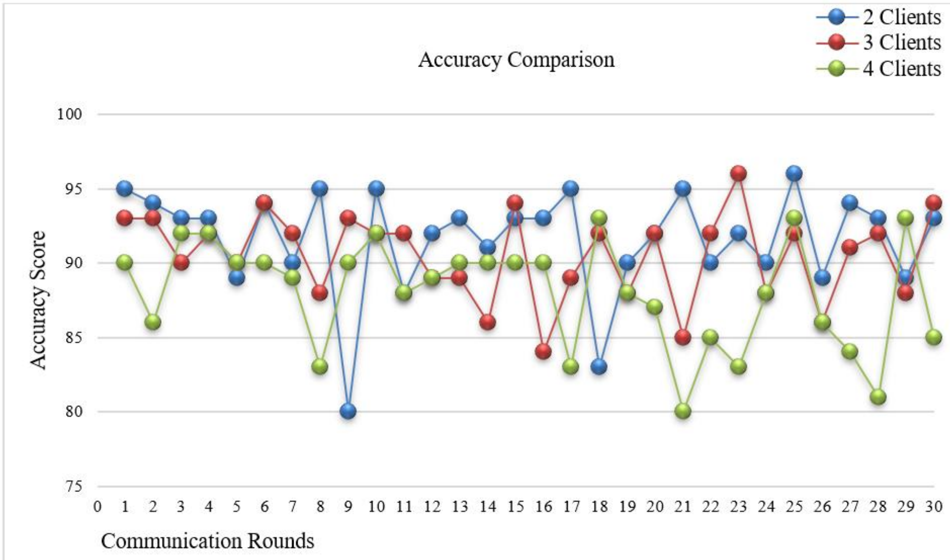

Figure 23. Accuracy comparison for 2,3,4 clients in 30 communication rounds.

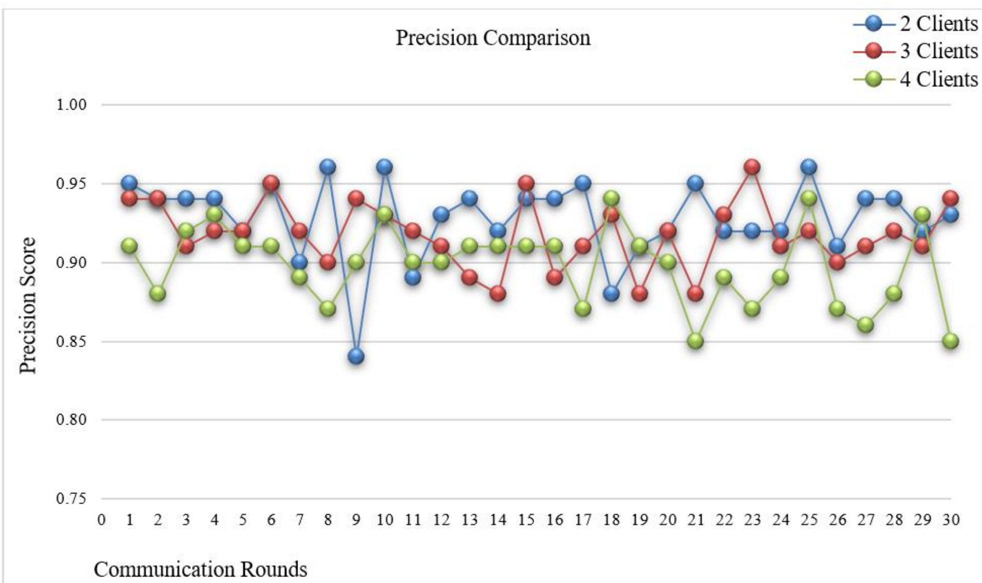

Figure 24. Precision comparison for 2, 3, 4 clients in 30 communication rounds.

The recall score comparison of 2, 3 and 4 clients for a privacy-preserving federated learning paradigm, is graphically drawn in **Figure 25** below. In any model, if the recall score is high, it suggests that the model effectively captures all the relevant instances whereas a low recall score is suggestive of significant misses of the relevant instances. The graphical comparison provides insights into the recall value comparison, these are highest for the 2-client implementations and the lowest for the 4-client implementations. The values although are better for 2 and 3 clients but the 4 client implementations witness more peaks and lesser dips which is indicative of better learning and stability of the model.

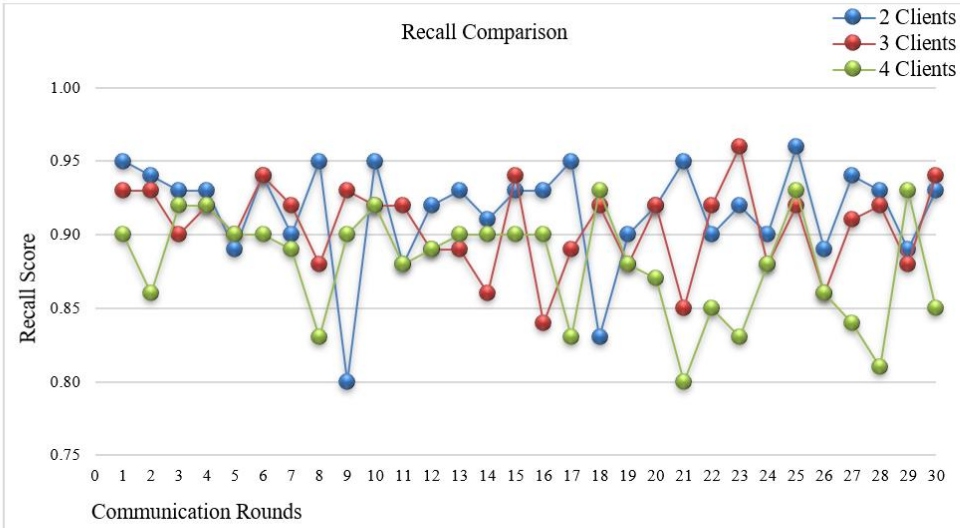

Figure 25. Recall comparison for 2, 3, 4 clients in 30 communication rounds.

The F1-score is one of the important decisive indicators of any machine learning model since it is the harmonic mean of the recall and precision scores. It shows the balance between the two metrics. The graph for F1-score score comparison across 30 communication rounds is given in **Figure 26** below. The inferences which are evident from the graph are that the F1-score values for 2 clients are the highest and the stability with more peaks and fewer dips is seen in the second half of the communication rounds. The 3-client implementation although shows more peaks in the second half of communication rounds but considering the stability factor the results are more stable in the first half of the communication rounds with very few dips. However, there is more balance in the first half of communication rounds in the 4-client scenario and overall, there is less variation in the scores in the 4-client scenario.

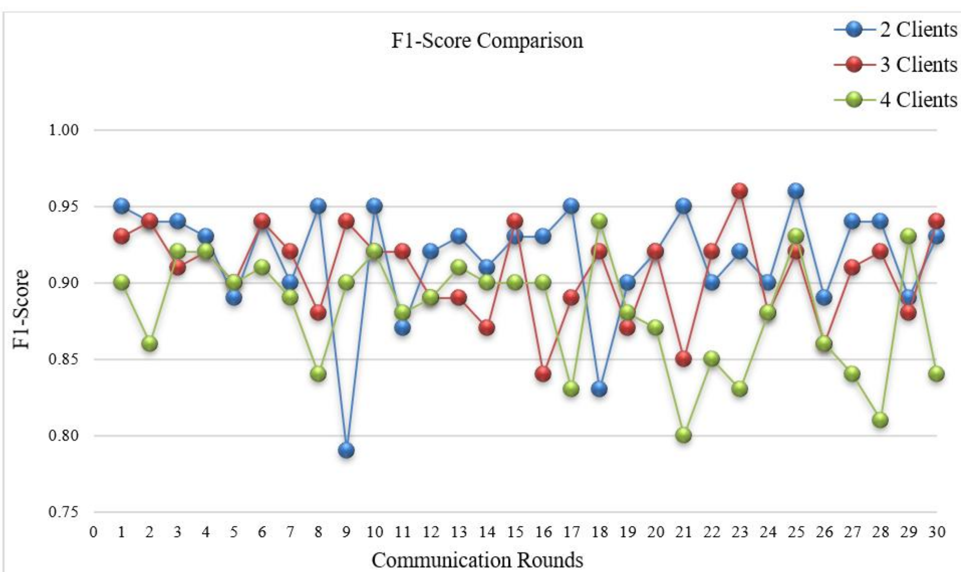

Figure 26. F1-Score comparison for 2, 3, 4 clients in 30 communication rounds.

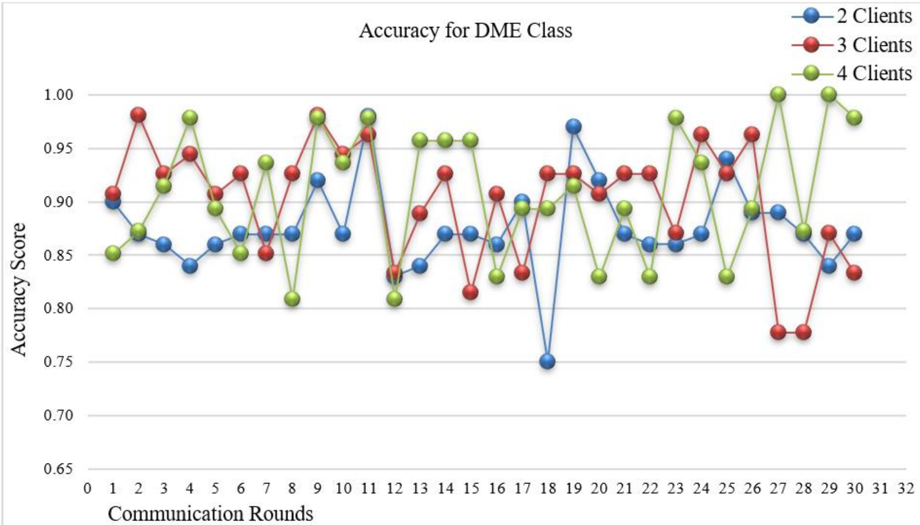

Figure 27. Accuracy comparison of DME class for 2, 3, 4 clients in 30 communication rounds.

The next metric is the class-wise accuracy for the two classes DME and DR. For a comparative analysis of the class-wise accuracy of 2, 3 and 4 client implementation **Figures 27** and **28** can be referred to. The class-wise accuracy graph for DME class shows that the highest accuracy for DME class is obtained in 4-client implementation at 27th and 29th communication round and the second highest accuracy is in 3-client implementation at 2nd and 9th communication round in the 2-client implementation at the 11th communication round. The class-wise accuracy comparison for the DR class reveals that the accuracy is 100% in many communication rounds in 2-client implementation and in 3-client 100% is achieved at the 7th and 21st communication rounds. However, the 4-client implementation doesn't attain 100% accuracy and goes up to 98% and the overall scores are much lesser. Therefore, for DR class 2-client implementation performs the best and for DME class 4-client implementation performs the best.

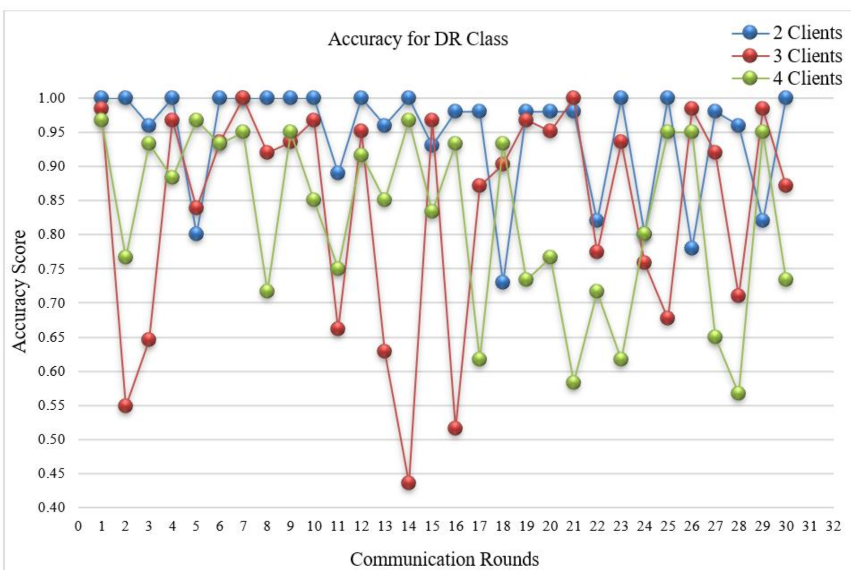

Figure 28. Accuracy comparison of DME class for 2, 3, 4 clients in 30 communication rounds.

Table 4 presents the performance metrics of the federated learning model with varying numbers of clients. The model demonstrates high and consistent accuracy, precision, recall, and F1 scores across 2 and 3 clients, with a slight decrease observed with 4 client implementations. The results indicate the robustness of the model in detecting ocular diseases while preserving data privacy.

Table 4. Summary of the results obtained from implementing the federated learning model.

Model used	Nature of data	Clients	Accuracy (in %)	Precision (in %)	Recall (in %)	F1-score (in %)
Federated Learning with MobileNetV2	IID	2	96	96	96	96
		3	96	96	96	96
		4	93	94	93	94

A comparison of the average accuracy obtained in the proposed work with existing work is depicted in **Figure 29**. The work proposed in the article shows superior performance than the other existing works. This is due to the different learning strategies, augmentation and optimisation techniques that helped in achieving a better accuracy score.

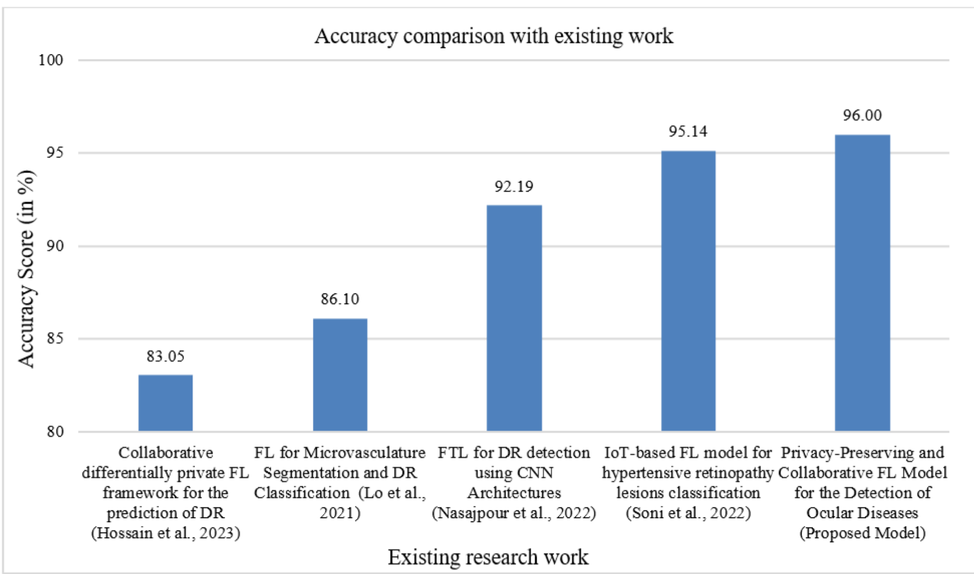

Figure 29. Average accuracy comparison of proposed work with the existing related work.

The proposed federated learning model with MobileNetV2 architecture performs better than other existing works. This confirms the robustness and efficacy of the model and highlights the potential of the approach to significantly enhance the predictive capabilities in ocular disease detection while ensuring data privacy and reducing computational overhead. The adoption of a collaborative federated framework can transform the way to approach data privacy in machine learning, particularly in sensitive healthcare data.

5. Conclusion

The article explains why privacy is important and how federated learning preserves the privacy of sensitive healthcare data. Federated learning allows collaboratively trained models to be built with multiple participating clients. The data of these clients is localised and never leaves the host location. Thus, ensuring privacy while enhancing the efficiency of training. Federated learning is especially useful in healthcare scenarios where privacy regulations of organisations limit data sharing. This is due to the underlying framework of federated learning which allows collaborative training in a distributed manner without the need of creating data banks. This article evaluates the results of implementing a federated learning framework for detecting diabetic retinopathy and diabetic macular edema using a MobileNetV2 architecture as the base deep learning model with 2, 3 and 4 clients. For each of the three scenarios, metrics such as accuracy, precision, recall, F1-score, and class-wise accuracy were analysed over 30 communication rounds. For the 2-client scenario, the model performed exceptionally well with prediction accuracy reaching 96% and occasionally dropping. Precision and recall scores followed similar trends, with optimal performance in multiple rounds. In the 3-client scenario, the model maintained high accuracy, peaking at 96% and showing stability despite slight decreases after the 23rd round. Precision and recall scores also exhibited consistent performance with minor variations. In the 4-client scenario, the average accuracy stayed high but showed more fluctuations, with peaks at 93% and dips to around 80%. Precision and recall scores followed similar patterns, with some variability but overall stable performance. When different client setups are compared, higher accuracy was achieved at the 2-client scenario but it also shows more variability in prediction accuracy. The 3-client setup exhibited a balance between accuracy and stability, while the 4-client setup demonstrated stable but slightly lower performance. Additionally, class-wise accuracy varied, with the 2-client setup performing best for the DR class and the 4-client setup performing best for the DME class. Overall, while the 2-client setup showed higher individual metrics, the 3-client setup provided a good balance of performance and stability, making it a favourable option.

In the future this research could involve expanding the federated learning framework to include more diverse and larger datasets, potentially incorporating non-IID data to further evaluate the robustness and scalability of the model. Further, differential privacy techniques can also be investigated to strengthen data security and compliance with privacy regulations.

Conflict of Interests

No conflicts of interest have been associated with this publication.

Acknowledgments

The authors would like to thank the editors and reviewers for their valuable comments and suggestions. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Aamir, M., Irfan, M., Ali, T., Ali, G., Shaf, A., Al-Beshri, A., Alasbali, T., & Mahnashi, M.H. (2020). An adoptive threshold-based multi-level deep convolutional neural network for glaucoma eye disease detection and classification. *Diagnostics, 10*(8), 602. https://doi.org/10.3390/diagnostics10080602.

Adnan, M., Kalra, S., Cresswell, J.C., Taylor, G.W., & Tizhoosh, H.R. (2022). Federated learning and differential privacy for medical image analysis. *Scientific Reports, 12*(1), 1953. https://doi.org/10.1038/s41598-022-05539-7.

Al-Fahdawi, S., Al-Waisy, A.S., Zeebaree, D.Q., Qahwaji, R., Natiq, H., Mohammed, M.A., Nedoma, J., Martinek, R., & Deveci, M. (2024). Fundus-DeepNet: multi-label deep learning classification system for enhanced detection of multiple ocular diseases through data fusion of fundus images. *Information Fusion, 102*, 102059. https://doi.org/https://doi.org/10.1016/j.inffus.2023.102059.

Alharbi, M. (2024). Multi-classification of eye disease based on fundus images using hybrid Squeeze Net and LRCN model. *Multimedia Tools and Applications, 83*(27), 69197-69226. https://doi.org/10.1007/s11042-024-18281-1.

Bajwa, M.N., Malik, M.I., Siddiqui, S.A., Dengel, A., Shafait, F., Neumeier, W., & Ahmed, S. (2019). Two-stage framework for optic disc localization and glaucoma classification in retinal fundus images using deep learning. *BMC Medical Informatics and Decision Making*, *19*(136), 1-16. https://doi.org/10.1186/s12911-019-0842-8.

Banabilah, S., Aloqaily, M., Alsayed, E., Malik, N., & Jararweh, Y. (2022). Federated learning review: fundamentals, enabling technologies, and future applications. *Information Processing & Management*, *59*(6), 103061.

Davidson, J.A., Ciulla, T.A., McGill, J.B., Kles, K.A., & Anderson, P.W. (2007). How the diabetic eye loses vision. *Endocrine*, *32*(1), 107-116. https://doi.org/10.1007/s12020-007-0040-9.

Gulati, S., Guleria, K., & Goyal, N. (2022a). Classification and detection of coronary heart disease using machine learning. In *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering* (pp. 1728-1732). IEEE. Greater Noida, India. https://doi.org/10.1109/icacite53722.2022.9823547.

Gulati, S., Guleria, K., & Goyal, N. (2022b). Classification of migraine disease using supervised machine learning. In *2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions*) (pp. 1-7). IEEE. Noida, India. https://doi.org/10.1109/icrito56286.2022.9964524.

Gulati, S., Guleria, K., & Goyal, N. (2023a). Classification and detection of diabetic eye diseases using deep learning: a review and comparative analysis. *AIP Conference Proceedings*, *2916*(1), 020005.

Gulati, S., Guleria, K., & Goyal, N. (2023b). Classification of diabetic retinopathy using pre-trained deep learning model- DenseNet 121. In *2023 14th International Conference on Computing Communication and Networking Technologies* (pp. 1-6). IEEE. Delhi, India. https://doi.org/10.1109/icccnt56998.2023.10308181.

Gulati, S., Guleria, K., & Goyal, N. (2023c). Classification of diabetic retinopathy using pre-trained deep learning model-DenseNet 121. *2023 14th International Conference on Computing Communication and Networking Technologies* (pp. 1-6). IEEE. Delhi, India. https://doi.org/10.1109/icccnt56998.2023.10308181.

Hartmann, F. (2018). *Federated learning*. Freie Universität Berlin.

Hossain, I., Puppala, S., & Talukder, S. (2023). Collaborative differentially private federated learning framework for the prediction of diabetic retinopathy. In *2023 IEEE 2nd International Conference on AI in Cybersecurity* (pp. 1-6). IEEE. Houston, TX, USA. https://doi.org/10.1109/icaic57335.2023.10044122.

Kairouz, P., McMahan, H.B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A.N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R.G.L., Eichner, H., El Rouayheb, S., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P.B., Gruteser, M., Harchaoui, Z., He, C., He, L., Huo, Z., Hutchinson, B., Hsu, J., Jaggi, M., Javidi, T., Joshi, G., Khodak, M., Konecný, J., Korolova, A., Koushanfar, F., Koyejo, S., Lepoint, T., Liu, Y., Mittal, P., Mohri, M., Nock, R., Özgür, A., Pagh, R., Qi, H., Ramage, D., Stich, S.U., Sun, Z., Suresh, A.T., Tramèr, F., Vepakomma, P., Wang, J., Xiong, L., Xu, Z., Yang, Q., Yu, F.X., Yu, H., & Zhao, S. (2021). Advances and open problems in federated learning. *Foundations and Trends® in Machine Learning*, *14*(1-2), 1-210. http://dx.doi.org/10.1561/2200000083.

Kaur, A., Guleria, K., & Trivedi, N.K. (2022). A deep learning based model for rice leaf disease detection. *2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)*, 1-5.

Kaur, H., Rani, V., Kumar, M., Sachdeva, M., Mittal, A., & Kumar, K. (2023). Federated learning: a comprehensive review of recent advances and applications. *Multimedia Tools and Applications.*, *83*(18), 54165-54188. https://doi.org/10.1007/s11042-023-17737-0.

Li, T., Sahu, A.K., Talwalkar, A., & Smith, V. (2020). Federated learning: challenges, methods, and future directions. *IEEE Signal Processing Magazine*, *37*(3), 50-60. https://doi.org/10.1109/msp.2020.2975749.

Lo, J., Yu, T.T., Ma, D., Zang, P., Owen, J.P., Zhang, Q., Wang, R.K., Beg, M.F., Lee, A.Y., Jia, Y., & Sarunic, M.V. (2021). Federated learning for microvasculature segmentation and diabetic retinopathy classification of OCT data. *Ophthalmology Science*, *1*(4), 100069. https://doi.org/10.1016/j.xops.2021.100069.

Matta, S., Hassine, M.B, Lecat, C., Borderie, L., Le Guilcher, A., Massin, P., Cochener, B., Lamard, M., & Quellec, G. (2023). Federated learning for diabetic retinopathy detection in a multi-center fundus screening network. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society* (pp. 1-4). IEEE. Sydney, Australia. https://doi.org/10.1109/embc40787.2023.10340772.

Nasajpour, M., Karakaya, M., Pouriyeh, S., & Parizi, R.M. (2022). Federated transfer learning for diabetic retinopathy detection using CNN architectures. In *SoutheastCon* (pp. 655-660). IEEE. Mobile, AL, USA.

Ng, D., Lan, X., Yao, M.M.S., Chan, W.P., & Feng, M. (2021). Federated learning: a collaborative effort to achieve better medical imaging models for individual sites that have small labelled datasets. *Quantitative Imaging in Medicine and Surgery*, *11*(2), 852-857. https://doi.org/10.21037/qims-20-595.

Nguyen, T.X., Ran, A.R., Hu, X., Yang, D., Jiang, M., Dou, Q., & Cheung, C.Y. (2022). Federated learning in ocular imaging: current progress and future direction. *Diagnostics, 12*(11), 2835.

Peyvandi, A., Majidi, B., Peyvandi, S., & Patra, J.C. (2022). Privacy-preserving federated learning for scalable and high data quality computational-intelligence-as-a-service in Society 5.0. *Multimedia Tools and Applications*, *81*(18), 25029-25050. https://doi.org/10.1007/s11042-022-12900-5.

Pfitzner, B., Steckhan, N., & Arnrich, B. (2021). Federated learning in a medical context: a systematic literature review. *ACM Transactions on Internet Technology*, *21*(2), 1-31. https://doi.org/10.1145/3412357.

Qi, P., Chiaro, D., Guzzo, A., Ianni, M., Fortino, G., & Piccialli, F. (2024). Model aggregation techniques in federated learning: a comprehensive survey. *Future Generation Computer Systems*, *150*, 272-293.

Safi, H., Safi, S., Hafezi-Moghadam, A., & Ahmadieh, H. (2018). Early detection of diabetic retinopathy. *Survey of Ophthalmology*, *63*(5), 601-608. https://doi.org/10.1016/j.survophthal.2018.04.003.

Shanmugarasa, Y., Paik, H., Kanhere, S.S., & Zhu, L. (2023). A systematic review of federated learning from clients' perspective: challenges and solutions. *Artificial Intelligence Review*, *56*(2), 1773-1827. https://doi.org/10.1007/s10462-023-10563-8.

Saini, A., Guleria, K., & Sharma, S. (2023). A deep learning-based convolutional neural networks model for white blood cell classification. In *2023 4th International Conference for Emerging Technology* (pp. 1-5). IEEE. Belgaum, India. https://doi.org/10.1109/incet57972.2023.10170666.

Sanghavi, J., & Kurhekar, M. (2024). Ocular disease detection systems based on fundus images: a survey. *Multimedia Tools and Applications*, *83*(7), 21471-21496. https://doi.org/10.1007/s11042-023-16366-x.

Sharma, R., Kukreja, V., & Kadyan, V. (2021). Rice diseases detection using convolutional neural networks: a survey. In *2021 International Conference on Advance Computing and Innovative Technologies in Engineering* (pp. 995-1001). IEEE. Greater Noida, India. https://doi.org/10.1109/icacite51222.2021.9404620.

Sharma, R., Kukreja, V., Kaushal, R.K., Bansal, A., & Kaur, A. (2022). Rice leaf blight disease detection using multi-classification deep learning model. In *2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)* (pp. 1-5). IEEE. Noida, India.

Sharma, S., Guleria, K., Kumar, S., & Tiwari, S. (2023). Deep learning based model for detection of vitiligo skin disease using pre-trained inception V3. *Article in International Journal of Mathematical Engineering and Management Sciences*, *8*(5), 1024-1039. https://doi.org/10.33889/ijmems.2023.8.5.059.

Soni, M., Singh, N.K., Das, P., Shabaz, M., Shukla, P.K., Sarkar, P., Singh, S., Keshta, I., & Rizwan, A. (2023). IoT-based federated learning model for hypertensive retinopathy lesions classification. *IEEE Transactions on Computational Social Systems*, *10*(4), 1722-1731. https://doi.org/10.1109/tcss.2022.3213507.

Srivastav, S., Guleria, K., & Sharma, S. (2023). Tea leaf disease detection using deep learning-based convolutional neural networks. In *2023 IEEE World Conference on Applied Intelligence and Computing* (pp. 569-574). IEEE. Sonbhadra, India. https://doi.org/10.1109/aic57670.2023.10263835.

Tang, Z., Wong, H.S., & Yu, Z. (2024). Privacy-preserving federated learning with domain adaptation for multi-disease ocular disease recognition. In *IEEE Journal of Biomedical and Health Informatics*, *28*(6), 3219-3227.

Thotad, P.N., Bharamagoudar, G.R., & Anami, B.S. (2023). Diabetes disease detection and classification on Indian demographic and health survey data using machine learning methods. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, *17*(1), 102690. https://doi.org/10.1016/j.dsx.2022.102690.

ur Rehman, M.H., & Gaber, M.M. (2021). *Federated learning systems: Towards next-generation AI*. Springer Nature. ISBN: 9783030706036(p), 9783030706043(e).

Zhang, H., Bosch, J., & Olsson, H.H. (2020). Federated learning systems: architecture alternatives. In *2020 27th Asia-Pacific Software Engineering Conference* (pp. 385-394). IEEE. Singapore.