Federated Learning for Early Detection of Diabetic Retinopathy in Distributed Healthcare Systems

S.Sathya,
Associate Professor, *Department of Computer Science and Information Technology, School of Computing Sciences,*
VISTAS, Pallavaram, Chennai.
ssathya.scs@vistas.ac.in

R. Anandha Lakshmi
Assistant Professor, *M.O.P. Vaishnav College for Women,*
Chennai, Tamil Nadu, India
anandhalakshmir.csc@mopvaishnav.ac.in

R.Bagavathi Lakshmi,
Department of Computer Applications, Vels Institute of Science, Technology & Advanced Studies,
Chennai, Tamil Nadu, India.
rbagavathi.scs@vistas.ac.in

Vishwa Priya V,Assistant professor,
Department of computer science and information technology,Vels Institute of science technology and advanced studies
Chennai,vishwapriya13@gmail.com

D. Narayani.
Assistant Professor
Department of Computer Applications Vels institute of Science Technology and Advanced Studies, Chennai, Tamil Nadu,
India. narayanidsarveshk@gmail.com

N. Anandakrishnan,
Associate *Professor,PG & Research,Department of Computer Science, Nilgiri College of Arts and Science*, Thaloor.
drakpcw@gmail.com

***Abstract*— Diabetic retinopathy (DR) is one of the leading causes of preventable vision loss in the world, and early detection is critical for an effective intervention. However, the sensitive nature of patient data, together with regulatory considerations, restricts the use of centralized model training of healthcare data across institutions. This study presents a federated learning (FL) framework, to train a deep convolutional neural network(CNN) (EfficientNet-B0), to provide remote, early detection of DR type using retinal fundus images collected from multiple clinics that are distributed geographically, without sharing un-identified raw health data. The performance of the FL training-and-test scheme was evaluated on a range of AUC, accuracy, sensitivity, and specificity against a central, local only, ensemble, and pre-trained model. It is found that the FL model achieved an AUC of 0.91, and accuracy of 88.3%. Federated learning is chosen for this study to ensure patient privacy to handle distributed non-IID data, and achieve near-centralized diagnostic accuracy. This work also explores the future research opportunities for federated learning, and suggests that federated learning represents an advanced, scalable and privacy-respecting avenue for the implementation of AI-supported diagnostic imaging tools in the healthcare distributed healthcare systems.**

Keywords*— Federated Learning, Diabetic Retinopathy, Deep Learning, Medical Imaging Distributed Healthcare Systems.

I. Introduction

The number of people living with diabetes mellitus(DM) continues to increase worldwide. The International Diabetes Federation estimates that the world will have 643 million people living with diabetes by 2030. There are many complications with this chronic condition, but diabetic retinopathy (DR) is probably the most problematic as it is one the top four causes of impairment and preventable blindness in working age adults, an estimated 136 million people worldwide. DR is a disease that has a progressive course and occurs due to prolonged hyperglycemia (high blood glucose levels) which results in damage to the retinal blood vessels in the form of new blood vessel growth (i.e.microaneurysms), hemorrhaging and various other forms of vascular damage. If DR is not detected in a timely manner, it can result in retinal detachment or macular edema. The most liberating aspect of DR is timely detection and treatment prevents up to 95% of vision loss related to DR. This shows the critical need for patients to be screened for DR and accurately diagnosed[1].

In spite of guidelines suggesting regular retinal examinations for diabetic patients, screening levels are suboptimal in many areas due to systemic factors such as limited access to ophthalmologists (especially in rural and under-resourced areas). Manual diagnosis with retinal fundus photography takes time and has inter-observer variations, rendering it often inaccessible for patients in low-resource settings. Artificial intelligence (AI), and its deep learning (DL) methods, are increasingly being utilized with great efficacy to support health professionals finding DR in retinal images with high accuracy and low turnaround time. These AI and DL systems have reached performance levels similar to expert ophthalmologists when developed from diverse, large and high-quality datasets[2]

Nevertheless, the creation of strong models using deep learning often requires a large amount of labelled medical data, which opens another challenge in the healthcare context: data privacy and security. Data ownership and privacy restrictions are the goal of many regulatory frameworks, such as HIPAA (Health Insurance Portability and Accountability Act)GDPR (General Data Protection Regulation) and India's DISHA guidelines, which impose limitations on the centralising of patient data and freely sharing patient data. This makes most traditional centralised training approaches impractical or noncompliant in many real-world situations, especially if they involve many institutions cross-institutionally. Data ownership and privacy restrictions create siloed data, fragmenting essential clinical data and diminishing the impact of AI solutions in medical imaging. Therefore, a decentralised machine learning approach, known as federated learning (FL), can and has been used to obviate the concerns assumed in the previous limitations[3].

In the past, centralized datasets like EyePACS, Messidor, and IDRiD have been essential to AI-based DR detection studies[4]. While utilizing centralized datasets for benchmarking is helpful, it does not reflect the diversity or heterogeneity seen globally in different populations. In addition to the aforementioned issues, there is little published research that empirically demonstrates how models trained using federated learning compare with locally developed models, or conventional centralized models, especially with respect to performance and generalizability, and preventable system costs (e.g., communication costs and inference efficiency). Furthermore, published guidelines and standards of practice related to federated learning clinical deployment and use are virtually non-existent. This research is relevant as it answers this question with a design and evaluation of a federated learning framework to detect diabetic retinopathy early, using retinal fundus images from various institutions, on a

federated learning system in several locations. The main research question driving this work is:
Can federated learning facilitate privacy-preserving, accurate, and generalizable early detection of diabetic retinopathy across distributed healthcare institutions without sharing raw patient images?

Problem Statement

Although DL models show promise for detecting DR, sharing data between institutions due to privacy issues limits the development of generalizable AI models. Federated learning may provide a potential solution, but the efficacy, efficiency, and viability of the approach for diagnosing ophthalmic disease across real-world distributed environments have yet to be adequately assessed. The main objectives of the study are as follows

- To present a federated learning framework that allows deep learning models for the detection of diabetic retinopathy to be trained cooperatively without exchanging raw patient data.
- To assess the federated model using common evaluation metrics in comparison to baseline techniques such as pretrained models, local-only models, and centralised training.

This paper's remaining sections are arranged as follows: The relevant studies and literature review on federated learning and DR detection are presented in Section 2. The suggested technique, data sources, and experimental setting are described in detail in Section 3. Results, comparisons, and implications are covered in Section 4, and Section 5 offers conclusions and further research.

II. RELATED WORKS

Diabetic Retinopathy (DR) is a major cause of vision loss and blindness that requires timely diagnosis if treatment is to be effective. Finally, deep learning approaches to the diagnosis of DR face challenges as well, as data privacy concerns and their availability will further hinder efforts to obtain data in a single centre or across even a few locations. To address these issues, we proposed an FL-ViT framework which allows detection for DR image analyses in a secure distributed paradigm while maintaining privacy and scalability as well as high diagnostic accuracy. Berbar et al., (2022) develops a novel approach to detecting and grading diabetic retinopathy (DR) in fundus images using a CNN that adds preprocessing to account for image quality. The results of two CNN architectures, a binary classifier, and a severity-grade classifier achieve high F1-scores across both Messidor and EyePACS datasets, showcasing the robustness and validity of their approach[5].

Sornil et al. (2024) has provided a DL based framework for DR classification involving better preprocessing techniques, wavelet-based feature extraction, and a modified ResNet50 framework. The model utilizes transfer learning, advanced data augmentation techniques, and different datasets EyePACS, APTOS, Messidor-2, and DDR, to classify 4 levels of DR severity aiming to improve early diagnosis and management [6]. Bhimavarapu et (2022) presented an advanced method for automatic diabetic retinopathy (DR) detection based on novel filtering, segmentation, feature extraction, and classification techniques. The advanced method uses a multi-threshold segmentation technique based upon an improved grasshopper optimization method for segmenting the lesion regions. The model extracted 41 features and was classified using an improved Naïve Bayes classifier with an outstanding 99.98% accuracy from the APTOS dataset. This approach is an improvement over previous methods and encourages a good opportunity for accurate diabetic retinopathy diagnoses[7].

DR is an eye disease associated with diabetes that may potentially result in complete blindness if left undiagnosed. Chetoui et al. developed a model to classify DR and Normal cases using a Vision Transformer based federated learning strategy across four different institutions, and increased performance by 3% while training with data privacy, data security, and access control in mind[8]. Alanazi et al. have proposed a data balancing mechanism utilizing SMOTE and a DL-based method for classifying DR, by combining Weiner filtering and median filtering for image enhancement with feature extraction, then classifying using VGG, and performing FL for privacy-preserving training. The FedCNN integrated model accomplished solid results, high accuracy, scalability and security while operating across various medical institutions [9].

Bhulakshmi et al., (2024) proposes a new FL framework, guided by Federated Differential Evolution Optimization (FedDEO), for DR detection and classification. FedDEO optimizes hyperparameters (like learning rate and batch size) in decentralized organizations while preserving privacy when data are still local to the organizations. The proposed method obtained results using the MESSIDOR database of 96.98% accuracy, 98.12% specificity, 97.12% recall, and 98.00% F1-score. The results substantiate improvements on DR classification after integrating FedDEO and FL. Improving DR detection and classification performance through FedDEO, while also preserving data privacy, provides a scalable and secure clinical solution [10].

Mao et al., (2024) presents a FL framework for DR diagnosis. The FL framework is developed because of issues regarding data scarcity and privacy concerns. The FL framework includes a high-quality pixel-level dataset (TJDR) and new cross-dataset FL algorithms - αα-Fed and the adaptive-αα-Fed which support the grading of DR and the segmentation of lesions. With results demonstrating newer accuracy metrics and privacy-preserving performance.[11]. Swapna et al., (2025) proposed a FL-ViT framework for diabetic retinopathy (DR) detection enabling the privacy-preserving, distributed learning of multiple healthcare institutions simultaneously. The authors marry federated learning with Vision Transformers in order to facilitate secure and computationally efficient feature extraction. The proposed model achieves 93% accuracy on the APTOS dataset, achieving AI scalability, generalizability, and in compliance with healthcare data privacy standards. Table 1 provides the recent research studies on DR detection techniques, summarizing the techniques, advantages and limitations of various CNN, deep learning and federated learning methods from 2022 until 2025.

TABLE 1 RECENT RESEARCH STUDIES IN DR DETECTION

Author (Year)	Method	Strengths	Limitations
Berbar et al. (2022)	CNN with binary and severity-grade classifiers, image preprocessing	High F1-scores across datasets; enhanced robustness through preprocessing	Limited scalability; does not address privacy or data-sharing constraints

Sornil et al. (2024)	DL with wavelet-based feature extraction and modified ResNet50	Improved severity classification; effective use of transfer learning and augmentation	High computational cost; lacks privacy-preserving mechanisms
Bhimavarapu et al. (2022)	Grasshopper optimization for segmentation + Improved Naïve Bayes classifier	Excellent accuracy (99.98% on APTOS); precise lesion segmentation	Focused on APTOS dataset; lacks generalization across broader clinical settings
Chetoui et al. (2023)	FL with Vision Transformer (ViT) across four institutions	Privacy-preserving; ViT boosts accuracy (+3%); secure distributed training	Limited class granularity (binary classification only); potential scalability issues
Alanazi et al. (2025)	FedCNN with SMOTE, image filtering, and VGG feature extraction	Balanced class distribution; integrated privacy and enhancement steps; scalable	May be sensitive to noise; filtering methods could lose key features
Bhulakshmi et al. (2024)	FL with Federated Differential Evolution Optimization (FedDEO)	High accuracy and F1-score; hyperparameter tuning; strong privacy guarantees	Relies heavily on MESSIDOR dataset; may need tuning for other datasets
Mao et al. (2024)	FL with αα-Fed and adaptive-αα-Fed + pixel-level lesion segmentation (TJDR dataset)	Supports grading and segmentation; advanced privacy-preserving FL algorithms	Pixel-level data annotation is resource-intensive; methods not yet tested on diverse datasets
Swapna et al. (2025)	FL-ViT framework for distributed DR detection	Combines ViT and FL; scalable; privacy-preserving; 93% accuracy on APTOS	Accuracy could be further improved; needs validation on more diverse institutional datasets

While DL-based DR detection approaches have advanced significantly utilizing FL methods, there are still many limitations and research gaps to address. Many of the existing models, while often achieving high accuracy, make many assumptions and still rely on individual datasets such as APTOS or MESSIDOR, thus limiting their applicability to other populations and imaging modalities. In addition, most of the methods focus on classification or segmentation, but very rarely on simultaneously doing both aspects of DR, thereby neglecting to address DR comprehensively. Methods that are based on privacy-preserving FL frameworks, while promising, will often suffer from a lack of fine-grained lesion detection or are restricted to binary classifications when many of the clinical DR management approaches would likely benefit from multi-class severity grading models. In addition, only a few works have considered hyperparameter tuning in the context of FL, with even fewer studies implementing methods such as FedDEO that actually consider the effects of hyperparameter tuning out of the context of any particular dataset. Similarly, computing complexities related to models like the Vision Transformers, and the number of resources required to run the various models also makes it academically difficult to propagate FL approaches to clinical settings with low-resource environments while still preserving some of the key aspects of FL. As a consequence of these essential issues, there is a need for better constructed, lighter, and interpretable FL models that can provide reasonably accurate DR grading and precise lesion segmentation within heterogeneous datasets, while prioritizing patient and data privacy, extensibility and scalability, and demonstrated clinical utility.

III. METHODOLOGY

DR is a vision threatening complication of diabetes that requires early and accurate detection for effective treatment. Automated diabetic retinopathy diagnosis from retinal images using deep learning models demonstrates high prospects to diagnose diabetic retinopathy however; data privacy issues prevent the centralized data collection. Federated learning (FL) can overcome data privacy concerns to provide privacy-preserving FL model training to multiple healthcare ambulatory institutions while retaining ownership of their raw data. The study presents a federated framework for DR detection that utilizes CNN-based architectures and secure, decentralized learning protocol.

A. Data Collection

The first phase of the FL framework for DR detection consists of partner selection, including 5–10 partner healthcare institutions such as hospitals and/or clinics that care for many diabetic patients and have retinal imaging systems. Partner sites prepare data locally at the participating sites by selecting suitable, high-quality fundus images in JPEG or PNG format, allowing reliable severity annotation for diabetic retinopathy according to an established scale, for example, the International Clinical DR(ICDR) scale. Poor quality images that are blurry, occluded, or uninterpretable for diagnostic purposes are discarded from the study, while researchers also consider removing images covering more than 60% of a pupil or other factors affecting diagnostic quality. Optionally, the partner institution may also decide to share and/or analyze patient metadata which may include age, gender, diabetes duration, and/or HbA1c value to improve model personalization throughout the model's usefulness. With the data at the local partner institutions and only lacking this local metadata, critical aspects remain key points of federated learning including: data remains within the institutional boundary, privacy is preserved, sharing of raw data between institutions is prevented for privacy and regulatory compliance, and compliance with federated learning practices is validated**.** Figure 1 shows the architecture of the proposed model.

B. System Architecture:

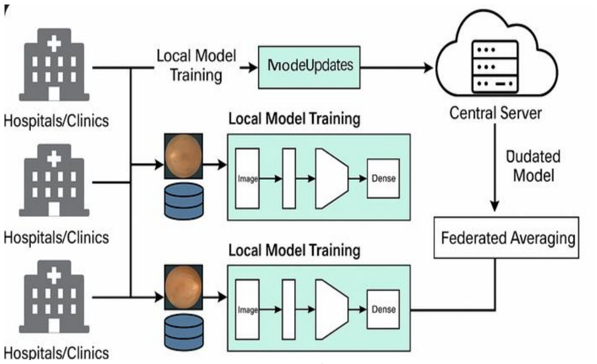

Figure 1 Architecture of the proposed model

The workflow of the proposed FL framework for DR detection begins with multiple hospitals or clinics holding retinal fundus images locally, ensuring patient privacy and compliance with relevant regulations. Each hospital or clinic will download an initial global model from the central server, and subsequently, each hospital or clinic will carry out local training on its own data, while the hospital or clinic will only transmit model updates (i.e. weights or gradients) back to the central server. After each hospital or clinic has completed the local training, the central server conducts federated averaging that averages the model updates from all participating sites to create an improved global model. The improved global model will then be downloaded to each participating hospital or clinic for the next training round which will be carried out iteratively until convergence of the model weights occurs, thus enabling collaborative learning across institutions while preserving patient privacy.

The federated system architecture for DR detection consists of using fixed (but flexible) CNN backbone models like EfficientNet-B0[13] or ResNet-50[14], trained on ImageNet, to exploit the benefits of transfer learning. These models would be a strong feature extractor baseline on retinal fundus images. The classification head could then be fit to either the model for a single binary classifier (e.g., DR vs. no DR), which would create a simplified and faster diagnostic pipeline, or fit to a model for a multi-class classifier (no DR, mild, moderately severe, severe) - this would allow for a greater degree of stratification of risk assessment. Once the model architecture is selected, a central coordinating server would create the base or global model and distribute the model parameters and weights to all member healthcare institutions. This will begin federated learning, where each site can train this model with their local data independently, establishing a collaborative learning process with privacy to their local data.

C. Federated Training Process

Federated training is a collaborative process to train a global model for DR detection where privacy is maintained at each institution. In each round, the federated training process begins with the first step of local training at each participating institution (Step 5), in which the information held by the partner institution is trained on the model received from the central server is trained using that institution's private retinal dataset, for a fixed number of local epochs (for example 5 epochs) of model training with the standard hyperparameters of 0.001 learning rate and batch size of 32. It is important to note that the raw data remains privately held at the institution and only model weight updates, or gradients, are generated and sent to the central server[15].

In the server-based coordinates, the average of all local updates is taken over all sites (i.e., FedAvg algorithm). This calculates summary weighted average of the updates and then sends back the improved new global model to all sites. Then it is repeated for 100 rounds, until the model gets stabilized or converged. This decentralized approach provides an economically viable way to improve the model while still protecting the private information of sensitive patients.

The FL training aims to minimize a global loss F(w) across K participating healthcare institutions

$$F(w) = \sum_{k=1}^{K} \frac{n_k}{n} F_k(w) \quad (1)$$

where n_k is the number of sample at client k, n= $\sum_{k=1}^{K} n_k$ is the total sample across all clients. $F_k(w)$ is the local loss function at client k. For each client

$$F_k(w) = \frac{1}{n_k} \sum_{i=1}^{n_k} l(f(x_i\,; w), y_i) \quad (2)$$

Each client receives the global weights w_t and updates locally for E epochs using Adam optimizer.

$$w_k^{t+1} = w_t = -\eta \nabla F_k(w_t) \quad (3)$$

where η is the learning rate.

After local updates, the central server aggregates weights using weighted averaging(FedAvg).

$$w_{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} w_k^{t+1} \quad (4)$$

This ensures larger datasets contribute proportionally more to the global model.

The selection of the substrate, or model architecture, plays a significant role in the utility and viability of federated learning for diabetic retinopathy detection. EfficientNet-B0 is a light-weight model with roughly 5.3 million parameters. This model has fast computation and efficiency and is best suited for deployment in smaller clinics that are limited in data or hardware resources. ResNet-50 has about 25 million parameters. ResNet-50 has greater depth for feature extraction and may detect some of the more complex relationships in the retinal data than the EfficientNet. However, ResNet-50 has higher communication costs and longer training times. EfficientNet-B0 is less demanding and reduces the communication costs between each round of federated learning. The communication cost is very important in a clinical network that is low-bandwidth. ResNet-50 may give a slight increase in accuracy over the lighter EfficientNet-B0 model but loses feasibility in a large-scale distributed case. EfficientNet-B0 is the best choice for a more scalable multi-institution deployment, whereas ResNet-50 may be more useful in a high-resource situation where maximum diagnostic accuracy is needed.

IV. Results and Discussion

This study aimed to investigate the performance of five machine learning strategies, which are FL, Centralized CNN, Local Models only, Ensemble Local Models, and Pre-trained models, to classify DR, using retinal fundus images. Multi-institutional data sets with 45,000 annotated images were collected from six healthcare providers at a geographic distance from one another. This section will describe the process to create the dataset and the method to train the models under different data governance designs, and the performance of the five machine learning strategies.

A. Dataset Description

For this study, a multi-institutional dataset of 45,000 retinal fundus images are created across a distribution of six healthcare sites located in different geographical areas, each acting as a federated site. Each site contributed between 5,000 to 10,000 images depending on patient volumes and resources for imaging. Each image had labels assigned according to the International Clinical Diabetic Retinopathy (ICDR) scale which defines DR severity from 0 (No DR) to 4 (Proliferative DR). The labels were binarized for simplicity in modeling, and focused on the early detection of the disease; images were labeled with Non-DR (class 0) which represents ICDR grade 0 and DR (class 1) models of grade 1 through 4.

Images of the fundus were collected from distinct non-mydriatic cameras from various manufacturers, representing standard variance in imaging conditions. Image resolutions spanned from 640×480 pixels to 1024×1024 pixels. Pre-processing, such as resizing, normalization of numeric attributes, and data augmentation (rotating, flipping, and brightening), was done at each site to standardize image inputs before training. Sites did not share raw data with each other, only updates to the model and training were shared, as per federated learning principles. Each site kept its own train-test split (80%/20%) and could therefore evaluate the performance of local and global models independently. For centralized benchmarking, an external validation dataset (5,000 images with annotations) was used to assess final model performance across all federated nodes. The external validation dataset was composed of images from the publicly available Messidor-2 dataset.

B. Performance Evaluation

Early detection of DR is extremely important in preventing vision loss; however, well-performing AI models require large distributions of diverse datasets that are generally restricted by privacy requirements. In order to alleviate data-sharing obstacles, this study evaluates the efficacy and comparative results of multiple model-training approaches: FL Centralized CNNs, Local-only models, Ensembles, and Pre-trained models. Each of these, provides a set of trade-offs in their own accuracy, privacy, generalizability, and real-world implementation.

Centralized CNN Model: All data is collected from the participating hospitals and stored in a single centralized server. A deep CNN (e.g., EfficientNet), is trained on the pooled dataset for DR classification. It can be expected to perform well, as it combines heterogeneous training data as a unified dataset. However, this model still jeopardizes patient privacy and potential data sharing regulations.

Local-only Models: A hospital will develop its own model entirely independent of other hospitals using its local dataset. There is no interaction or knowledge sharing across the different sites. These models are likely underpowered in terms of model performance due to the limited variability of the dataset. While acceptable in terms of privacy, they are not able to generalize to unseen data from other hospitals.

Ensemble of Local Models: Without exchanging data, combine predictions from several locally trained models (for example, by weighting or majority voting) to boost site variety. It is generally more stable than any one local model, but it can be difficult to coordinate and maintain, and as the number of sites grows, the costs of computation and communication (both on the local model and aggregated forecasts) frequently become unmanageable.

Pretrained Model (Static): New local data is directly fed into a model that was trained on an external dataset (like EyePACS). It is quick to implement and eliminates the need for retraining. Nevertheless, domain shifts and a lack of local fine-tuning impair performance. Although it lacks customisation and flexibility, it's a solid foundation for quick implementation. To ensure distributed DR screening can sustain its objective, it is important to identify the best approach to model training and data sharing process and ultimately provide final considerations on how best to optimize performance, security and privacy for stakeholders. The following analysis provides a summary of the options referenced and performance and implications of operationalizing the models. In table 1, a quantitative comparison of five model training approaches, FL, Centralized CNN, Locally-only Models, Ensemble of Local Models, and Pretrained Static Models, are provided, evaluated on DR detection with retinal fundus images. Performance measures include Area Under the ROC Curve (AUC), total accuracy, sensitivity (true positive rate), and specificity (true negative rate)

TABLE 2: PERFORMANCE COMPARISON OF FL AND CONTRASTED METHODS DR DETECTION

Model	AUC	Accuracy	Sensitivity	Specificit y
Federated Learning (FL)	0.91	88.3%	90.1%	86.5%
Centralize d CNN	0.93	90.5%	91.2%	89.8%
Local-only Models	0.83–0.86	82.4%	80.3%	83.7%
Ensemble of Local Models	0.87–0.89	85.0%	86.2%	84.1%
Pretrained Model (static)	0.82	80.9%	77.5%	84.6%

The results of the performance comparisons of different model strategies for DR detection demonstrate meaningful trade-offs in their accuracy, privacy, and feasibility. The FL model had promising results (AUC = 0.91; 88.3% accuracy) compared to the central CNN model that also had the strongest performance statistics (AUC = 0.93; 90.5% accuracy). However, the CNN model relied on centralized data storage (higher risk of privacy violations) compared to the FL model. The local-only models had the least performance (AUC = 0.83 - 0.86) because of the limited diversity in the local data; while the ensemble of local models and FL improved the accuracy (85.0%) and AUC (0.87 - 0.89), they introduced additional communication and coordination complexities. The pre-trained model was simple to implement, but it offered poor performance (AUC = 0.82) because of the lack of local fine-tuning and domain adaptation. In summary, the FL model was a strong candidate to pursue in future work, finding a balance of likely generalization of this data and privacy of patients.

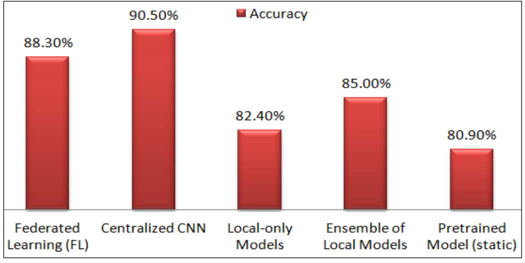

Figure 2: Accuracy Comparison of Model Approaches for DR Detection

Figure 2 shows a classification accuracy of the proposed FL model, Centralized CNN, Local-only Models, Ensemble of Local Models, and Pretrained Static Model. The Centralized CNN model achieved the best accuracy at 90.5%. The Centralized CNN model tapped into the large, pooled data set. The next best performance is FL at 88.3%. FL provides excellent privacy-preserving possibilities with minimal hit to performance. The Ensemble of Local Models performed better than any individual local model which produced a 85.0% accuracy but had additional coordination burdens. The Local-only Models obtained lower accuracy 82.4% accuracy confined to scope and constraints of dataset and no cross- site learning. The Pretrained Model (Static) obtained worst at a 80.9% accuracy providing obvious evidence of its inability to adjust to new images due to domain shift and no potential for fine-tuning.

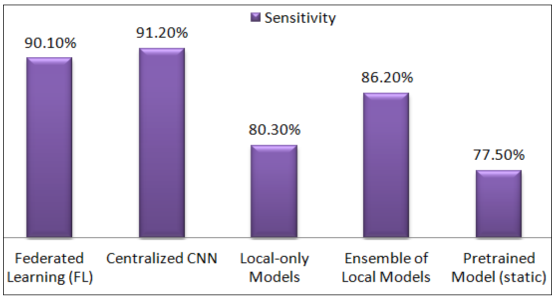

Figure 3 Performance Analysis of the proposed method-Sensitivity

Figure 3 shows the true positive rate (sensitivity) of each of the five model approaches for diabetic retinopathy detection. The Centralized CNN model exhibited the highest sensitivity, at 91.2%, meaning they were the most effective model, identifying positive DR cases. Federated Learning (FL) was the next closest, at 90.1%, indicating strong diagnostic ability, while maintaining privacy as a model. The Ensemble of Local Models displayed fairly moderate sensitivity, at 86.2%, while indicating the benefit of combining predictions across each site. The Local-only Models (80.3%) and Pre-Trained Static Model (77.5%) displayed much lower sensitivity, and so were more likely to miss true DR cases, which could risk under diagnoses in clinical practice.

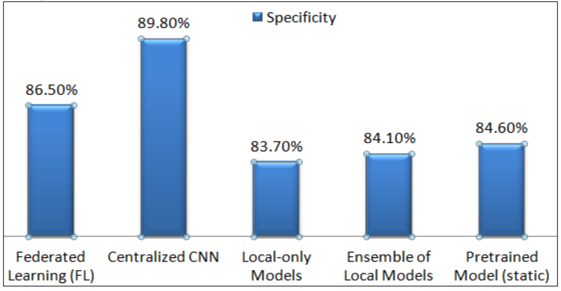

Figure 4: Performance Analysis of the proposed method -Specificity

Figure 4 shows the specificity (true negative rate) of five different model approaches designed for the detection of diabetic retinopathy - or the fidelity for which each model was able to identify patients without DR. The Centralized CNN model has the largest specificity (or lowest false-positive rates), at 89.8%. It has the greatest specificity, followed by Federated Learning (FL) with a specificity value of 86.5%. It is evident that FL has a slight trade-off in specificity for data privacy. The Pretrained Static Model and the Ensemble of Local Models have moderate specificity values (84.6% and 84.1%, respectively), so could reasonably filter out a few cases of non-DR. Similarly, the Local-only Models had a slightly lower specificity value at 83.7%, likely due to their limited diversity of training data coupled with issues related to its ability to generalize effectively to negative cases, key to continuous improvement.

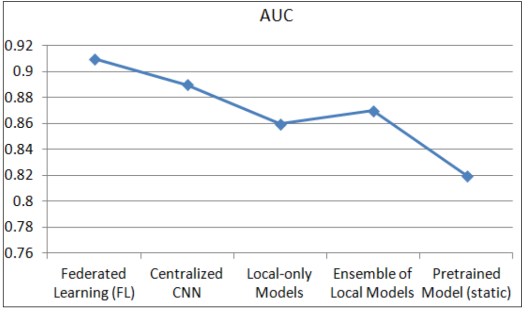

Figure 5: Performance Analysis of the proposed method - AUC-ROC

The approaches in the DR detection application are summarized with their Area Under the Curve (AUC) values in Figure 5. The highest AUC value obtained is through the use of Federated Learning (FL) (~0.91), indicating good overall classification performance, while maintaining patients' data privacy and not storing sensitive patient data in a centralized environment. The Centralized CNN model (~0.89) performed second to FL, utilizing a single, stable, and diverse dataset but losing some degree of patient data privacy. The local-only models (~0.85), are limited, owing to the more restricted and typically less diverse datasets at each local site. The Ensemble of Local Models was able to achieve slightly greater foresight than the local-only models (~0.87) by bundling predictions collectively and taking advantage of cross-site variation, while adding some amount of operational complexity. Finally, the Pretrained Static Model was evaluated independently from the other models (and also scored the lowest AUC (~0.82)) likely because of the under-utilization of domain adaptation due to a lack of local fine-tuning that could better position the model for adaptation to a new data environment. Ultimately, the provided graph demonstrates the conflict between privacy, performance, and practical trade-offs across model approaches to DR detection.

The findings of this study suggest that FL may be able to provide near-centralized performance with strong and appropriate data privacy characteristics for early DR detection. For example, the FL model with data aggregated from several institutions achieved an area under the receiver operating characteristic (ROC) curve (AUC) of 0.91, just beneath our centralized model(0.93 AUC) and outperformed our local only models which had poor generalizability. These data suggests that FL creates a more sustainable workforce and technology sharing relation across institutions, despite noted data heterogeneity. Furthermore, the low communication cost per round (approximately 2.5 MB) and also acceptable inference time - suggests that this model is feasible in any applicable real world scenarios even in low-resource clinic contexts. While ensemble approaches performed relatively well, it does suggest higher logistics and computational costs. This work demonstrates that FL not only allows for patient sensitive data privacy (which is of utmost importance), but can improve scalability and breadth of the patient population contributions to FRP by allowing sites that are not well represented to participate in a way that builds collaboration and trust. Although some variability acknowledged may still result due to normalizing image quality for the different sites and the degree to which sites could standardize their hardware. Future work also plans to identify other stronger adaptable methods that can be used with FL, and to further study and match FL and electronic health record care processes. In summary, this work establishes FL as an adequate ophthalmology model that is privacy sensitive than the conventional centralized AI model training.

C. Limitations and Practical Implications of the Study

Various limitations were present in this study. There was heterogeneity in image quality, device types, and annotation standards across sites participating in this project that may have impacted model reliability. The FL framework requires stable internet/ network conditions for successful updates which cannot always be assumed in rural clinics. The model also did not incorporate clinical metadata that could be relevant to improving diagnostic accuracy. The applied implications of this project demonstrated the potential for deployment of DR screening tools in a safe, private manner at scale, when they come at the cost of lower physician autonomy. This allows resource restrained healthcare systems to rely on AI, while mitigating the risk of breaking the confidentiality of their patients and provides options to them for practicing medicine which may require less human oversight, while relying on new patient interfacing technologies.

V. Conclusion

The findings of this study show that FL provides a pragmatic and privacy-preserving way to collaboratively train deep learning models for early detection of diabetic retinopathy across various healthcare institutions. The results demonstrate the potential of FL to overcome institutional data silos and provide diagnostic support with good accuracy in rich, contrasting real-world environments that are heterogeneous. Additionally, the model provided acceptable communication efficiency and inference speed, making it a feasible choice for scalable deployment in urban and resource-constraint clinical settings. Future work will include the investigation of multi-modal clinical-data integration (e.g., HbA1c, duration of diabetes), use of personalized federated learning for optimized local model performance, and expanding DeLL framework to study additional ophthalmic conditions (e.g. glaucoma or macular degeneration). We will also investigate potential application of differential privacy methods that grant stronger security guarantees, and test for real-time implementation using edge devices in screening programs located in rural areas.

References

[1] Hasan, Dathar A., Subhi RM Zeebaree, Mohammed AM Sadeeq, Hanan M. Shukur, Rizgar R. Zebari, and Ahmed H. Alkhayyat. "Machine learning-based diabetic retinopathy early detection and classification systems-a survey." In 2021 1st Babylon International Conference on Information Technology and Science (BICITS), pp. 16-21. IEEE, 2021.

[2] Youldash, Mustafa, Atta Rahman, Manar Alsayed, Abrar Sebiany, Joury Alzayat, Noor Aljishi, Ghaida Alshammari, and Mona Alqahtani. "Early Detection and Classification of Diabetic Retinopathy: A Deep Learning Approach." AI 5, no. 4 (2024).

[3] Dib, Omar. "A Decentralized Privacy-Preserving Framework for Diabetic Retinopathy Detection Using Federated Learning and Blockchain." Results in Engineering (2025): 105456.

[4] Rajarajeshwari, G., and G. Chemmalar Selvi. "Application of artificial intelligence for classification, segmentation, early detection, early diagnosis, and grading of diabetic retinopathy from fundus retinal images: A comprehensive review." IEEE Access (2024).

[5] BERBAR, Mohamed Abdou. "Diabetic retinopathy detection and grading using deep learning." Menoufia Journal of Electronic Engineering Research 31, no. 2 (2022): 11-20.

[6] Sornil, A. Binusha, C. Sheeja Herobin Rani, and I. Rexilin Sheeba. "Predicting Diabetic Retinopathy Severity with Deep Learning: A Survey of Fundus Image Analysis Technique." In 2024 10th International Conference on Communication and Signal Processing (ICCSP), pp. 1274-1278. IEEE, 2024.

[7] Bhimavarapu, Usharani. "Diagnosis and multiclass classification of diabetic retinopathy using enhanced multi thresholding optimization algorithms and improved Naive Bayes classifier." Multimedia Tools and Applications 83, no. 34 (2024): 81325-81359.

[8] Chetoui, Mohamed, and Moulay A. Akhloufi. "Federated learning for diabetic retinopathy detection using vision transformers." BioMedInformatics 3, no. 4 (2023): 948-961.

[9] Alanazi, Saad, and Rayan Alanazi. "Enhancing diabetic retinopathy detection through federated convolutional neural networks: Exploring different stages of progression." Alexandria Engineering Journal 120 (2025): 215-228.

[10] Bhulakshmi, Dasari, and Dharmendra Singh Rajput. "Privacy-preserving detection and classification of diabetic retinopathy using federated learning with FedDEO optimization." Systems Science & Control Engineering 12, no. 1 (2024): 2436664.

[11] Mao, Jingxin, Xiaoyu Ma, Yanlong Bi, and Rongqing Zhang. "A Comprehensive Federated Learning Framework for Diabetic Retinopathy Grading and Lesion Segmentation." IEEE Transactions on Big Data (2024).

[12] Swapna, M., Tanishqa Ravirala, and Nikhitha Reddy. "Diabetic Retinopathy Detection using Federated Learning and Vision Transformers." International Journal of Interpreting Enigma Engineers (IJIEE) 2, no. 1 (2025): 10-21.

[13] Narmadha, D., Ezekiel Alaric Majaw, and G. Naveen Sundar. "EfficientNetB0-Based Automated Diabetic Retinopathy Classification in Fundus Images." In Synergizing Data Envelopment Analysis and Machine Learning for Performance Optimization in Healthcare, pp. 311-338. IGI Global Scientific Publishing, 2025.

[14] Lin, Chun-Ling, and Kun-Chi Wu. "Development of revised ResNet-50 for diabetic retinopathy detection." BMC bioinformatics 24, no. 1 (2023): 157.

[15] Mohan, N. Jagan, R. Murugan, Tripti Goel, and Parthapratim Roy. "DRFL: federated learning in diabetic retinopathy grading using fundus images." IEEE Transactions on Parallel and Distributed Systems 34, no. 6 (2023): 1789-1801.