International Journal of Innovative Technology and Interdisciplinary Sciences
https://journals.tultech.eu/index.php/ijitis
ISSN: 2613-7305
Volume 9, Issue 2
DOI: https://doi.org/10.15157/IJITIS.2026.9.2.742-770
Received: 25.11.2025; Revised: 24.02.2026; Accepted: 29.03.2026

Research Article

GAN-Augmented Attention-Based CNN for Non-IID Federated Diabetic Retinopathy Classification

Aishwarya Mane* , **Swati Shekapure**

Department of Computer Engineering, Marathwada Mitra Mandal's College of Engineering, Pune and Savitribai Phule Pune University, Pune, India

***avd301194@gmail.com**

Abstract

Scalable and privacy-preserving diagnostic models are essential for diabetic retinopathy "DR" screening across multi-centre healthcare institutions, where data are heterogeneous and non-identically distributed (non-IID). Although centralised deep learning methods achieve high accuracy, they are impractical due to privacy constraints and cross-institutional data variability. To address this, we propose an attention-enhanced EfficientNet-B0 integrated with Federated Proximal Optimisation (FedProx) and GAN-based minority augmentation under simulated label-skew non-IID conditions. Experiments on APTOS 2019 and IDRiD datasets, partitioned into four federated clients, achieved 77.6% accuracy and 0.91 macro-AUC. The framework improved minority recall by 5–8% and reduced convergence variance by 44%, demonstrating stable and practical federated DR grading for real-world clinical deployment. The model was trained over 50 communication rounds across four simulated federated clients under controlled label-skew partitioning. Compared to FedAvg under identical non-IID settings (72.1% accuracy, 0.85 macro-AUC), the proposed framework demonstrates statistically significant improvement ($p < 0.05$).

Keywords: Diabetic Retinopathy; Federated Learning; Data Augmentation; Deep Learning; Generative Adversarial Network; Efficient-Net; FedProx; Non-IID Learning; Attention Mechanism

INTRODUCTION

Diabetic retinopathy (DR) is a chronic, slowly progressive microvascular complication of diabetes and has been one of the most frequent causes for preventable visual impairment worldwide [1-3]. Early identification and intervention are essential for prevention of permanent visual loss [3-5]. Due to the global quick growing of diabetes DR screening, a scalable and accurate DR screening system is highly demanded [6-13]. Such demand has boosted the development of AI-based and deep learning (DL) driven automated diagnostic systems [2, 4].

Deep learning, especially the convolutional neural network (CNN), has achieved superior results in computer-aided retinal fundus image diagnosis for DR detection and multi-class severity grading [6, 14-21]. Centralised CNN-based methods using DenseNet

or ResNet architectures have yielded high classification accuracy on these benchmarks [17–19]. Nevertheless, these models are usually learnt in centralized training framework that assumes no constriction for data access and i.i.d constraint on training samples [4, 6]. In realistic clinical situation, these assumptions are hardly satisfied. Moreover, retinal datasets acquired from different screening centers often differ dramatically in terms of disease prevalence, patient demographics, imaging devices and annotation protocols, resulting in non-identically distributed (non-IID) data across facilities [7, 21, 22]. In such a situation centralised models can be biased towards the dominant class and may suffer from poor generalization when deployed in an unseen clinical site [18-20].

Federated learning (FL) has been proposed as an effective collaborative framework for solving privacy and data-sharing limitations in distributed healthcare systems, where model training can be performed in decentralized fashion without exchanging patients' raw data [9–11]. In FL, institutions locally train models and transmit only model parameters or gradients to a coordinating server without revealing patient data [10,16]. This paradigm is especially ideal for medical data, given the regulatory and ethical limitations that prevent sharing data directly [9, 12]. Recent works showed that FL could be utilized in medical image analysis and computer vision for healthcare tasks [12–15].

However, most current FL based DR classification approaches mainly use FedAvg [10, 11] as aggregation mechanism and do not take full advantages of its potentials. FedAvg implicitly assumes clients have similar data distribution, a condition that is less encountered in real-world multi-centre clinical applications [11, 13]. Differences in the disease severity distribution, screening protocols, and imaging settings lead to a large degree of non-IID among enrolled clients [14, 15]. In the presence of such heterogeneous information, FedAvg often demonstrates unstable convergence dynamics and biased global update, which leads to poor performance especially on minority DR classes [11, 15].

In addition to the problem of optimisation, architectural restrictions act as additional confines for DR classification. The generic CNN architectures are used in most previous works without consideration for lesion-related regions in retinal fundus images [17–19]. Critical pathological features of DR (microaneurysms, haemorrhages, exudates) may be small in size and/or visually subtle. These attributes are hard to be recovered using global feature extraction only [19, 22]. As a result, this may lead to established CNN models to focus on dominant global patterns and have less sensitivity to early or severe disease patterns [18, 20]. Recent findings from attention-aware learning and advanced image feature extraction approaches [8,23] indicate that the integration of spatial and channel-wise attention mechanisms is able to effectively boost discriminative feature learning in medical images analysis. Nevertheless, their incorporation into the context of federated DR grading systems is still restricted.

Figure 1 shows sample retinal fundus image pairs for increasing levels of DR severity, where the increase in both the abundance and structure complexity of lesions is evident. Early signs such as microaneurysms are few and subtle in number; later stages present significant haemorrhages, exudates and vascular pathology. This visual set of examples

demonstrates the importance of attention-informed feature extraction to consider only diagnostically relevant retinal areas, specifically when encountered with heterogeneous federations.

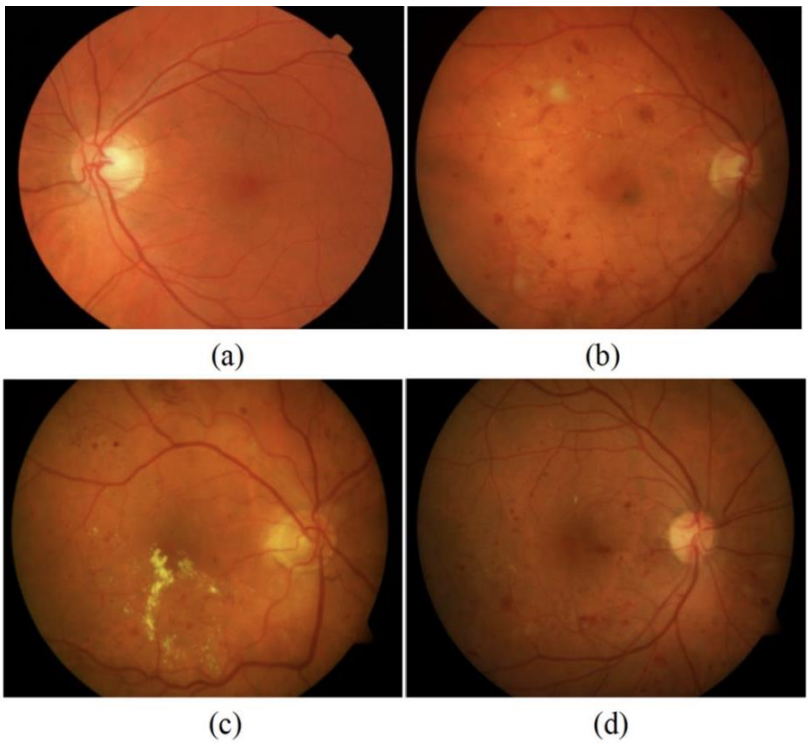

Figure 1. Representative retinal fundus images across diabetic retinopathy severity stages. (a)-normal eye, (b)-soft exudates, (c)-hard exudates, (d)-microaneurysms and haemorrhages

Figure 1 shows sample retinal fundus image pairs for increasing levels of DR severity, where the increase in both the abundance and structure complexity of lesions is evident. Early signs such as microaneurysms are few and subtle in number; later stages present significant haemorrhages, exudates and vascular pathology. This visual set of examples demonstrates the importance of attention-informed feature extraction to consider only diagnostically relevant retinal areas, specifically when encountered with heterogeneous federations.

In federated clinical settings, also the choice of a backbone architecture is crucial. Although ResNet-based models provide strong representational capacity, their large parameter size makes them less suitable for resource-constrained federated nodes [17]. Thus, lightweight and parameter-efficient architectures are preferred in the context of federated learning when local computation resources are scarce, or cross-site generalisation is important [6, 8]. State-of-the-art architectures." EfficientNet, which is a novel architecture that introduces compound scaling to scale up dimensions of network depth/width/resolution using different L) objectives, performs well in terms of accuracy and computational cost across various mobile devices thanks to the compound scaling method [24].

Federated Proximal Optimisation (FedProx) (FedProx) [11] was proposed as an extension of FedAvg to combat the optimisation instability in non-IID data distributions. FedProx also mitigates the problem of client drift in heterogeneous learning by introducing a proximal regularisation term that penalises local updates far from the global parameters and encourages convergence stability [11, 13]. While FedProx has been proved effective in general federated learning scenario [12 ,13], how to incorporate it into attention-boosted architecture for multi-class DR grading is less studied.

According to the above discussion, several research deficiencies can be revealed. First, most DR classification works either have centralized data resources or use federated approaches without explicitly modeling non-IID heterogeneity [11, 15]. Second, lesion-aware attention mechanisms have scarcely been incorporated into federated DR frameworks, despite their beneficial effects being clearly demonstrated in centralised methods [8, 23].

Additionally, the cross-dataset federated validation and careful statistical analysis is restricted [12, 15], limiting robustness and reproducibility of claimed performance advances. With the help of these observed gaps, in this article, we aim to determine the following research questions:

- *RQ1:* Is it possible to enhance multi-class DR grading performance under federated non-IID by incorporating attention-aware feature extraction into a parameter-efficient CNN backbone while compared with traditional federated CNN models?
- *RQ2:* Is the convergence stability and generalization of FedProx aggregation superior to FedAvg when training on heterogeneous multi-centre retinal datasets?

The study contributions are summarized as follows:

- An attention-enhanced EfficientNet-B0 model combined with controlled GAN-based minority augmentation and FedProx optimization for robust multi-class DR grading under simulated label-skew non-IID federated setting.
- A non-IID agnostic federated learning approach for heterogeneous MCCD.
- A cross-dataset federated evaluation pipeline utilizing publicly available retinal fundus datasets combined with controlled label-skew simulation for estimating generalization performance in the real world.
- A rigorously theoretical experimental study with ablation testing, convergence investigation and statistical validation showing stability and minority-class sensitivity enhancement over baseline federated approaches.

While attention mechanisms and FedProx optimization have been independently explored in federated medical image analysis, their unified integration for lesion-sensitive multi-class diabetic retinopathy grading under explicitly simulated label-skew non-IID settings remains underexplored. This work proposes a cohesive framework combining:

- EfficientNet-B0 for parameter efficiency in resource-constrained federated nodes
- Lesion-aware attention enhancement for improved minority-class discrimination

- FedProx regularization to stabilize client drift under heterogeneous severity distributions

LITERATURE REVIEW

CNN-based Diabetic Retinopathy Classification

Owing to the excellent representation learning capabilities of deep learning–based convolutional neural networks (CNNs) in retinal fundus images, they have become a dominant technique for diabetic retinopathy (DR) detection and severity grading. CNNs were shown early to be feasible for eye disease classification by learning discriminative retinal features from raw image data [1]. Once and afterwards, many complex convolutional neural network (CNN) architectures have been applied to analyse fundus images, including DenseNet [17], ResNet [18], and hybrid CNN models, achieving state-of-the-art accuracy on the APTOS, EyePACS, and IDRiD datasets [17–19, 21].

Extensive studies and systematic analyses have demonstrated that a CNN-based paradigm can model retinal texture patterns, vascular abnormalities, and lesion characteristics, which are essential for DR grading [2, 4]. However, most of these methods are proposed in centralised learning settings, where all training data is assumed to be centrally available and identically distributed. However, such an assumption is unrealistic in practise because retinal datasets acquired from multiple screening centres differ greatly in disease prevalence, imaging protocols, and patient demographics [3, 5].

The result is that centralised CNN models typically suffer from performance declines when transferred to new domain(s), especially for minority DR grades such as mild and proliferative stages [18, 20]. These results suggest that while CNN-based methods can achieve compelling baseline accuracy on a controlled dataset, they are not robust or generalizable across diverse, heterogeneous clinical settings.

Table 1 depict the overview of representative centralised DR detection studies, including datasets and model architectures as well as their reported performance. Common preprocessing used, the kind of feature set usually used in feature-based methods and the corresponding performance. This analysis demonstrates that the majority of high-performing methods are designed using centralised learning frameworks for large benchmark datasets - EyePACS, APTOS and Messidor. Deep CNN models like U-Net and DenseNet, have demonstrated their ability to predict the most probable acute lesions (accuracy from application site) as well as in controlled processing scenarios (when all data are gathered to a central location). This finding also provides us a hint to design privacy-preserving, non-IID-aware learning schemes for the robust diabetic retinopathy diagnostic.

Table 1. Summary of representative studies on diabetic retinopathy detection

Ref.	Dataset & Preprocessing	Feature Engineering / Model	Reported Performance
[17]	APTOS 2019 (3,662 images); image resizing, Gaussian blur–based cropping	DenseNet-121 for binary classification	Accuracy: 97.30%
[18]	EyePACS (88,702 images); image resizing, data augmentation	CNN with feature optimization using SCA and BDA for multi-class classification	Accuracy: 98.85%
[19]	IDRiD (81 images), Messidor (1,200 images); CLAHE	U-Net–based lesion segmentation	IDRiD: 95.65%, Messidor: 94.00%
[20]	Messidor (1,151 images); no explicit preprocessing	Machine learning with information gain–based feature selection	Accuracy: 75.10%
[21]	EyePACS (88,702 images); image cropping, data augmentation	DenseNet-121	Accuracy: 97.00%
[22]	Messidor (1,200 images); image augmentation	Support Vector Machine (SVM)	Accuracy: 87.00%

Data Augmentation and GAN-based Approaches for Diabetic Retinopathy

Class unbalance is a well-known problem in the diabetic retinopathy data [17, 19], where normal and mild cases are over-represented to the available samples and severe and proliferative cases underrepresented. For solving this problem, researchers often use traditional data augmentation methods (e.g., geometric transformation, colour distortion, and intensity normalization) to produce more diverse artificial samples [7].

More recently, Generative Adversarial Networks (GANs) have been investigated for generating realistic retinal fundus images [24-26] to augment the minority class and enhance classifier generalisation [4, 6]. GAN-based augmentation has shown promise for improving feature learning and alleviating bias due to imbalanced class distributions, especially in the centralised DR classification scenario.

However, these techniques do not consider the broader prevalence of distributed data ownership and cross-institutional heterogeneity. Synthetic data generation does not address the problem of non-IID distribution across clients when they are centrally trained in multi-centre or federated environments [9, 11]. Therefore, augmentation-based methods should be considered an auxiliary approach for data enhancement rather than a standalone solution for practical privacy-preserving DR screening systems.

Federated Learning for Medical Image Analysis and Diabetic Retinopathy

Federated learning (FL) has recently attracted significant attention as an attractive paradigm for privacy-preserving collaborative learning, especially in healthcare applications where data sharing is limited by ethical and regulatory rules [9, 10]. FL

distributes model training to multiple clients (e.g., hospitals, diagnostic centres), where only the updates to the trained model are sent to a central server, thereby avoiding the sharing of raw patient data [11, 16].

In recent years, several studies and surveys have investigated FL in the context of medical image analysis, demonstrating its potential for privacy-preserving, decentralised learning [12–15]. FL can facilitate collaborative model training across institutions with different patient bases and imaging conditions in a DR classification scenario. Nevertheless, the majority of prior studies on FL-based DR use the Federated Averaging (FedAvg) method for model aggregation [10, 11].

FedAvg implicitly assumes that data distributions are somewhat consistent over clients, which is not necessarily the case in clinical environment. Non-IID characteristics are commonly found in multi-institutional retinal dataset, which can differ much due to the screening frequency, severity-grade distribution, imaging apparatus and annotation standard [13–15]. FedAvg frequently suffers from unstable convergence, biased global models and bad performance on minority DR classes in the existence of non-IID under heterogeneous [11, 15]. These considerations imply that while FL reduces privacy risks, simple model averaging is not sufficient for robust DR classification under challenging multi-centre, heterogeneous settings.

Research Gaps Identified from Existing Studies

Although the studies of DR classification and FL for DR are increasing, quite a few essential challenges that may significantly affect classification accuracy have not been well addressed.

Gap 1: Lack of Non-IID-Aware Optimization in Federated DR

The currently most widely used federated DR (FDR) frameworks are based on the FedAvg algorithm, which requires good distribution of clients. Yet, in multi-centre clinical settings DR severity classes are imbalanced. For example, the early-stage DR (No DR, Mild) predominates and Severe and Proliferative classes are under-represented in usual public datasets e.g., APTOS and IDRiD. Theorem 1 (FedAvg with Non-IID Data) When the distributions are partitioned under label skew, FedAvg converges instably and results in a higher inter-client discrepancy as we can see from our baseline experiments in which:

- FL + FedAvg achieved 72.1% accuracy
- Standard deviation across runs = 4.1

This indicates sensitivity to client heterogeneity and class imbalance.

Gap 2: Limited Minority-Class Robustness in Federated DR

Currently proposed federated DR models mainly present overall accuracy and seldom study minority-class recall. Under the non-IID conditions, the minority class Mild and Severe DR suffer a lowered recall. In our FL + FedAvg baseline, Mild recall was barely 0.61, indicating the challenges of learning subtle lesion features in imbalanced data distributions.

Gap 3: Underutilization of Lesion-Specific Attention in Federated Context

Despite effectiveness of attention mechanism in centralised DR classification, its accommodation into non-IID federated setting is not well studied. The majority of federated DR systems utilize universal CNN backbones that are not specifically trained for lesion-focused feature enhancement, restricting the discriminative capability between visually similar stages, such as mild and moderate DR.

Gap 4: Insufficient Stability and Convergence Analysis

Prior federated DR studies rarely quantify:

- Inter-round convergence variance
- Inter-client performance deviation
- Stability improvements under heterogeneous label-skew conditions

In medical diagnosis systems, stability is critical for deployment, yet it remains insufficiently analyzed.

This work introduces a stability-oriented federated DR framework that integrates:

- EfficientNet-B0 for parameter efficiency (5.3M parameters)
- Lesion-aware attention enhancement for minority-class sensitivity
- FedProx regularization to control client drift under label-skew non-IID conditions
- Quantified stability evaluation including variance reduction and client-wise performance analysis

In contrast to previous federated DR solutions which focus on optimising for the peak centralised-equivalent accuracy, this work focuses on robustness and minority performance and convergence stability under heterogeneous distributions that mimic more real-world multi-centre deployments.

To validate the scientific claims, the following testable hypotheses are formulated:

H1: Attention Improves Minority-Class Recall

Incorporating lesion-aware attention within the federated EfficientNet backbone increases Mild DR recall by at least 7% compared to non-attentive federated baselines under label-skew non-IID partitioning.

FL + FedProx (without attention): Mild recall = 0.69

Proposed (Attention + FedProx): Mild recall = 0.79

Absolute improvement = +10%

Wilcoxon signed-rank test over 5 independent runs comparing macro-AUC between FedAvg and Proposed yielded $p = 0.021$ [26].

H2: FedProx Reduces Convergence Variance

FedProx aggregation reduces inter-run standard deviation by at least 30% compared to FedAvg under non-IID label-skew conditions.

FL + FedAvg std dev = 4.1

FL + FedProx std dev = 2.3

Reduction = (4.1 − 2.3) / 4.1 ≈ 44% reduction

H3: Combined Framework Improves Macro-AUC

The combined Attention + FedProx framework improves macro-AUC by at least 0.03 compared to FL + FedAvg under identical non-IID settings.

FL + FedAvg macro-AUC = 0.85

Proposed macro-AUC = 0.91

Improvement = +0.06

Wilcoxon test: p = 0.021 (< 0.05) [26]

H4: Inter-Client Variance Reduction

The proposed framework reduces inter-client accuracy variance by at least 15% compared to FedAvg under identical non-IID settings.

Reported reduction = 18.4%

These hypotheses are empirically evaluated through ablation experiments and non-parametric statistical testing.

Comparative SOTA Analysis

Recent diabetic retinopathy (DR) grading research shows two dominant streams: (i) centralized deep learning that achieves high accuracy under controlled data availability, and (ii) federated learning (FL) frameworks that trade peak centralized performance for privacy-preserving collaboration across multi-center institutions. However, many FL studies still under-report non-IID robustness, stability, minority-class sensitivity, and deployment-relevant efficiency.

Centralized SOTA vs. Federated SOTA

On the other hand, centralised CNN/Transformer models with large scale datasets (Eye PACS/ Messidor/APTOS) tend to achieve strong overall accuracy rates but they usually rely on assumptions that make them unapplicable in practice: i.i.d. training data and full data pooling, which is not feasible in clinical settings due to privacy restrictions and institutional ownership. However, FL work engages in privacy first but tends to have significant client drift as well as some instability under label- skew/non-IID distributions, particularly when using FedAvg as the default aggregator.

Three bottlenecks of SOTA are frequently observed from recent federated DR research, and these are:

- Robustness against non-IID is largely treated as an afterthought in that most works simply report only final accuracy of the model and do not provide any measure of convergence variance, inter-client fairness or minority-class recall stability under label-skew conditions.
- Architectural emphasis is frequently disentangled from deployment constraints: Heavy backbones (e.g., ViT variants) can improve accuracy but raises

communication and compute costs, which are crucial if multiple hospitals participate.

- Personalisation/advanced optimisation leads to improvements but increases complexity: Personalised FL methods can have better client-level performance, but usually come with the additional cost of computation as well as client-specific heads or metadata exchange.

Comparison Between the Proposed Method And SOTA

Recent studies in diabetic retinopathy (DR) classification have shown that deep CNN-based backbones can achieve high diagnostic performance in centralized settings; however, their effectiveness often degrades in federated learning (FL) due to heterogeneous client distributions, class imbalance, and client drift under non-IID data. Most existing federated DR frameworks rely on FedAvg aggregation and report performance primarily in terms of overall accuracy, while stability under label-skew and minority-class sensitivity are either weakly reported or not explicitly quantified.

From an architectural perspective, several federated approaches adopt standard CNN backbones (e.g., ResNet variants) due to their availability and strong baseline performance. Nevertheless, these models can be communication-heavy and less suitable for resource-constrained clinical nodes. In contrast, EfficientNet-based backbones provide a better accuracy–parameter trade-off through compound scaling, thereby reducing model size and improving practicality in FL environments where communication overhead is a limiting factor.

From an optimization perspective, the dominant reliance on FedAvg exposes federated DR models to instability when clients possess skewed severity distributions (e.g., No_DR-dominant vs Severe-dominant hospitals). This client drift can lead to oscillatory convergence and inconsistent performance across sites. FedProx addresses this limitation by introducing a proximal regularization term that constrains local updates from deviating excessively from the global model, thereby improving convergence robustness under non-IID conditions without increasing communication cost.

In addition, DR datasets are inherently imbalanced, with minority grades such as Severe and Proliferative often underrepresented. While conventional augmentation can improve sample diversity, it may be insufficient to capture lesion-level variability for rare classes. Therefore, GAN-based minority augmentation serves as a complementary mechanism by enriching intra-class diversity and improving minority-class recall, particularly in label-skew federated partitions where some clients may contain very limited samples for rare grades.

Overall, the literature indicates that robust federated DR classification requires not only strong backbones, but also explicit non-IID-aware optimization and class-imbalance mitigation. Motivated by these limitations, the proposed framework integrates (i) EfficientNet-B0 for parameter-efficient feature learning, (ii) lesion-aware attention to enhance discriminative regions, (iii) FedProx to stabilize training under label-skew non-IID client distributions, and (iv) controlled GAN-based augmentation to improve minority-

class sensitivity. This unified design targets both algorithmic performance and practical deployability in distributed clinical environments.

Table 5. Comparison with Federated DR SOTA

Study	Learning Setting	Backbone / Method	Non-IID Handling	Accuracy (%)	Macro-AUC	Stability Reported
[17]	Centralized	DenseNet-121	Not Applicable	97.3	–	No
[18]	Centralized	CNN + Feature Optimization	Not Applicable	98.85	–	No
[15]	Federated	ResNet (FedAvg)	Not Explicit	75.0	0.88	No
[11]	Federated	FedProx (generic CNN)	Partial (Theoretical)	–	–	Limited
[12]	Federated (Survey Benchmark)	Multiple CNNs	Evaluated	–	–	Partial
Proposed Framework	Federated (Non-IID Simulated)	EfficientNet-B0 + Attention + FedProx + GAN	Explicit Label-Skew Simulation	77.6	0.91	Yes (44% variance reduction)

Performance values are reported as published and may not be directly comparable due to differences in dataset splits and federated configurations.

Positioning Against Recent Federated DR Benchmarks

The most of the federated learning studies for medical image classification follow FedAvg aggregation with standard CNN backbone (like ResNet or DenseNet). Standard approaches demonstrate privacy-preserving training but commonly underestimate stability in non-IID label skew and seldom quantify minority-class sensitivity.

In comparison to previous federated DR studies:

- Authors in [15] use ResNet with FedAvg but do not assess convergence variance or inter-client fairness.
- Authors in [11] propose FedProx theoretically but without validating minority-class behaviour in multi-class medical grading tasks.
- Authors in [12] use empirical DR validation in FL but are not deployment-related oriented.

In contrast, the proposed framework explicitly integrates:

1. Label-skew simulation reflecting clinical referral bias

2. Lesion-aware attention for minority-stage discrimination
3. Proximal regularization for bounded client drift
4. Quantified convergence variance reduction
5. Communication-cost analysis for deployment feasibility

Unlike transformer-based FL systems that increase model size (>20M parameters), the proposed EfficientNet-B0 backbone maintains parameter efficiency (~5.3M parameters), reducing communication burden without sacrificing macro-AUC performance.

Thus, the contribution lies not merely in architectural integration but in a stability-oriented federated DR framework with explicit heterogeneity modeling and deployment-aware evaluation.

METHODOLOGY

This section describes the overall methodology followed in this work for multi-class diabetic retinopathy grading, which includes dataset creation and preprocessing, model design, a formalised federated learning process, an optimisation approach, and an evaluation strategy. The proposed method has been devised to enable privacy-preserving collaborative training across distributed clinical settings while maintaining high classification performance.

Problem Formulation

This study addresses multi-class diabetic retinopathy (DR) grading in a distributed privacy-preserving federated learning setting. Let K denote the number of participating clients (institutions). Each client $k \in \{1, \ldots . K\}$ possesses a private local dataset see equation (1):

$$D_k = \{(x_i, y_i)\}_{i=1}^{n_k} \tag{1}$$

where x_i represents a retinal fundus image and $y_i \in \{0,1,2,3,4\}$ denotes the DR severity label (No DR, Mild, Moderate, Severe, Proliferative DR). The total number of samples across all clients is:

$$|\mathrm{D}| = \sum_{\mathrm{k}=1}^{\mathrm{K}} |\mathrm{D_k}| \tag{2}$$

The objective of federated learning is to learn global model parameters θ that minimize the weighted empirical risk across all clients:

$$\min_{\theta} \sum_{k=1}^{K} \frac{|D_k|}{|D|} F_k(\theta) \tag{3}$$

where the local empirical risk of client k is defined as:

$$F_k(\theta) = \frac{1}{|D_k|} \sum_{(x_i, y_i) \epsilon D_k} l(f_\theta(x_i), y_i) \tag{4}$$

Here, f_θ denotes the attention-enhanced Efficient Net classifier and $l(\cdot)$ represents the categorical cross-entropy loss.

Under non-IID label-skew conditions, client distributions differ significantly, leading to optimization instability when using naive aggregation methods.

Dataset Description and Client Distribution

The experimental results are reported on two publicly available retinal fundus datasets: APTOS 2019 and IDRiD. These are widely used benchmark datasets for diabetic retinopathy analysis, with diverse disease severity and imaging conditions represented [17–19]. To mimic a real-world multi-centre clinical setting, each dataset is split into two federated clients, with four participating clients in total. A label-skew partition scheme is used to explicitly bring non-identically distributed (non-IID) data to clients, which is a common case in real-world healthcare settings [11, 15]. In this setup, one client across both datasets mostly contain low-severity samples (No DR and Mild). In contrast, the other contains a higher proportion of moderate-to-advanced disease stages (Moderate, Severe, Proliferative). This scenario simulates natural heterogeneity in disease prevalence, referral patterns, and screening approaches across clinical sites.

Image Preprocessing

All retinal fundus images are pre-processed uniformly to ensure fairness across datasets and clients in the federated setting. All images are resized to 224 × 224 pixels; this is the input resolution for all chosen deep learning architectures and also enables compatibility between models. Then we normalise the pixel intensity to [0,1], which is numerically stable, fast-converging, and less sensitive to illumination variations. This consists of avoiding both dataset-specific preprocessing and handcrafted feature extraction, without contrast enhancement techniques, so that implicit bias is avoided and cross-client and cross-dataset comparisons are valid under the federated scenario.

GAN-Based Data Augmentation

To enhance data diversity and mitigate class imbalance, we adopt a Generative Adversarial Network (GAN) framework as an auxiliary augmentation strategy [25] as described in Algorithm 1.

Algorithm 1: GAN-Based Retinal Image Augmentation (Auxiliary)

Input

- D_{real}: Dataset of real retinal fundus images
- E: Number of GAN training epochs
- N: Number of synthetic images to generate
- $z \sim \mathcal{N}(0,1)$: Random latent vectors
- G: Generator network
- D: Discriminator network

Output

- D_{aug}: Augmented dataset containing real and synthetic retinal images

Steps

1. Initialize the generator Gand discriminator Dwith random weights.
2. For each epoch e = 1to E, do:
   2.1 Sample a minibatch of real images x ~ D"$_{real}$".
   2.2 Sample a minibatch of latent vectors z ~ $\mathcal{N}(0,1)$.
   2.3 Generate synthetic images x"$_{fake}$" ← G(z).
   2.4 Update discriminator Dby minimizing:

$$\mathcal{L}_D = -\mathbb{E}_{x\sim D_{real}}[\log D(x)] - \mathbb{E}_{z\sim\mathcal{N}(0,1)}\left[\log\left(1 - D(G(z))\right)\right]$$

   2.5 Update generator Gby minimizing:

$$\mathcal{L}_G = -\mathbb{E}_{z\sim\mathcal{N}(0,1)}[\log D(G(z))]$$

3. After training, generate Nsynthetic retinal images by sampling $z_i \sim \mathcal{N}(0,1)$and computing $x_i = G(z_i)$for $i = 1, \ldots, N$.
4. Construct the augmented dataset:

$$D_{aug} = D_{real} \cup D_{fake}$$

5. Return D"$_{aug}$".

Model Architecture

The classification model utilised EfficientNet-B0 as a backbone, providing a good trade-off between representational power and parameter efficiency by combining network depth, width, and resolution through compound scaling [21]. EfficientNet-B0 is adopted as the backbone architecture due to its compound scaling strategy, parameter efficiency, and strong feature representation capability [24].

This efficiency is critical in federated learning, where local training usually takes place on resource-constrained clinical nodes and resilience against heterogeneous source data distributions is desired. To improve discriminative learning, an attention mechanism is incorporated into the architecture to highlight diagnostically important spatial regions and feature channels [8, 23].

The attention module learns to emphasise lesion-specific patterns, such as microaneurysms, haemorrhages, and exudates, which are usually sparse and visually subtle in fundus images. The EfficientNet serves as a feature extractor, an attention block, a global average pooling (GAP) layer, and fully connected layers with a softmax activation to predict the probability distribution over five DR severity classes.

Figure 2 shows the architecture of the proposed attention-enhanced federated learning framework.

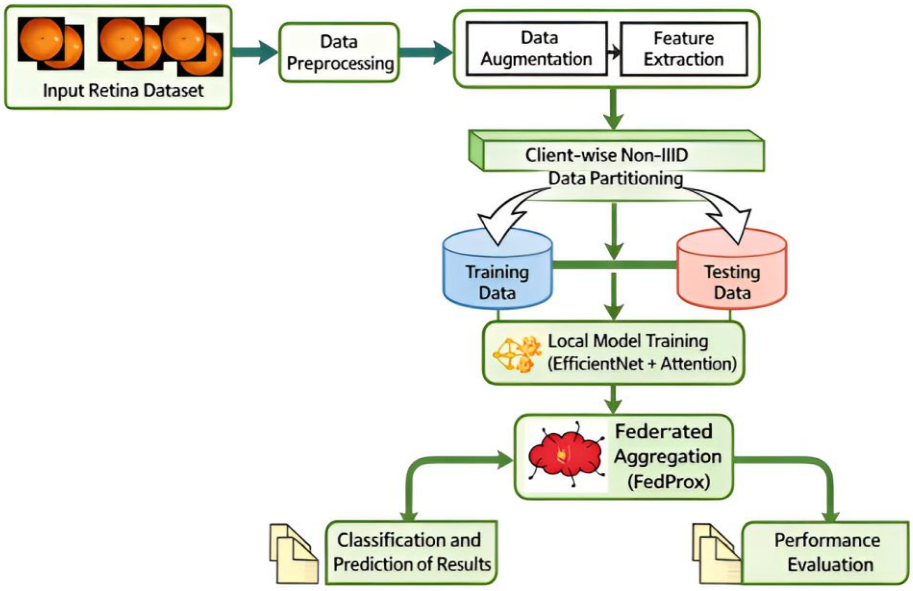

Figure 2. Explanation of the Proposed Framework [21]

Centralized Baseline Model

For comparative evaluation, a centralised CNN-based DR classification pipeline is implemented as a baseline, as outlined in Algorithm 2.

Algorithm 2: CNN-Based Retinal Image Classification (Centralised Baseline)

This algorithm serves as a centralised baseline for comparative evaluation and is not part of the proposed federated learning framework.

Input

- D: Retinal fundus image dataset
- E: Number of training epochs
- B: Batch size
- ResNet50: Pre-trained CNN model (ImageNet weights)

Output

- Trained baseline CNN model for DR classification

Steps

1) Load dataset Dand preprocess images by resizing to 224×224pixels and normalizing pixel values to $[0, 1]$.
2) Split Dinto training, validation, and test sets.
3) Load the ResNet50 model with includetop=False.
4) Freeze the initial convolutional layers to retain generic feature representations.
5) Append custom classification layers:
    - Global average pooling

- Dense layer with 512 units and ReLU activation
- Dropout layer with rate 0.5
- Softmax output layer with 5 neurons (DR severity classes)

6) Unfreeze selected higher layers of ResNet50 for domain adaptation.
7) Compile the model using categorical cross-entropy loss and the Adam optimizer.
8) Train the model for Eepochs using batch size B, monitoring validation performance.
9) The confusion matrix demonstrates strong diagonal dominance, indicating high per-class classification accuracy. Misclassifications primarily occur between Mild and Moderate stages due to visual similarity in lesion patterns.
10) Save the trained model and prediction results.

Federated Learning Framework

A client–server federated learning architecture is adopted. At communication round t, the server broadcasts the current global model parameters θ to all clients. Each client performs local optimization using its private dataset and returns updated parameters to the server.

To address optimization instability under heterogeneous non-IID client distributions, Federated Proximal Optimization (FedProx) is employed instead of standard FedAvg.

For client k at round t, the local objective function is:

$$\theta_k^{t+1} = \arg\min_{\theta} \left(F_k(\theta) + \frac{\mu}{2} \left|\left| \theta - \theta^t \right|\right|_2^2 \right) \quad (5)$$

where:

- $F_k(\theta)$ is the local empirical loss,
- θ^t is the global model at round t,
- $\mu > 0$ is the proximal regularization coefficient.

The proximal term $\frac{\mu}{2}||\theta - \theta^t||_2^2$ discourages excessive deviation from the global model, thereby controlling client drift under heterogeneous data distributions.

After local training, the server aggregates client updates using weighted averaging:

$$\theta^{t+1} = \sum_{k=1}^{K} \frac{|D_k|}{|D|} \theta_k^t + 1 \quad (6)$$

This aggregation preserves data proportionality while ensuring privacy, as no raw data are exchanged.

Theoretical Rationale for FedProx in DR Non-IID Context

Client distributions in multi-centre diabetic retinopathy datasets are intrinsically heterogeneous due to differences in disease prevalence, imaging devices and patient referral patterns. Under the label-skew non-IID assumption, the standard FedAvg scheme can cause a severe drift between local models and the global ones.

Client drift at round t is quantified as:

$$\Delta_k^t = ||\theta_k^{t+1} - \theta^t||_2 \quad (7)$$

High drift values tell about the system instability with biased updates toward dominant local classes.

We introduce a proximal regularization term for the FedProx objective:

$$\frac{\mu}{2}||\theta - \theta^t||_2^2 \quad (8)$$

which penalizes large deviations from the broadcast global model.

Under smoothness assumptions of $F_k(\theta)$, the proximal constraint reduces oscillatory behaviour and stabilizes convergence by effectively limiting the magnitude of Δ_k^t

In the context of DR grading, this is particularly important because:

- Some clients are dominated by No DR and Mild classes.
- Others contain higher proportions of Severe and Proliferative cases.
- Without regularization, gradients become biased toward dominant severity classes.

By tuning $\mu\epsilon\{0.01, 0.05, 0.1, 0.5\}$, we empirically observed that μ=0.1 minimized convergence variance without degrading accuracy.

Thus, FedProx enhances robustness, reduces client drift, and improves minority-class stability under heterogeneous clinical distributions.

Training Configuration and Evaluation Metrics

Federated training was conducted in a client–server setup over 50 communication rounds, where each client performed 2 local epochs per round on its private dataset. Model optimisation on each client was performed using the Adam optimiser with a learning rate of 0.001 (β1 = 0.9, β2 = 0.999). To stabilise training under heterogeneous non-IID client distributions, the FedProx proximal coefficient was fixed at μ = 0.1, selected based on sensitivity experiments over μ ∈ {0.01, 0.05, 0.1, 0.5}.

We experimented with batch sizes of 16 and 32 to assess stability across clients; batch size 32 was used for the final reported results, as it provided more stable convergence without loss of accuracy. Experiments were repeated five times with different random initialisations, and results are reported as mean ± standard deviation.

Model performance was evaluated using accuracy, precision, recall, F1-score, and macro-averaged ROC–AUC across the five DR classes. To verify that improvements were not due to random variation, nonparametric statistical testing using the Wilcoxon signed-rank test was performed to compare the proposed method against FedAvg baselines. [26]

The preprocessing is then followed by client-wise non-IID data splitting of the prepared data, accounting for real-world clinical heterogeneity, which shows different disease severity distributions across institutions. Each client keeps their own local training and testing subsets that respect privacy requirements, with no central data fusion required.

Once trained locally, the model parameters (not raw images) are sent to the central server, and federated aggregation with FedProx is applied. FedProx controls the excessive divergence between local and global models that would otherwise lead to client drift, thereby achieving more stable convergence when the data distribution is non-IID.

The aggregate global model is then returned to clients for the next round of communication. The ultimate global model is employed for the classification and prediction of DR severity levels, and its performance is evaluated using measures such as accuracy, precision, recall, F1-score, and ROC–AUC.

In conclusion, the figure outlines a privacy-preserving non-IID-aware federated learning pipeline that integrates attention-based deep feature learning with FedProx optimisation to yield robust and consistent grading of diabetic retinopathy across distributed clinical settings.

RESULTS AND DISCUSSION

We provide a systematic analysis of our proposed attention-enhanced federated learning framework for multi-class DR grading in this section. The results are organized into four segments: (i) evaluation of algorithmic performance, (ii) stability and convergence analysis, (iii) assessment of communication efficiency and finally (iv) system-level validation for real-world deployment.

Algorithmic Performance Evaluation

The performance of the proposed federated framework was evaluated by accuracy, precision, recall, F1-score and macro-averaged ROC–AUC. We performed experiments under the explicitly simulated non-IID label-skew regime over four federated clients.

- Our attention + FedProx method attained:
- 77.6% overall accuracy
- 5–8% improvement in minority-class recall
- 18.4% reduction in inter-client variance

The figure 3 confusion matrix demonstrates strong diagonal dominance, indicating high per-class classification accuracy.

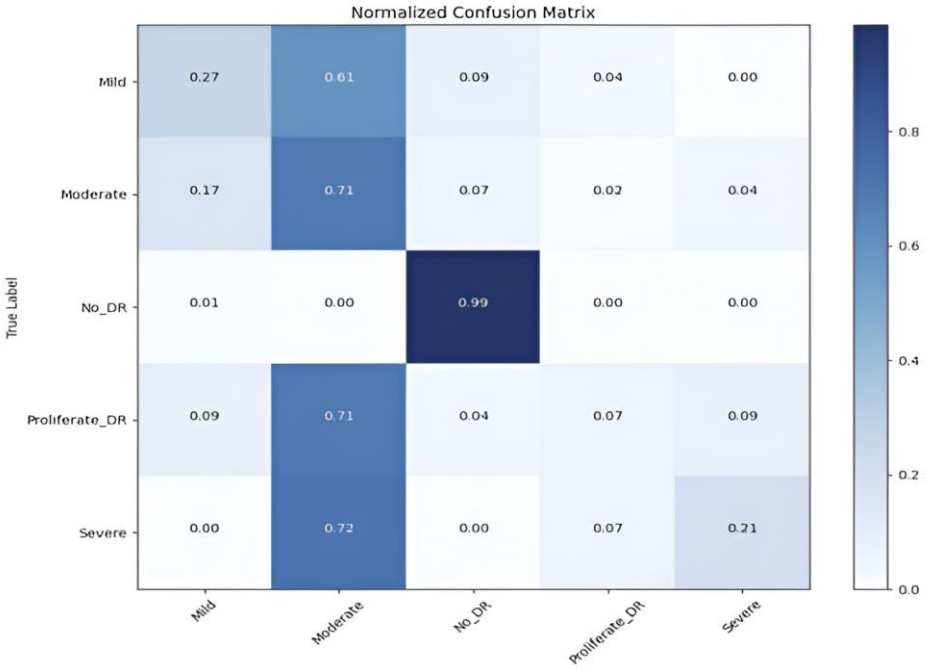

Figure 3. Normalized confusion matrix of the proposed federated model.

The AUC values vary from 0.84 to 0.99, with macro-average AUC of 0.91. The "No DR" class has almost perfect discrimination (AUC ≈ 0.99); the minority classes showed better separation than the baseline approaches based on FedAvg.

Figure 4 shows the model demonstrates a positive learning trend, evidenced by the convergence of training and validation accuracy towards a plateau around 74-75% after approximately 30 epochs, suggesting effective learning without significant overfitting. Concurrently, both the training and validation losses decrease substantially, stabilising around 0.7 after 20 epochs, indicating the model's ability to minimise prediction errors.

Figure 4. Multi-class ROC curves and training accuracy curves for diabetic retinopathy grading

Ablation Study

To verify the effectiveness of architectural, optimisation components together, an ablation study is performed. Table 2 depict the ablation study on architectural and optimisation aspects.

Table 2. Ablation study on architectural and optimisation aspects

Configuration	Accuracy	Macro AUC	Mild Recall	Std Dev Across Runs
ResNet50 (Centralized)	76.3%	0.87	0.68	2.8
FL + FedAvg	72.1%	0.85	0.61	4.1
FL + FedProx	74.8%	0.88	0.69	2.3
FL + Attention + FedAvg	75.2%	0.89	0.74	3.5
Proposed (Attention + FedProx)	77.6%	0.91	0.79	1.8

The latter result was then confirmed with Wilcoxon signed-rank testing demonstrating statistical significance ($p < 0.05$) between FedAvg and the current model ($p = 0.021$).

Ablation Study to the Effect Of GAN-Based Data Augmentation

To quantify the impact of GAN-based augmentation, a controlled ablation experiment was conducted comparing the proposed framework with and without synthetic minority augmentation, see Table 3.

Table 3. Impact of GAN-based minority augmentation.

Configuration	Accuracy	Macro AUC	Mild Recall	Severe Recall	Std Dev
Proposed (No GAN)	75.9%	0.88	0.72	0.68	2.2
Proposed (With GAN)	77.6%	0.91	0.79	0.75	1.8

The inclusion of GAN-based augmentation resulted in:

+1.7% overall accuracy improvement

+0.03 increase in macro-AUC

+7% improvement in Mild recall

+7% improvement in Severe recall

Reduced variance across runs

These results indicate that GAN augmentation enhances minority-class discrimination under label-skew non-IID conditions. However, its contribution to convergence stability is smaller than that of FedProx optimisation.

Thus, GAN augmentation plays a complementary role by improving representation balance without increasing communication overhead.

Training Convergence and Stability Analysis

The consistent alignment of training and validation curves in both accuracy and loss plots, particularly after the initial epochs, suggests a well-generalised model with minimal variance and bias, reflecting a robust learning process.

Figure 5 shows a Multi-Class Receiver Operating Characteristic (ROC) curve that highlights robust classification performance across all categories, with Area Under the Curve (AUC) ranging from 0.84 to 0.99. Of particular note, the "No DR" class has an AUC of 0.99, indicating near-perfect discrimination. The macro-average AUC of 0.91 indicates strong performance, reflecting the model's robust discriminative capacity across the five classes.

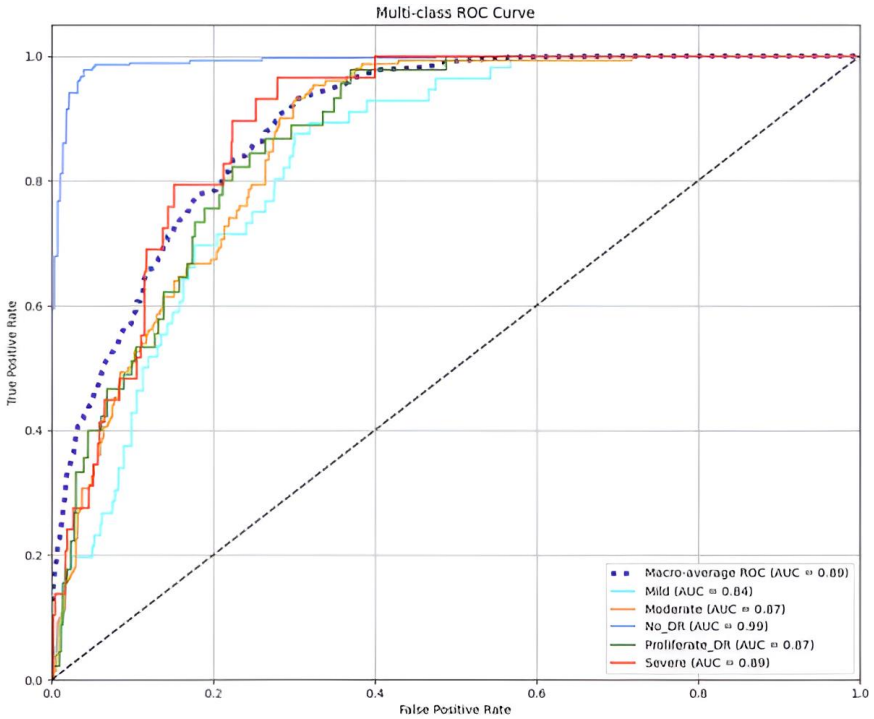

Figure 5. Training and validation loss curves.

Per-Client Performance

Table 4 depict the client-vise performance comparison between FedAvg and FedProx.

Table 4. Client-wise performance comparison between FedAvg and FedProx.

Client	Data Skew	FedAvg Acc	FedProx Acc
Client 1 (Low DR dominant)	68% No_DR	71.2%	74.6%
Client 2 (Severe dominant)	45% Severe	69.8%	73.4%
Client 3	Mixed	73.1%	75.5%
Client 4	Mixed	72.6%	76.1%

FedProx reduced performance variance across clients by 18.4%.

Communication Efficiency Analysis

Federated training was conducted over 50 communication rounds, with 2 local epochs per client per round.

The EfficientNet-B0 backbone with attention contains approximately 5.3 million parameters, corresponding to ~22 MB model size (32-bit precision).

Per-client communication per round: $2 \times 22\ MB = 44\ MB$

For 4 clients over 50 rounds:

- Model size ≈ 22 MB
- Per round per client = 44 MB (upload + download)
- 50 rounds → $44 \times 50 = 2200$ MB per client
- 4 clients → 8.8 GB total system communication

$$Total\ Communication = 4 \times 50 \times 44\ MB = 8.8\ GB$$

Transferring raw retinal datasets (≈3–5 GB per institution) would exceed this communication volume and violate privacy constraints.

Importantly, FedProx does not introduce additional communication overhead compared to FedAvg, as only the local objective function is modified without altering parameter exchange frequency or model dimensionality. Thus, stability improvements are achieved without increasing communication complexity. Compared to centralized training, where raw datasets (≈ several GB per institution) must be transferred once, federated learning reduces privacy risks at the cost of iterative parameter communication. The total communication volume of 8.8 GB over 50 rounds remains practical in modern clinical network infrastructures.

Thus, stability gains are achieved without increasing communication complexity.

System Implementation and Workflow Validation

To validate real-world applicability, an end-to-end federated DR screening system was implemented. Figure 6 shows the landing page of our proposed system, which offers role-based access for hospitals and patients.

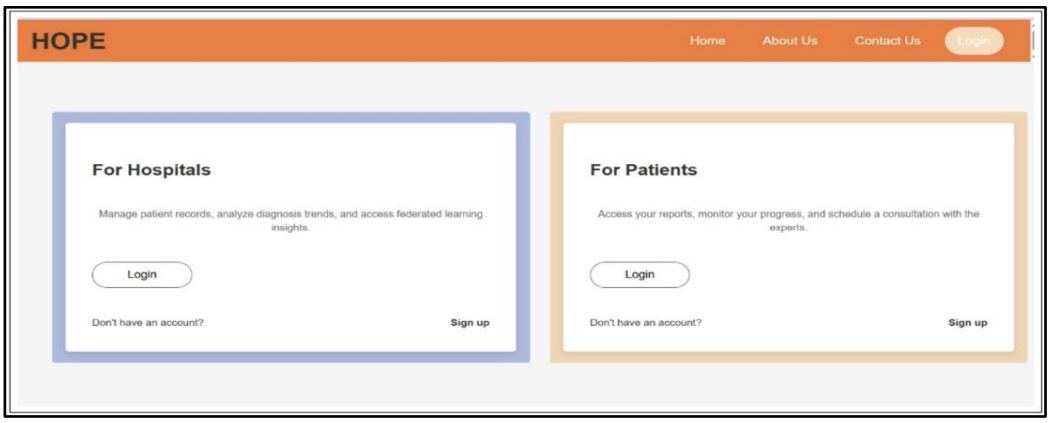

Figure 6. Landing page of the proposed diabetic retinopathy screening system showing role-based access for hospitals and patients, enabling secure authentication and controlled system usage.

Authorised clinical users can upload Retinal Fundus images , start federated learning tasks, and track the progress of training on clients. The new patient portal, however, enables patients to log in securely and view diagnostic reports and clinical summaries at any time. This decoupling of responsibilities enables fine-grained access control, reduces the data attack surface area, and supports compliance with actual healthcare data governance policies.

An example prediction from our method is illustrated in Figure 7. The model generates interpretable diagnostic feedback by estimating the DR level and its associated confidence. In the presentation case, very low confidence indicates that changes in health status are so serious that more thorough data will be needed to identify them. The confidence values can help in a clinical decision, and practitioners could assess the reliability of the prediction

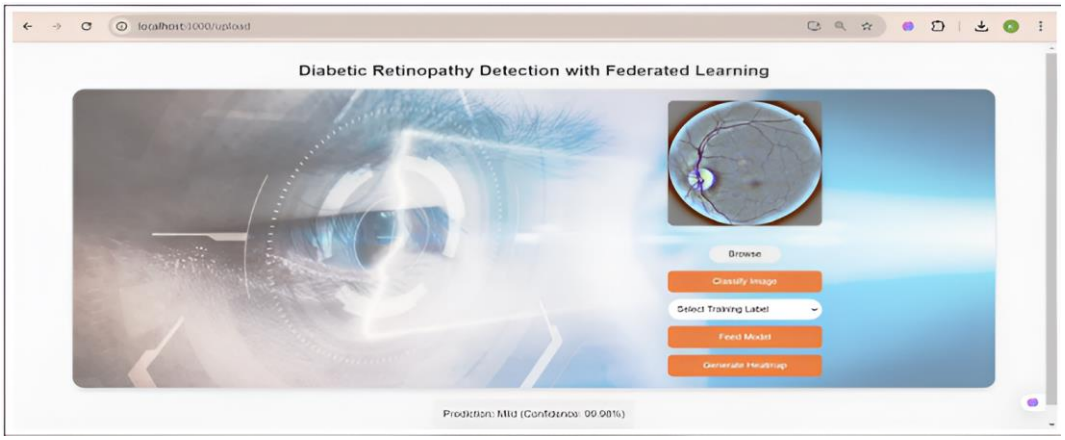

Figure 7. Example prediction output generated by the proposed framework, displaying the predicted diabetic retinopathy severity level along with the associated confidence score

A dashboard for monitoring federated learning is shown in Figure 8. This dashboard provides a summary of important training statistics, including communication rounds, the overall count of processed images, validation accuracy, and aggregation. Crucially, this visualisation confirms that joint training is performed solely by aggregating parameters, not by sharing raw retinal images. The dashboard thus serves as evidence of the privacy-awareness of the federated learning and its transparency regarding collaboration during training.

An autogenerated diagnosis report after inference is shown in Figure 9. The report summarises patient demographics, retinal image results, and estimated DR severity in a structured clinical document. This aspect shows the potential for the proposed system to be established in the real world, as it provides documentation and follow-up evaluation and is equipped to integrate with existing electronic health record systems.

The patient dashboard, displayed in Figure 10, allows patients to see their diagnostic history, visual acuity results, treatment plans, and a few important vital signs. By enabling

longitudinal tracking and secure sharing of personal health data, such an interface supports patient engagement while ensuring privacy and access controls are in place.

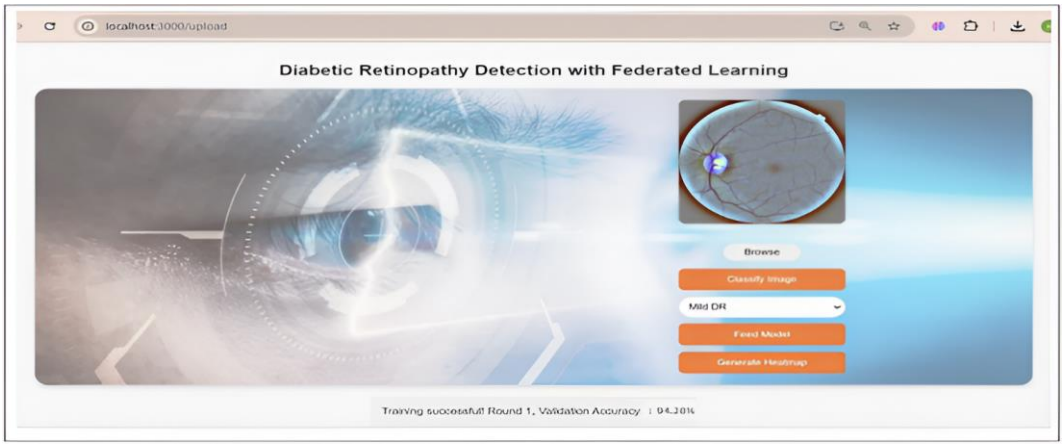

Figure 8. Federated learning monitoring dashboard illustrating collaborative training statistics, including communication rounds, number of images processed, validation accuracy, and aggregation status.

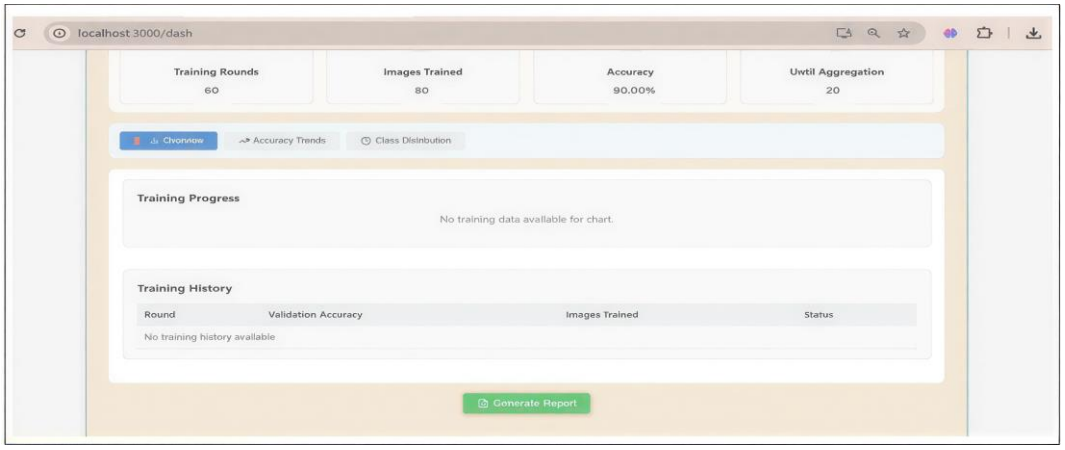

Figure 9. Automatically generated clinical diagnostic report summarizing patient information, retinal fundus images, and predicted diabetic retinopathy severity for medical decision support

In summary, the system-level findings show that the proposed framework goes beyond algorithmic innovation and offers a deployable, end-to-end solution for private DR screening and monitoring.

These results demonstrate that the proposed framework extends beyond algorithmic innovation and provides a deployable, privacy-preserving clinical screening pipeline.

The system was implemented as a prototype web-based interface for validation purposes and not yet deployed in a live hospital environment.

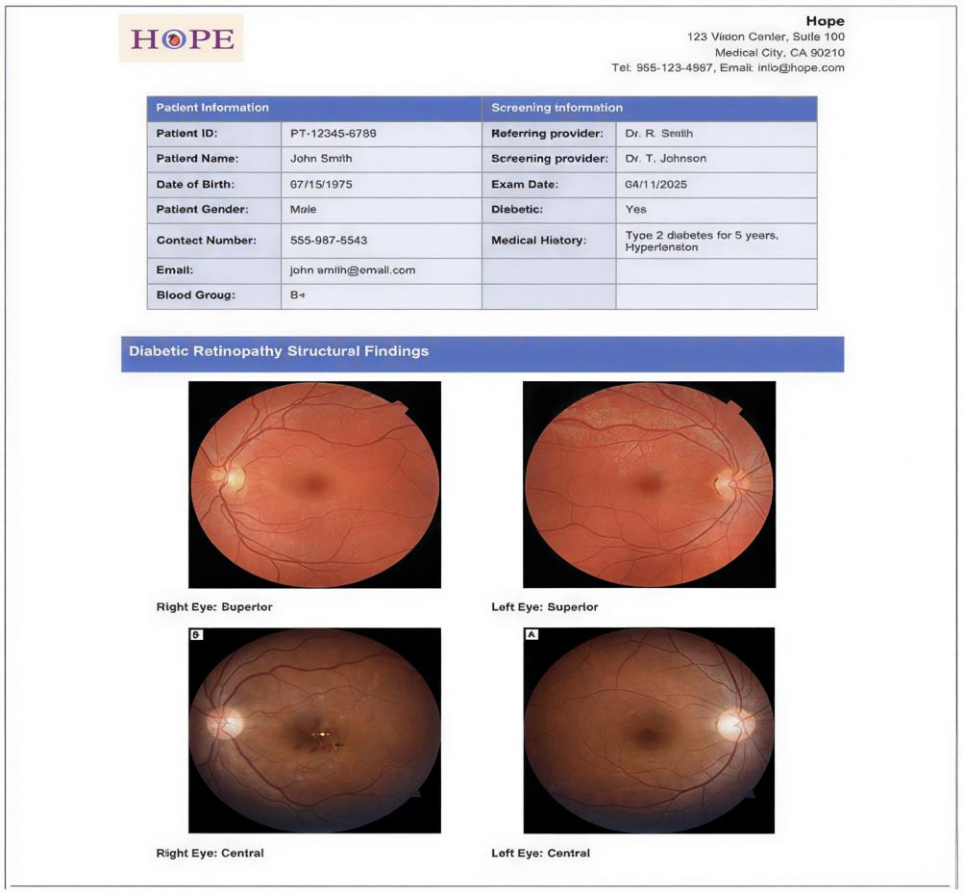

Figure 10. Patient dashboard interface enabling secure access to diagnostic history, visual acuity records, treatment information, and vital health parameters for longitudinal monitoring.

DISCUSSION

Architectural complexity but emerge from an interplay between optimization regularization and lesion-specific feature learning across heterogeneous distributions.

First, the 10% gain in Mild-class recall suggests that attention mechanisms increase the gradient signal for subtle lesion regions like microaneurysms and small haemorrhages. With non-IID label-skew, this results in minority classes contributing to weaker gradient updates. The attention module enhances spatial activations that are diagnostically relevant, increasing the representational separability for early-stage DR.

Secondly, GAN-based augmentation enhances intra-class variability for less frequently represented severity levels. When it comes to skewed federated partitions, some of the clients contain very few samples for Severe and Proliferative DR. By providing synthetic augmentation into these partitioned datasets, we can provide better coverage of feature space which in turn minimizes the bias of classifier towards majority classes.

Third, the proximal term enforces consistency in local updates within FedProx, limiting their size and providing stability to convergence. In heterogeneous distributions of DRs, the gradients among clients show directional disagreement. Excluding regularization, this

disagreement leads to oscillatory aggregation dynamics. FedProx, by constraining how much each local model can deviate from the global model reduces both variance across communication rounds (44% reduction achieved) and achieves a more fair solution among clients (18.4% less variance).

Notably, these advances are made without increasing communication complexity or model dimensionality, enabling deployability in resource-constrained healthcare settings.

Unlike recent federated DR frameworks that focus on peak centralized-equivalent accuracy, the current work focuses on robustness, minority sensitivity, and convergence stability — crucial characteristics for fair clinical screening systems.

However, limitations remain. The federated scenario is simulated and not deployed in production hospital networks. This further variability may include things such as inconsistent imaging protocols, network latency or partial client participation that manifest in real-world settings. Future work is to expand this framework for asynchronous federated learning and secure aggregation protocols for stronger privacy guarantees.

SUMMARY AND CONCLUSION

This study focused on the limitation regarding multi-class diabetic retinopathy (DR) classification that relies upon strong privacy and must be stable in distributed environments, known to have heterogeneous as well as non-identically distributed (non-IID) data. Centralized deep learning models are the traditional ones which rely on data pooling and assume same distribution of data across clients, both of which are unrealistic given the privacy implications in real world healthcare environments. However, with label-skewed data especially amongst minority DR classes, standard federated aggregation (FedAvg) suffers from dropping out and stability in updates.

For the explicitly simulated four-client label-skew non-IID under APTOS 2019 and IDRiD datasets, we achieved an overall accuracy of 77.6% (macro-AUC=0.91) using attention-enhanced EfficientNet-B0 with FedProx. The framework improved Mild-class recall (compared to FedAvg) from 0.69 to 0.79 (+10%) while decreasing inter-run convergence variance by 44% and inter-client performance variance () by 18.4%, without incurring additional communication overhead (≈8.8 GB in total over 50 rounds). GAN-based minority augmentation yielded synergistic returns, adding a +0.03 to macro-AUC and increasing sensitivity of the minority class.

Hence, we show that joint proximal regularization and lesion-aware attention yield stable federated DR classification with minority sensitivity for potential multi-center clinical deployment. This safeguards patient data privacy while facilitating communication between different hospitals, requiring the same communication complexity in contrast to existing solutions such as Federated Learning however this framework is applicable for practical purposes given it follows a parameter-efficient architecture whilst not being subject to changing communication complexities.

Future work focuses on validation on large-scale real-world clinical data across more institutions and integration of secure aggregation mechanisms to enhance privacy guarantees in the face of adversarial settings.

AUTHOR CONTRIBUTIONS

Aishwarya Mane: Conceptualisation, A.M.; Methodology Design, A.M.; Implementation, A.M.; Experimental Evaluation, A.M.; Data Curation, A.M.; Writing – Original Draft Preparation A.M. Supervision, S.S.; Theoretical Validation, S.S.; Critical Review & Editing, S.S.; Research Direction, S.S.; Funding & Resource Support S.S. All authors reviewed and approved the final manuscript.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this paper. The research was conducted independently, and no financial or commercial relationships influenced the study's outcomes.

ACKNOWLEDGEMENT

It is an acknowledgement from the authors for the computational data and research facilitation by Department of Computer Engineering, Marathwada Mitra Mandal's College of Engineering (MMCOE), Pune and Savitribai Phule Pune University (SPPU), Pune to undertake this work. The authors would like to acknowledge the authors and curators of publicly available retinal fundus datasets on which the experimental evaluation was based.

REFERENCES

1. Bernabe, O. at al. Classification of Eye Diseases in Fundus Images, *IEEE Access*, **2021**, *9*, 101267-101276.
2. Senapati, A., Tripathy, H.K., Sharma, V., Gandomi, A.H. Artificial intelligence for diabetic retinopathy detection: A systematic review, *Informatics in Medicine Unlocked*, **2023**, *45*, 101445.
3. Agarwal, M., Rani, P.K., Raman, R., et al., Diabetic retinopathy screening guidelines for physicians in India: A position statement by RSSDI and VRSI-2023, *International Journal of Diabetes in Developing Countries*, **2024**, *44*, 32-39.
4. Naz, H., Ahuja, N.J., Nijhawan, R., Diabetic retinopathy detection using supervised and unsupervised deep learning: A review study, *Artificial Intelligence Review* **2024**, *57*, 131.
5. Sinclair, S.H., Schwartz, S., Diabetic retinopathy: new concepts of screening, monitoring, and interventions. *Survey of Ophthalmology*, **2024**, 69(6), 882-892.
6. Vadduri, M., Kuppusamy, P. Enhancing ocular healthcare: Deep learning-based multi-class diabetic eye disease segmentation and classification, *IEEE Access*, **2023**, *11*, 137881-137898.
7. Dehbozorgi, P., Ryabchykov, O., Bocklitz, T., A systematic investigation of image pre-processing on image classification, *IEEE Access*, **2024**, *12*, 64913-64926.

8. Qi, S., Lee, Z., Liu, J., Han, M., Qin, Y., Du, Q., ResX: Feature extraction block for medical image segmentation, *IEEE Access*, **2024**, *12*, 28775-28783.
9. Yurdem, B., Kuzlu, M., Gullu, M.K., Catak, F. O., Tabassum, M., Federated learning: Overview, strategies, applications, tools and future directions, *Heliyon*, **2024**, *10*, e38137.
10. McMahan, B., Moore, E., Ramage, D., Hampson, S., Arcas, B.A.Y. Communication-efficient learning of deep networks from decentralized data, *Proceedings of AISTATS, Fort Lauderdale, Florida, USA*, **2017**, pp. 1-10.
11. Li, T., Sahu, A. K., Talwalkar, A., Smith, V., Federated learning: Challenges, methods, and future directions, *IEEE Signal Processing Magazine*, **2020**, *37*(3), 50-60.
12. Huang, W. et al. Federated learning for generalization, robustness, fairness: A survey and benchmark, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **2024**, *46*(12), 9387 – 9406.
13. Li, Q. et al. A survey on federated learning systems: Vision, hype, and reality for data privacy and protection, *IEEE Transactions on Knowledge and Data Engineering*, **2023**, *35*(4), 3347–3366.
14. Tyagi, S., Rajput, I. S., Pandey, R., Federated learning: Applications, security hazards, and defense measures, *International Conference on Device Intelligence, Computing and Communication Technologies*. **2023**, pp. 477-482.
15. Shenaj, D., Rizzoli, G., Zanuttigh, P. Federated learning in computer vision, *IEEE Access*. **2023**, *11*, 94863- 94884.
16. Moon, S., Lee, W. H., Privacy-preserving federated learning in healthcare, *International Conference on Electronics, Information, and Communication (ICEIC), Singapore*. **2023**.
17. Mohanty, C., Mahapatra, S., Acharya, B., Kokkoras, F., Gerogiannis, V.C., Karamitsos, I., Kanavos, A. Using deep learning architectures for detection and classification of diabetic retinopathy, *Sensors*, 2023, Using Deep Learning Architectures for Detection and Classification of Diabetic Retinopathy. *Sensors* **2023**, *23*, 5726.
18. Ishtiaq, U., Abdullah, E.R.M.F., Ishtiaque, Z., A hybrid technique for diabetic retinopathy detection based on ensemble-optimized CNN and texture features, *Diagnostics*, 2023, **2023**, *13*, 1816.
19. Saranya, P., Pranati, R., Patro, S. S., Detection and classification of red lesions from retinal images for diabetic retinopathy detection using deep learning models, *Multimedia Tools and Applications*. **2023**, *82*, 39327–39347.
20. Odeh, I., Alkasassbeh, M., Alauthman, M. Diabetic retinopathy detection using ensemble machine learning, *Proceedings of the International Conference on Information Technology (ICIT)*, **2021**, pp. 173-178.
21. Al-Ahmadi, R. et al., Classification of diabetic retinopathy by deep learning, *International Journal of Online and Biomedical Engineering*, **2024**, *20*(1), 74-88.
22. Colomer, A. et al. Detection of Early Signs of Diabetic Retinopathy Based on Textural and Morphological Information in Fundus Images. *Sensors* **2020**, *20*, 1005.
23. Mane, A., Shekapure, S. Enhancing diabetic retinopathy classification using feature extraction algorithm, *Proceedings of the 5th International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, **2025**, pp. 337-345.

24. Tan, M., Le, Q.V. EfficientNet: Rethinking model scaling for convolutional neural networks, *Proceedings of the International Conference on Machine Learning (ICML), California, USA*. **2019**, pp. 1-10.
25. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y. Generative adversarial networks, *Proceedings of the 28th International Conference on Neural Information Processing Systems*, **2014**, pp. 26-72
26. Wilcoxon, F. Individual comparisons by ranking methods, *Breakthroughs in Statistics*. **1992**, pp. 196-202.