

FedDDR: A Federated Improved DenseNet for Classification of Diabetic Retinopathy

Akansha Singh^a, Krishna Kant Singh^b

^a SCSET, Bennett University, Greater Noida, India

^b Delhi Technical Campus, Greater Noida, India

Abstract

Damage to the retina and other eye blood vessels is the result of diabetes, a condition known as diabetic retinopathy (DR). Affected individuals may have retinal clots, lesions, or hemorrhaging. Exudates and lesions in the retina may cause visual loss in people with diabetic retinopathy. Diabetic retinopathy identification is essential for effective patient care. This research proposes a federated version of an enhanced DenseNet deep learning model for use in the detection and classification of Diabetic Retinopathy in retinal fundus pictures. With the dense blocks performing concatenation, the upgraded DenseNet model improves the feature utilization efficiency. The model is trained using federated learning algorithm. Federated learning enables distributed training of the model using remotely hosted datasets without the need to gather data and, subsequently, damage it. This overcomes the limitations posed by the data silos and makes full advantage of the existing medical data. The proposed model improves the performance and ensures patient privacy by not gathering the data at a central dataset. The federated average learning algorithm is used to train the model. The model uses Maximum Probability Based Cross Entropy (MPCE) loss function. The proposed method's outcomes are evaluated and contrasted with those of similar approaches. The results of this comparison demonstrate that the suggested technique is superior to the others in terms of accuracy, precision, and recall when applied to the categorization of retinal pictures.

Keywords 1

diabetic retinopathy, deep learning, Federated Learning, DenseNet

1. Introduction

Use Deep learning has emerged as a promising strategy for automated clinical diagnosis. The most prevalent complications of diabetes are well-known to the general population. Many incidences of avoidable blindness are caused by diabetic retinopathy, an eye condition that diabetics are prone to, but which is not as well-recognized as other diabetes consequences. Diabetic retinopathy affects around 60% of people with type 2 diabetes and over 100% of those with type 1. The illness progresses through four distinct phases, with the first two being the most manageable thanks to early detection and subsequent preventive care. High blood sugar levels may damage blood vessels in the retina, leading to diabetic retinopathy, an eye condition, as described by the American Academy of Ophthalmology. There is a risk of a blockage and subsequent lack of blood flow if the afflicted blood vessels expand and leak or seal up completely. Diabetic retinopathy may be very damaging to a person's eyes if left untreated, thus finding it early is crucial. Possessing a reliable method for early detection of the illness is crucial. The hazards associated with each stage of diabetic retinopathy are discussed here, as well as the symptoms experienced at each stage and the medicinal interventions available to prevent further progression of the disease. To diagnose and treat diabetic retinopathy in its earliest stages, it is essential to take preventative measures, such as arranging yearly diabetic retinal exams.

IDDM'2023: 6th International Conference on Informatics & Data-Driven Medicine, November 17 - 19, 2023, Bratislava, Slovakia

EMAIL: akanshasing@gmail.com (A. 1); krishnaiitr2011@gmail.com (A. 2)

ORCID: 0000-0002-5520-8066 (A. 1); 0000-0002-6510-6768 (A. 2);



© 2023 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

These vital retinal examinations may discover hazardous problems before they cause significant vision loss, giving the patient and their doctors the time to devise a treatment strategy. Patients and doctors may use this action plan as a road map to better comprehend and address the far-reaching effects of diabetes on a person's health. Photos of a healthy eye and a DR-affected eye are shown side-by-side in Figure 1.

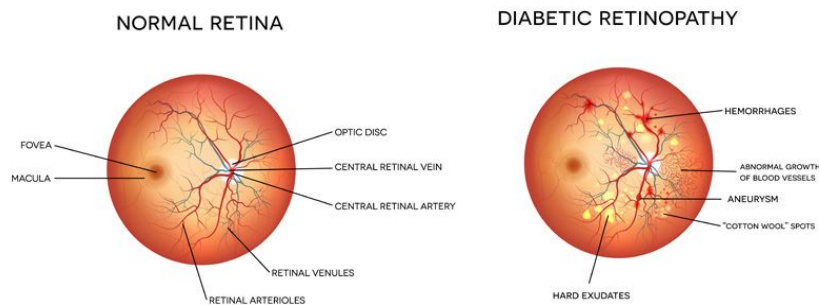


Figure 1: Eye structure and presence of DR [Image credit <https://www.eyepops.com/>]

Thus, to enable early detection of diabetic retinopathy computer aided methods are being widely developed. In this paper, a federated deep learning model is proposed to detect DR using retinal images. The use of Artificial Intelligence to help radiologists with computer-assisted patient diagnosis has been generally successful, it is still difficult to create robust models with tiny datasets at specific locations. The Early Treatment Diabetic Retinopathy Study (ETDRS) is only one of many widely used diabetic retinopathy grading systems [1]. ETDRS uses a multi-tiered system to categorize the finer, more nuanced aspects of DR. All seven fovea of the retinal fundus are evaluated in this manner (FOV). The ETDRS is the standard, although the International Clinical Diabetic Retinopathy (ICDR) [2] scale is employed instead because of its acceptance in both clinical and CAD contexts [3]. This is because the ETDRS is hard to execute and has technological limitations. There are less field-of-view (FOV) requirements for the ICDR scale, which specifies 5 severity levels and 4 levels for Diabetic Macular Edema (DME). It has been shown that convolutional neural networks are effective in detecting and classifying DR [4]. Transfer learning was used with CNN to significantly increase the performance of these networks [5]. The retinal pictures in the medical image collection may be used to fine-tune the pretrained models using transfer learning. These models were shown to be more precise than the standard CNN [6]. Ensemble approaches, which take the best features of several classifiers and combine them, have been suggested by several academics. There is a greater information gain in the ensemble models since they include the results of several individual models. Many different ensembling methods exist for integrating complementary model data. For the DR issue, several ensemble classifiers have been described and published in the literature [7-9].

The limited medical data availability is a major challenge for deep learning models. Collaboration across several hospital is important to attain excellent algorithm performance when the number of medical data samples is constrained. Due to technical, regulatory, or ethical issues, sharing patient data frequently has restrictions. A neural network model with small sample size of biomedical images could not be very generalizable. A multi hospital study could be used to overcome the difficulties because it can greatly expand the sample size and sample variety.

Conventionally, the algorithm is trained on all patient data at a central location. But this strategy has some drawbacks. First, sharing patient data that requires a lot of storage space (such as high-resolution photographs) may be difficult. Second, legal, or ethical constraints prevent sharing part or all patient data. Third, patient data is precious, thus institutions may not share it. Instead of sharing data to central location, Federated learning based deep learning model may be more useful and accurate.

With federated learning the repeated analysis of many databases and the exchange of mathematical parameters (metadata) rather than real data that may disclose possible patient identifiers is utilized [10]. Early applications of federated learning methods saw more uptake in image classification and the improvement of wireless communication systems. Predictions on healthcare outcomes such as mortality, ICU stay-time, hospitalization for cardiac events, dyspnea,

adverse medication responses, and more have recently been included into federated learning models in the healthcare domain [11]. However, most applications of federated learning in healthcare outcome prediction used relatively small datasets and partitioned the data theoretically (randomly) to simulate the properties of actual data. In this research, we apply our framework to the Health Facts data by using the information provided by the healthcare systems for each individual patient.

Most healthcare federated learning applications employed classification methods such logistic regression, artificial neural network, multi-layer perceptron, support vector machines, and random forest to construct federated predictive models. Existing methods for predicting complications from diabetes, such as retinopathy (eye disease), neuropathy (peripheral nerve disorder), and nephropathy (kidney disease), rely on centralized machine learning algorithms trained on small-size datasets from the US population, which contain fewer than ideal numbers of cases of complications and less than ideal patient information. In this research, we used a federated learning architecture to develop three different machine learning models for binary classification of the occurrence of three different diabetes-related complications: those affecting the eyes, the kidneys, and the peripheral nerves.

The existing deep learning and machine learning models have several limitations. These include limited medical data availability, patient privacy issues and training overhead at a centralized location. Therefore, in this paper a modified federated learning DenseNet model is proposed for the classification of diabetic retinopathy. In this architecture, several sites may work together to train a single global model. With federated learning, a global model is built by combining training results from many locations without the need to share datasets. The confidentiality of the patients is protected in this way. The global model's detection skills are further enhanced by the additional supervision received from the findings of collaborating locations. When training AI models with little data, this solves the problem of inadequate supervision. Thus, in this paper the above-mentioned limitations are removed with the proposed Federated Learning Dense net model. The initial model on the central server is initialized and the parameters are shared with the connected devices. The results are simulated using tensorflow federated learning module. More than 5,000 retinal pictures from the third biggest dataset APTOS19 are segmented for use in virtual testing. The DenseNet model used overcomes the vanishing gradient problem and strengthens the feature propagation as features are concatenated at each stage.

This paper is organized into five sections in the first section the introduction to the problem and literature review is discussed. The second section discusses the proposed methodology followed by results and discussion section. The last section gives the overall conclusion of the work presented in this paper.

2. Proposed Method

The traditional machine learning models that are trained centrally on one device pose some serious challenges when used for healthcare applications. The limited availability of data due to multiple constraints of data privacy and sharing is a major issue. Therefore, in this paper a DenseNet model with federated learning approach is presented. Data privacy, data security, data access rights, and access to heterogeneous data may all be addressed by using federated learning, which allows several hospitals to construct a shared, robust machine learning model without sharing data. Therefore, federated learning models may collect data from several sources (e.g., hospitals, electronic health record databases) to give more diverse data. In this section the steps involved in the proposed methodology are discussed in detail. In figure 2 the proposed methodology is shown. The central model is trained using N connected devices. Each device uses its own dataset for training and transmits the updates in the model weights.

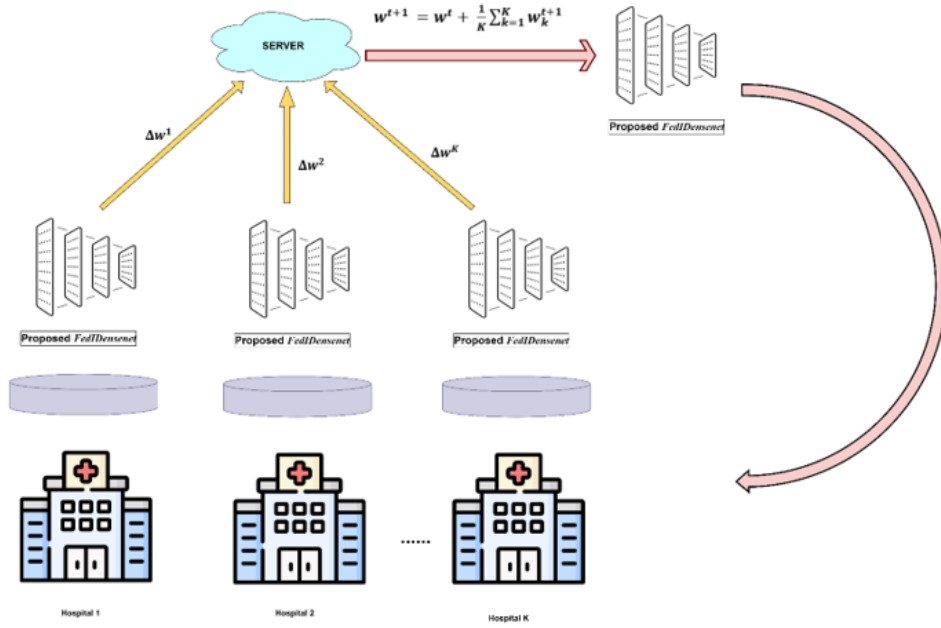


Figure 2: Proposed Methodology

The general training mechanism is shown in figure 3.

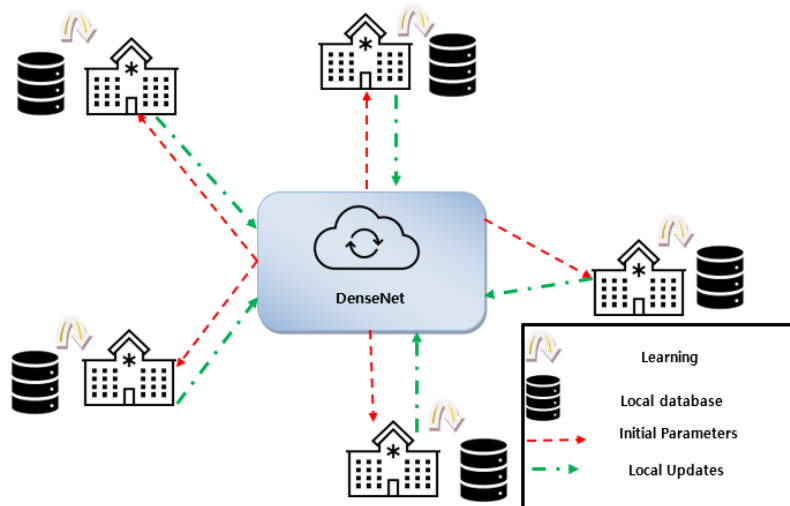


Figure 3: Training at central and connected devices

1. **Initial Model Configuration :** The DenseNet model [12] is initialized at the central server device. The training of the central model is done using the APTOS2019 dataset. The initial parameters of the model are then transmitted to each of the connected devices.
2. **Training at connected devices:** A copy of the model is available at each of the connected and it uses the parameters broadcasted by the server. The following steps are followed at each connected device.
3. **Input Retinal Images:** The retinal pictures are fundus images captured under a variety of lighting and camera angles. A doctor assigns a score from zero to four for five categories to each picture, reflecting the severity of diabetic retinopathy. The model is tested and trained using these pictures.
4. **Pre-processing of images:** The photos are shot in a variety of environments with varying levels of illumination. Before they may be utilized for model training, these photos need preprocessing. Due to the lack of contrast in retinal pictures, CLAHE is used to equalize the histograms [13]. The CLAHE histogram equalization is computed as follows:

$$X^{new} = \frac{X_{max} - X_{min}}{X_k} \quad (1)$$

The proposed deep learning network receives its input data from the pre-processed images acquired in this stage.

5. **Image Resizing:** As the images at different connected devices are of different size and thus, they need to be resized before feeding to the deep network. Thus, all images are resized to 224×224 pixels.
6. **Image Standardization:** Image standardization is a data transformation technique. Standardization rescales the image features so that the mean is 0 and standard deviation is 1. This improves the optimization and consequently the accuracy of the model.

$$X^{st} = \frac{X - \mu_X}{\sigma_X} \quad (2)$$

where μ_X is the mean and σ_X is the standard deviation.

7. **DenseNet Model:** A DenseNet model is used for the classification of the retinal images for identification of diabetic retinopathy. Using Dense Blocks, in which we directly link all layers (with matching feature-map sizes) with each other, a DenseNet is a sort of convolutional neural network that makes use of dense connections between layers. To maintain the feed-forward structure, each layer pulls in data from the layers below it and sends its own feature maps to the layers above it. DenseNets have outperformed traditional CNNs and ResNets on a wide variety of benchmark datasets, and their smaller model size is a consequence of both factors. The architecture of the DenseNet 121 used is as follows:

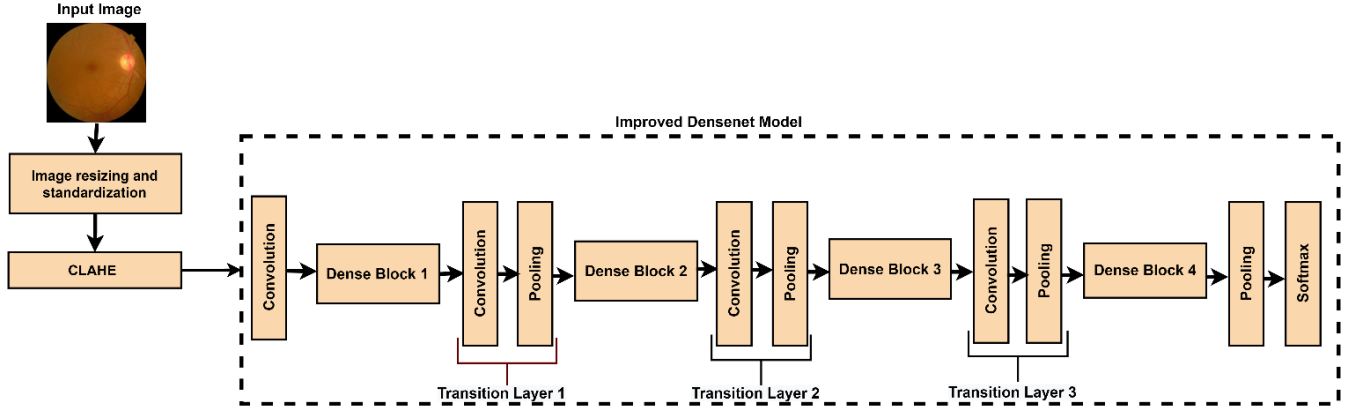


Figure 4: Model Architecture for the proposed method

In each layer the feature maps of the preceding layers are concatenated as input. Due to concatenation the features are not repeated, and redundant features are removed. Each l th layer receives the feature maps of the previous layers.

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]) \quad (3)$$

where $[]$ denotes concatenation operation and H_l is a composite function. It comprises of batch normalization (BN), a rectified linear unit (ReLU) and a convolution (Conv).

DenseBlocks are the building blocks of DenseNet; the size of the feature maps stay the same inside a block, but the number of filters varies. By removing one layer of transition between each block, we may cut the total number of channels in half.

The amount of information to be added in each layer is controlled by the growth rate (k) of DenseNet. Thus, in l th layer the amount of information added can be computed as:

$$k_l = k_0 + k * (l - 1) \quad (4)$$

where k_0 is the number of channels in the input layer.

8. **Maximum probability based cross entropy loss:** For the sake of fine-tuning the model-learning process an MPCE loss function is implemented [20]. Because of this, the convergence is accelerated, and the back propagation error is minimised. MPCE may be expressed mathematically as in eq (5).

$$f^t(W) = -\sum_{i=1}^m y'_i \log(y_i) = -\sum_{i=1}^m (y_{max} - y_u) \bar{y}_i \log(y_i) \quad (5)$$

where, y_{max} where the real class, u_{th} among m classes, is the largest. If \bar{y} is a vector of real classes, then the u th coordinate is 1, and \bar{y} is the vector of u th coordinates. The i -th coordinate of the vector y' is denoted by y'_i .

9. **Adam Optimization:** In order to maximise efficiency, the adam optimizer combines the benefits of both the Momentum and Root Mean Square propagation methods [14]. When the gradient hits its global minimum, ADAM slows the pace of descent such that there is little oscillation.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \left[\frac{\delta L}{\delta w_t} \right] v_t = \beta_2 v_{t-1} + (1 - \beta_2) \left[\frac{\delta L}{\delta w_t} \right]^2 \quad (6)$$

10. **Federated averaging Learning:** Federated averaging algorithm uses an averaging method to combine the updates at the central server [15]. A network of N devices available at N different hospitals, indexed $i \in \{1, 2, \dots, N\}$. Each device or hospital has its own dataset consisting of retinal images denoted by D_k . Each D_k comprises of an input vector x_d and an outcome variable y_d . The model will be trained using this network of devices. Thus,

$$g_w = x_d \rightarrow \widehat{y}_d \quad (7)$$

x_d is the input feature vector and \widehat{y}_d is the predicted output using vector w and loss function.

The local loss at each device can be computed as,

$$F_k(w) = (1/|D_k|) \sum_{d \in D_k} l_d \quad (8)$$

The assumption in this problem is ,

$$|D_i| = |D_j| \quad \forall i, j \quad (9)$$

Thus, the optimization is the average over the $F_k(w)$. The objective is to find w that minimizes $f(w)$ over the data $D = \cup_k D_k$

$$\min_w f(w), \text{ where } f(w) := \frac{1}{N} \sum_{k=1}^N F_k(w) \quad (10)$$

In case $|D_i| \neq |D_j|$ then $\frac{1}{N}$ can be replaced with $p_k = |D_k|/|D|$

The complete algorithm for the training process is as follows:

Algorithm: FedDDR learning
Input: K [Number of Hospitals/Devices], T [epochs], w^0 [Initial weight vector], α [learning rate of client], γ [Learning rate of server]
Start
Server broadcasts w^0 to K devices.
For $t=0, \dots, T-1$
For each device $k = 1, \dots, K$ computes w_k^{t+1}
Each devices sends the w_k^{t+1} back to the server
Server averages and updates the w as
$w^{t+1} = w^t + \frac{1}{K} \sum_{k=1}^K w_k^{t+1}$
Output the final model parameters $w^{trained}$

11. **Termination Condition:** The training is terminated when the number of iterations is complete, or the model has converged to the optimal solution.
12. **Grad Cam Visualization:** The Gradient based Class Activation Map (Grad-CAM) is an example of a class-discriminative localization map that draws attention to important parts of an image by calculating the gradient of the class score y^c for class c relative to the activations A^k of a convolutional layer's feature map $\partial y^c / \partial A^k$. a^c_k is the result of a global-average-pooled backflow of these gradients, which is used to derive the significance of neuronal weights [16].

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \overbrace{} \quad (11)$$

Grad-CAM is basically a weighted combination of forward activation maps followed by ReLU operation as follows:

$$L_{Grad-CAM}^c = ReLU\left(\sum_k \alpha_k^c A^k\right) \quad (12)$$

With the help of the Grad-CAM visualisation heatmap, we can see how the three categories our model predicts for test photos are distributed among a set of representative examples. Grad-heatmap CAM's depiction draws attention to the key pixel clusters used by the model's last convolution layer to make class distinctions. Here, we see how the GRAD-CAM visualisation system distinguishes between standard and DR photos by highlighting them in a variety of ways. Class activation maps for both normal and DR photos show that the centre of the image is emphasised more strongly in the former instance, while the top part of the image is illuminated more densely in the latter. Important visual features utilised by the model to make the concept prediction are highlighted in the class activation map. Figure 5 displays several example gradcam representations of retinal images.

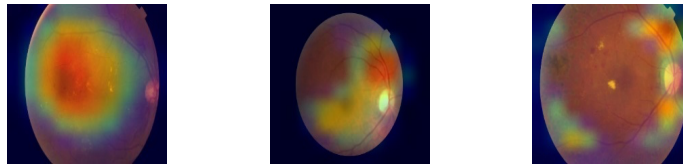


Figure 5: GradCam visualizations of Retinal images

3. Experiments

The proposed method is tested and its performance is measured by implementing it in Python. The dataset is large and thus GPU acceleration is used for the simulation purpose. Keras module in Python is used for developing the deep network and Tensorflow Federated Learning for training. The initial base model is trained using the following datasets. Thereafter, each client uses its own dataset. For simulation purpose the dataset was divided into different clients. The dataset used is from APTOS 2019, a dataset on diabetic retinopathy (<https://www.kaggle.com/c/aptos2019-blindness-detection>). Aravind Eye Hospital collects the data in rural India so that it may be used to create AI for DR detection. All the images fall into one of five categories: No DR, Mild, Moderate, severe and Proliferative DR. The severity, location, and frequency are taken into account when assigning a grade from 0 to 4.

The dataset comprises of a total of 3662 images. For experiments the images are split into training and testing in the ratio of 80:20. Thus, 2930 images are used for training and 732 images are used for testing. The sample images from the database are shown in figure 6.

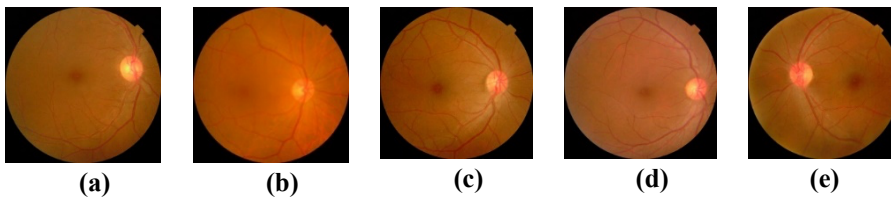


Figure 6: Retinal images for (a) Normal (b) Mild DR (c) Moderate DR (d) Severe DR (e) Proliferative DR

We use these metrics to measure how well the suggested technique performs. Specifically, these indicators are employed:

The number of incorrect predictions is the standard measure of accuracy, or precision. You can figure this out by

$$Precision = \frac{TP}{TP+FP} \quad (13)$$

For each given model, recall indicates how many true positives it generates. In eq.(14), we see the formula for calculating the recall:

$$Recall = \frac{TP}{TP+FN} \quad (14)$$

Total accuracy is calculated using the same method.

$$Overall Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (15)$$

where TP, FP and FN represents the true positive, false positive and false negative, respectively.

F1-score can be computed using eq.(16)

$$F_1 = 2 \times \frac{precision \times recall}{precision + recall} \quad (16)$$

4. Results and Discussion

In this section the results obtained from the proposed method and a comparative analysis is presented. The results are obtained by performing simulations and the data was split amongst different clients to observe the results. Federated data set, i.e., a collection of data from multiple users is required for demonstrating the proposed method. Thus, to facilitate experimentation, the data set was split amongst five users. Due to individual differences in data consumption behaviors [21, 22], federated data is often not identically distributed among users. Due to data scarcity on the device, some customers may have fewer training instances than others, while other clients may have more than enough. Because this is a simulated environment, we have access to all the data required to do such a comprehensive examination of a client's data. In a fully operational federated setting, it is impossible to see the data of a single client.

Table 1:
Data Splitting for five devices

Dataset Splitting for 5 devices						
Grade	D1	D2	D3	D4	D5	Total Images
0 -No DR	144	360	285	304	351	1444
1 – Mild	40	56	62	66	72	296
2-Moderate	125	184	116	196	178	799
3- Severe	20	45	44	34	34	155
4- Proliferative	110	54	57	42	32	236

An extremely large number of user devices may be involved in a typical federated training scenario, yet only a subset of these devices may be accessible for training at any one moment. For instance, when the client devices are mobile phones, they can only take part in the training when they are fully charged, not using the network, and not being charged. Since this is a simulation, all the information we need is already on-hand. So, when we conducted simulations, we would usually choose a new group of customers to train with each time.

The parameters used for simulating the federated learning environment is as follows:

Table 2:
Parameters Used for Simulation

Number of Clients	5
Client Optimizer (Local model Updates)	Adam Optimization
Server Optimizer (averaged update to the global model at the server)	Adam Optimization
Learning Rate (Client)	0.001
Learning Rate (Server)	0.001
epochs (Server)	60

All DR pictures in the dataset fall into one of five categories, labelled with the digits 0-4.

0 – No DR (NDR)
1: Mild

2: Moderate
3: Severe

4: Proliferative DR (PDR)

732 test photos from a range of grading levels were used in the analysis. Table 4 displays the distribution of photos by grade level.

Table 3:
Testing image distribution

The five-class confusion matrix is shown in table 6.

Grade	Testing images
0 -No DR(NDR)	353
1 – Mild	87
2-Moderate	205
3- Severe	40
4- Proliferative (PDR)	48
Total	733

Table 4
Confusion Matrix for five classes

		Predicted				
		NDR	Mild	Moderate	Severe	PDR
Actual	NDR	344	2	3	1	3
	Mild	3	76	4	2	2
	Moderate	3	2	195	3	2
	Severe	2	1	1	35	1
	PDR	1	2	2	3	41

Based on the confusion matrix the following metrics are computed

Table 5
Metrics computed from Confusion matrix

Class	Precision (%)	Accuracy (%)	Recall (%)	F1 Score
NDR	97	97.54	97	97
Mild	87	97.68	93	90
Moderate	95	97.27	95	95
Severe	88	98.09	80	83
PDR	85	97.95	84	85
Overall Accuracy	94.27%			

The resulting findings are compared to those of other cutting-edge approaches. Table 8 displays the outcomes of the various approaches used across the five categories.

Table 6

Results Comparison five classes

Method	Precision (%)	Recall (%)	Accuracy (%)
DRISTI (VGG16 + Capsule) [17]	91	88	82.06
EfficientNet-B3 [18]	59	66	84.86
Resnet50 + Capsule [19]	59	69	76.80
FedDDR (Proposed Method)	89	90	94.27

5. Conclusion and Future work

In this paper, a federated deep learning model for detection of diabetic retinopathy from retinal images. The retinal fundus images are classified into five classes. A modified DenseNet model with federated learning approach is proposed in this work. The federated learning makes the training process distributed and hence improving the overall performance of the classifier. The patient privacy is also intact as their data remains on their device. Also, the limitation of limited medical data for training is overcome as the data is used from multiple devices. The simulations are done using tensorflow federated learning module. The results show that the proposed method achieves 94.275 overall accuracy and class wise accuracy is also high. The comparison with other state of the art methods reveal that the proposed method outperforms the existing state of the art methods.

6. References

- [1] S. Kumar NC and R. Y, "Optimized maximum principal curvatures based segmentation of blood vessels from retinal images," *Biomedical Research*, vol. 30, no. 2, 2019.
- [2] G. Hassan, N. El-Bendary, A. E. Hassanien, A. Fahmy, S. Abullah M., and V. Snasel, "Retinal blood vessel segmentation approach based on mathematical morphology," *Procedia Computer Science*, vol. 65, pp. 612–622, 2015.
- [3] S. S. Mondal, N. Mandal, A. Singh, and K. K. Singh, "Blood vessel detection from retinal fundas images using GIFKCN classifier," *Procedia Computer Science*, vol. 167, pp. 2060–2069, 2020.
- [4] R. Reguant, S. Brunak, and S. Saha, "Understanding inherent image features in CNN-based assessment of diabetic retinopathy," *Scientific Reports*, vol. 11, no. 1, 2021.
- [5] J. Benson, J. Maynard, G. Zamora, H. Carrillo, J. Wigdahl, S. Nemeth, S. Barriga, T. Estrada, and P. Soliz, "Transfer learning for diabetic retinopathy," *Medical Imaging 2018: Image Processing*, 2018.
- [6] I. Kandel and M. Castelli, "Transfer learning with convolutional neural networks for diabetic retinopathy image classification. A Review," *Applied Sciences*, vol. 10, no. 6, p. 2021, 2020.
- [7] N. Sikder, M. Masud, A. K. Bairagi, A. S. Arif, A.-A. Nahid, and H. A. Alhumyani, "Severity classification of diabetic retinopathy using an ensemble learning algorithm through analyzing retinal images," *Symmetry*, vol. 13, no. 4, p. 670, 2021.
- [8] Z. Shen, Q. Wu, Z. Wang, G. Chen, and B. Lin, "Diabetic retinopathy prediction by ensemble learning based on biochemical and physical data," *Sensors*, vol. 21, no. 11, p. 3663, 2021.
- [9] G. T. Reddy, S. Bhattacharya, S. Siva Ramakrishnan, C. L. Chowdhary, S. Hakak, R. Kaluri, and M. Praveen Kumar Reddy, "An ensemble based machine learning model for diabetic retinopathy classification," *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, 2020.
- [10] Q. Yang, Y. Liu, Y. Cheng, Y. Kang, T. Chen, and H. Yu, "Federated learning," *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 13, no. 3, pp. 1–207, 2019.
- [11] N. Rieke, J. Hancox, W. Li, F. Milletari, H. R. Roth, S. Albarqouni, S. Bakas, M. N. Galtier, B. A. Landman, K. Maier-Hein, S. Ourselin, M. Sheller, R. M. Summers, A. Trask, D. Xu, M. Baust, and M. J. Cardoso, "The Future of Digital Health with Federated Learning," *npj Digital Medicine*, vol. 3, no. 1, 2020.
- [12] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *2017 IEEE Conference on*

Computer Vision and Pattern Recognition (CVPR), 2017.

- [13] Goyal, L., Dhull, A., Singh, A., Kukreja, S., & Singh, K. K. (2023). VGG-COVIDNet: A Novel model for COVID detection from X-Ray and CT Scan images. *Procedia computer science*, 218, 1926-1935.
- [14] Z. Zhang, "Improved adam optimizer for Deep Neural Networks," *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)*, 2018.
- [15] Konečný, J., McMahan, H.B., Yu, F.X., Richtárik, P., Suresh, A.T. and Bacon, D., 2016. Federated learning: Strategies for improving communication efficiency. arXiv preprint arXiv:1610.05492.
- [16] N. Sikder, M. Masud, A. K. Bairagi, A. S. Arif, A.-A. Nahid, and H. A. Alhumyani, "Severity classification of diabetic retinopathy using an ensemble learning algorithm through analyzing retinal images," *Symmetry*, vol. 13, no. 4, p. 670, 2021.
- [17] G. Kumar, S. Chatterjee, and C. Chattopadhyay, "Dristi: A hybrid deep neural network for diabetic retinopathy diagnosis," *Signal, Image and Video Processing*, vol. 15, no. 8, pp. 1679–1686, 2021.
- [18] A. Sugeno, Y. Ishikawa, T. Ohshima, and R. Muramatsu, "Simple methods for the lesion detection and severity grading of diabetic retinopathy by Image Processing and Transfer Learning," *Computers in Biology and Medicine*, vol. 137, p. 104795, 2021.
- [19] G. Kumar, S. Chatterjee, and C. Chattopadhyay, "Dristi: A hybrid deep neural network for diabetic retinopathy diagnosis," *Signal, Image and Video Processing*, vol. 15, no. 8, pp. 1679–1686, 2021.
- [20] Y. Zhou, X. Wang, M. Zhang, J. Zhu, R. Zheng, and Q. Wu, "MPCE: A maximum probability based cross entropy loss function for neural network classification," *IEEE Access*, vol. 7, pp. 146331–146341, 2019.
- [21] Y. Tolstyak and M. Havryliuk, 'An Assessment of the Transplant's Survival Level for Recipients after Kidney Transplantations using Cox Proportional-Hazards Model', *CEUR-WS.org*, vol. 3302, pp. 260–265, 2022.
- [22] Y. Tolstyak, V. Chopyak, and M. Havryliuk, 'An investigation of the primary immunosuppressive therapy's influence on kidney transplant survival at one month after transplantation', *Transplant Immunology*, vol. 78, p. 101832, Jun. 2023, doi: 10.1016/j.trim.2023.101832.