International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Ingénierie des Systèmes d'Information
Vol. 28, No. 5, October, 2023, pp. 1377-1386
Journal homepage: http://iieta.org/journals/isi

Enhanced Classification of Diabetic Retinopathy via Vessel Segmentation: A Deep Ensemble Learning Approach

Sanjay Tanaji Sanamdikar [1*], Mayura Vishal Shelke [2], Jyoti Prashant Rothe [3]

[1] Department of Instrumentation, PDEAs College of Engineering Manjari, Pune 411028, Maharashtra, India
[2] Department of AI and DS, All India Shree Shivaji Memorial Society, Institute of Information Technology, Pune 411001, Maharashtra, India
[3] Department of Electrical Engineering, St. Vincent Pallotti College of Engineering & Technology, Nagpur 441108, Maharashtra, India

Corresponding Author Email: sanamdikar123@gmail.com

https://doi.org/10.18280/isi.280526

Received: 22 June 2023
Revised: 3 September 2023
Accepted: 11 October 2023
Available online: 31 October 2023

Keywords:
diabetic retinopathy, vessel segmentation, deep ensemble learning, computer vision, DRIVE dataset, Canny operator, blood vessel segmentation

ABSTRACT

Diabetic Retinopathy (DR), a medical condition that impairs the blood vessels within the eye, is increasingly prevalent. Unchecked progression of DR can lead to significant visual impairment or total blindness. Traditional techniques for automatic DR detection, primarily reliant on computer vision systems, often fail to adequately encapsulate the inherent complexity of the disease, resulting in suboptimal categorization of DR stages, particularly the early ones. However, deep ensemble learning has emerged as a potent tool for the accurate detection and classification of DR using retinal images. In this study, deep ensemble models are proposed that initially segment the retinal image using the Canny operator and subsequently detect and classify all DR categories using the publicly available DRIVE dataset. Each model, crafted with subtle architectural distinctions or trained on distinct data subsets, was designed to capture varying disease attributes. A threshold was established to accurately categorize DR severity into mild, moderate, or severe cases. The results indicate a significant enhancement in the performance of both segmentation and DR detection through deep ensemble learning, compared to individual models. The ensemble approach effectively amalgamated the collective knowledge of the models, yielding superior accuracy, robustness to data variations, and improved generalization capabilities. This cost-effective computational method achieves an accuracy score of 98.65% in DR detection and classification. By synthesizing the predictions of multiple models, the ensemble captured a wider spectrum of disease patterns, thereby bolstering the system's overall effectiveness in DR diagnosis. The findings underscore the enhanced accuracy and robustness attained through the ensemble approach, surpassing the performance of individual models.

1. INTRODUCTION

Diabetic retinopathy (DR) stands as the leading cause of vision impairment among individuals aged between 30 and 55 years [1]. A staggering 90% of confirmed DR incidents could potentially be circumvented through adequate eye care and timely screening procedures [2]. Predominantly affecting long-term diabetes patients, DR alters the lens curvature, giving rise to visual disturbances, and ranks among the most severe ocular disorders [3]. Furthermore, DR emerges as a primary contributor to blindness, with its insidious progression rendering it treatable if identified at an early stage. However, delayed detection can inflict irreversible damage to the human retina, culminating in permanent blindness [4].

The onset of DR is attributed to alterations in the retinal blood vessels. The retina, a thin photosensitive layer lining the rear of the eye, is vulnerable to elevated blood glucose levels, which can inflict damage on blood vessels. Consequently, these vessels can thicken and leak, precipitating vision loss. DR is generally categorized into four stages: mild, moderate, severe non-proliferative, and proliferative. The disease enters the initial stage, known as mild non-proliferative, upon the manifestation of sac-like swelling in small sections of the retinal blood vessels [5, 6]. The condition progresses to the moderate non-proliferative stage when blockages occur in certain retinal blood vessels. The severe non-proliferative stage is marked by additional vessel blockages, depriving sections of the retina of adequate blood supply [7]. The final stage, proliferation, is characterized by the formation of new, yet unstable and distorted, blood vessels in the retina. These vessels may subsequently hemorrhage, potentially inducing total vision loss or blindness [8].

Routine retinal examinations of diabetic patients prove instrumental in the early detection of DR, thereby mitigating the risk of blindness. However, retinal screening is a time-intensive process that relies heavily on the expertise of ophthalmologists to dissect fundus images and examine retinal vessels. Figure 1 illustrates images of a healthy retina juxtaposed against a DR-afflicted retina. Diabetic eye disease encompasses two stages: Non-Proliferative Diabetic Retinopathy (NPDR) and Proliferative Diabetic Retinopathy (PDR).

Given the stealthy progression of DR, often remaining undetected until vision loss transpires, timely intervention is of paramount importance. Recognizing the criticality of early DR detection, the role of retinal blood vessel segmentation assumes greater significance. For the purpose of automated eye disease screening, fundus images need to be segmented into regions that correspond to specific ocular sections.

The primary objective of this study is to enable early detection and prevention of vision loss induced by diabetic retinopathy (DR). A crucial component in the process of fundus imaging is the segmentation of retinal blood vessels [9]. Manual recognition of these vessels requires the expertise of a seasoned professional, a task that is both time-consuming and labor-intensive [10]. This is where computer-assisted diagnostics gain prominence. The employment of computerized image analysis tools offers a swift, cost-effective means to identify DR. The automated system enables large-scale screenings at a reduced cost, with minimal human error, even in remote locations [11, 12]. In the process of detecting DR, segmentation of the retinal blood vessels emerges as a crucial step.

This paper introduces notable contributions as follows:

• An automated deep learning procedure is presented that effectively extracts pertinent features from retinal images.

• The proposed approach integrates advanced image segmentation techniques, enhancing the precision of retinal image segmentation, a pivotal step in DR diagnosis.

• The proposed deep ensemble model expands beyond the detection of moderate-stage DR to encompass all DR classes, ensuring that early-stage detection is an integral part of the comprehensive solution.

The structure of the paper is organized as follows. Section 2 provides a review of recent studies related to DR detection and classification. The proposed methodology is presented in Section 3. Section 4 delves into the result analysis of the proposed ensemble approach, juxtaposing the final results with state-of-the-art methods. Conclusions and prospective scope for future research are discussed in Section 5.

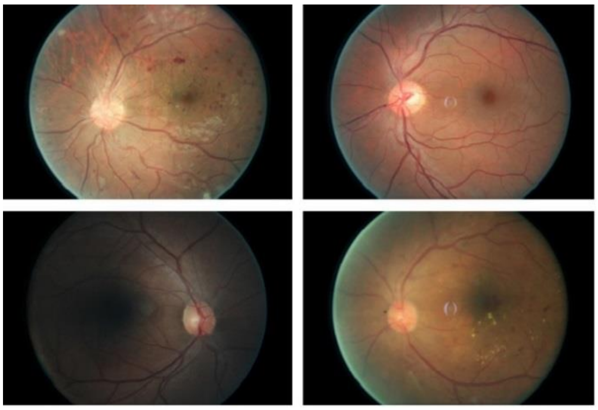

Figure 1. Samples images of retinal. Two retinal images on the bottom side represent diabetic retinopathy, although the top side retinal images represent normal retinal images [7]

2. RELATED WORK

This section provides a comprehensive review of the existing literature on diabetic retinopathy (DR) classification, feature extraction, and vessel segmentation. A comparative analysis of the results obtained through various methods is conducted based on key performance indicators such as accuracy, sensitivity, specificity, precision, and recall.

A deep learning approach employing convolutional neural networks (CNNs) for DR detection was proposed in the study of Gulshan et al. [13]. The ensemble-based methodology achieved a 31% recall and 96% specificity; however, it failed to detect the mild stage of the disease. A notable limitation was its inability to accurately classify the infected stage.

An ensemble deep learning system devised for DR diagnosis was proposed by Gao et al. [14]. This ensemble was composed of several CNN models with diverse architectures and configurations. Notably, this system surpassed the performance of individual CNN models on publicly available DR datasets, demonstrating high accuracy.

Deep ensemble learning method specifically for DR detection in the study of Ai et al. [15]. This ensemble was an amalgamation of multiple CNN models, each trained on different subsets of the dataset. The approach exhibited improved classification performance and robustness relative to individual CNN models.

Pak et al. [16] suggested a deep-learning approach predicated on various DR images and classifiers for DR screening. They amalgamated multiple features extracted from retinal images and trained an ensemble of classifiers. As a result, the DR detection accuracy was significantly enhanced compared to individual classifiers.

Johnson et al. [17] introduced an ensemble of deep-learning models for DR detection. This ensemble consisted of multiple CNNs trained on an extensive dataset, and it achieved state-of-the-art performance on the Kaggle Diabetic Retinopathy Detection Challenge dataset.

In the study of Asiri et al. [18], an examination of several deep neural network (DNN) techniques employed for DR detection was conducted. The study presented 18 DR databases suitable for four distinct DR recognition tasks, providing comprehensive data regarding DR.

In the study of Hosseinzadeh et al. [19], a comprehensive evaluation of the classification performance of Deep Neural Networks (DNNs) applied to DR images was conducted. Two tabular summaries encapsulating the features, findings, databases, and contributors of the examined research were utilized in the investigation.

An exhaustive review of Artificial Intelligence (AI) methodologies for DR detection was conducted by Ishtiaq et al. [20], incorporating 75 pieces of literature from nine scientific sources. The findings of this assessment were presented from various perspectives, including databases, image pre-processing techniques, machine learning (ML) strategies, DNN-based approaches, and performance analysis.

Alyoubi et al. [21] undertook a study of the most recent DNN methodologies used for DR image detection and classification. The study analyzed 34 pieces of research, which employed 12 publicly available DR datasets in conjunction with imaging pre-processing methods. These datasets spanned binary, multi-level, lesion-based, and other screening methodologies.

Stolte and Fang [22] provided an in-depth analysis of the current techniques employed for DR detection. A comprehensive overview of DR, existing technology, and related information was presented in conjunction with a description of the concepts used for DR identification and tracking. This study served as an extensive assessment of healthcare image analysis in the context of DR.

In the study of Badar et al. [23], the utilization of DNNs for the detection of retinal vessel anomalies in DR was thoroughly analyzed. The analysis evaluated existing literature based on publicly available databases of fundus and optical coherent tomography retinal vessels, as well as DNN architectures. The authors concluded that DNNs could potentially replace traditional classification techniques.

Chu et al. [24] conducted a systematic review that incorporated both DNN and DR, examining 16 distinguished studies. These studies were scrutinized based on the chosen database and performance parameters. Notably, this work diverged from other reviews by focusing on the potential algorithmic constraints inherent in each study. A comprehensive review was subsequently released, probing the potential application of DNN in identifying diabetic eye complications. This review encompassed 66 publications across 11 scientific databases and included an investigation into DR and three discrete ocular conditions. The databases, image processing methods, and deep classification techniques such as Transfer Learning, DNN, and combined DNN and ML classification employed for each ocular condition were detailed. Consequently, the authors demonstrated the efficacy of DNN in producing beneficial outcomes for DR detection.

Sarki et al. [25] offered a singular study in 2020, concentrating on DR classification via transfer learning. This study evaluated the model, databases, optimization techniques, and fine-tuning methods employed across 19 investigations. Ultimately, the researchers demonstrated the potential of transfer learning to significantly contribute to DR detection and monitoring.

The literature review indicates that researchers have devised and proposed a diverse array of methodologies for the detection and classification of different stages of Diabetic Retinopathy (DR). While numerous techniques have been put forth for DR detection, it is crucial to note the particular effectiveness of multiclassification techniques in determining the severity levels of DR stages. Ensemble learning emerges as a pivotal tool in DR detection, harnessing the combined strength of multiple models to enhance accuracy, robustness, generalization, and bias reduction. This ultimately elevates the efficacy of diagnosing and managing this sight-threatening disease.

There are numerous publicly available databases that contain images suitable for DR detection or classification.

3. PROPOSED METHODOLOGY

This section presents the details of implementation and result analysis. The deep learning method is implemented over the multiclass dataset for early detection of diabetic retinopathy. This study used the publicly available DRIVE dataset which contains five classes. Figure 2 shows the complete architecture of the proposed Model that involves multiple deep learning models, such as InceptionResNetV2, EfficientNetV2L, Xception, DenseNet201, and VGG19, with slight variations in architecture. These models are then combined through an ensemble approach, where their predictions are aggregated using weighted averaging. The ensemble of models harnesses the collective knowledge and diversity of the individual models, improving the system's ability to capture various DR patterns and reduce overfitting. By leveraging the power of deep ensemble learning, this methodology aims to provide a robust and accurate solution for the early detection and classification of DR, facilitating timely interventions and effective management of this sight-threatening disease.

3.1 Dataset description

In this study, a publicly available DRIVE dataset has been used. The DRIVE datasets, which include 2750 diabetic images of patients aged 25 to 90 that among the results of a testing program for diabetic retinopathy. It contains a total of five classes 'Healthy', 'Mild_ DR', 'Moderate_DR', 'Proliferate_DR', and 'Severe_DR' images. Each class shows the severity of DR with a scale of 0 to 4 as shown in Figure 3.

The size of each image is 224 × 224 pixels in dataset. The dataset is obtained from (URL: https://www.kaggle.com/datasets/sovitrath/diabetic-retinopathy-224x224-2019-data) was used to measure the performance of Proposed DR detection and classification model.

3.2 Data pre-processing

When inputting the DR images to a proposed deep ensemble model during the training and testing of the model, the intensity of pixel value in the DR images must be scaled. During the data pre-processing, a 1/255 rescale factor was used to transform each pixel value of the DR image from the range (0, 255) because converts all the DR images similarly and train the model with relevant features and patterns more effectively, and also leading to better performance in identifying DR.

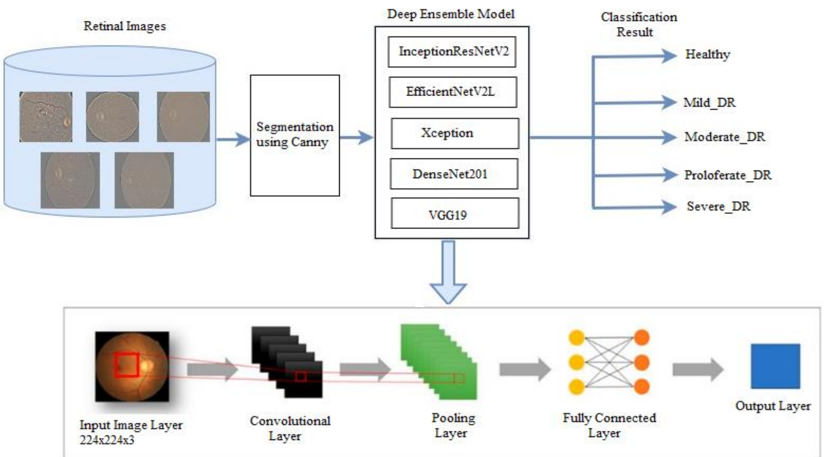

Figure 2. Complete architecture of proposed ensemble model

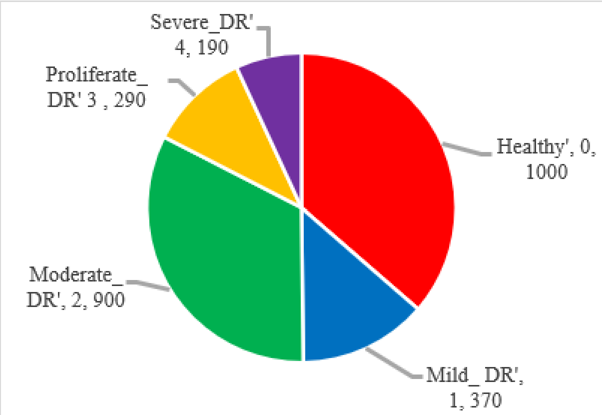

Figure 3. Grade-wise distribution of the DR dataset

3.3 Data augmentation

The absence of sufficient amounts of pertinent, class-specific data is a barrier to the development of an efficient ensemble model for the classification of retinal images. Positive disease symptoms are typically hard to identify. Using image augmentation techniques may artificially expand the quantity of the training and validation datasets, giving the proposed model much more data for training and validation. Making the proposed model more able to detect new variations of training and validation data, and may increase the accuracy of the proposed model and avoid the overfitting problem. The augmentation is performed by flipping and rotating the DR image by 90^0 degrees. Figure 4 shows the dataset distribution used for training the proposed ensemble model. The complete dataset is divided into 2750 images for Training, 540 images for Validation, and 540 images for Testing after data augmentation generating 11000 images for training data, and generating 2160 images for validation for 5 classes of diabetic retinopathy in the dataset.

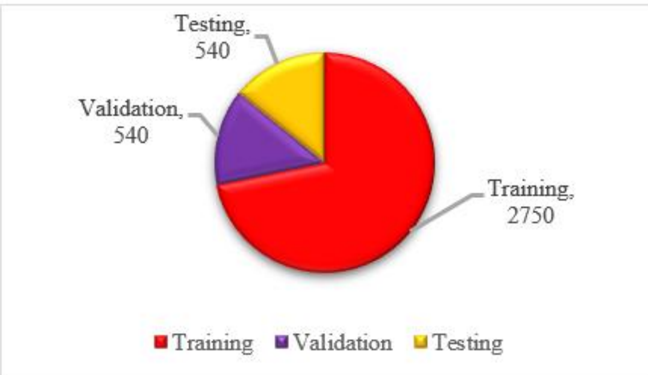

Figure 4. Distribution dataset

3.4 Segmentation using Canny operator

Apply the Canny operator to the pre-processed retinal images. This will help to detect edges and boundaries of blood vessels that are relevant to diabetic retinopathy. By detecting these edges, the operator can highlight potential areas of interest for further analysis. Once the edges are detected, proceed to segment the blood vessels of the retinal images.

$$I_D = \int_{-R}^{R} D(-x) f(x) dx \tag{1}$$

where, D(x) is the edges of retinal images, f(x) represents the filter response with the range -R to R. The RMS values can be measured by:

$$I_b = \sqrt{b_0^2 \int_{-R}^{R} f^2(x) dx} \tag{2}$$

where, b_0^2 represents the RMS values of distortion. The final outcome SNR ratio can be evaluated by

$$SNR = \frac{\int_{-R}^{R} D(-x) f(x) dx}{\sqrt{b_0^2 \int_{-R}^{R} f^2(x) dx}} \tag{3}$$

where, $I_b(x)$ represents the filter response to distortion. $k_s(x)$ is the filter response to edge of vessels. At $x = x0$ is the complete local maximum response.

$$I_b'(x0) + I_s'(x0) = 0 \tag{4}$$

Using Taylor expansion $I_s(0)=0$

$$I_s'(x0) = I_s'(0) + I_s''(x0) + Q\left(x_0^2\right) = 0 \tag{5}$$

Using Eqs. (4) and (5)

$$I_s''(x0) = I_s'(x0) \tag{6}$$

$I_s'(x0)$ is the Gaussian randomized samples

$$G\left[I_s'\left(x_0^2\right)\right] = x_0^2 \int_{-R}^{R} f'^2(x) dx \tag{7}$$

$$G\left(x_0^2\right) = \frac{x_0^2 \int_{-R}^{R} f'^2(x) dx}{\left[\int_{-R}^{R} D'(-x) f(x) dx\right]^2} = \partial x_0^2 \tag{8}$$

$$Localize = \frac{\int_{-R}^{R} D(-x) f'(x) dx}{b_0 \sqrt{\int_{-R}^{R} f'^2(x) dx}} \tag{9}$$

The combined result of Eqs. (3) and (9) must be maximized for the purpose to minimize the problem.

$$\frac{\int_{-R}^{R} D(-x) f(x) dx}{\sqrt{b_0^2 \int_{-R}^{R} f^2(x) dx}} = \frac{\int_{-R}^{R} D(-x) f'(x) dx}{b_0 \sqrt{\int_{-R}^{R} f'^2(x) dx}} \tag{10}$$

Figure 5 displays the sequential stages of segmentation results obtained from a retinal image using the Canny edge detection operator. This illustration underscores the effectiveness of the proposed approach in achieving improved segmentation outcomes. The segmentation, every blood vessel within the retinal image becomes distinctly discernible, enhancing the clarity of the visualization.

3.5 Ensemble model

To provide a result that is more effective, the ensemble model mixes a number of distinct deep-learning classifiers. Deep ensemble models integrate the benefits of deep learning models and improve the prediction results of the final ensemble model. In this paper InceptionResNetV2, EfficientNetV2L, Xception, DenseNet201, VGG19 architecture is used because we have the limited datasets so this model will help to reduce overfitting and high accuracy and ability to capture complex patterns for DR detection and classification. In the proposed model, in addition to ensemble elements, additionally, stack high-level features taken from subsequent convolutional layers to enhance the effectiveness of DR grade detection and classification. Figure 6 shows the deep ensemble model structure with the size of classifiers.

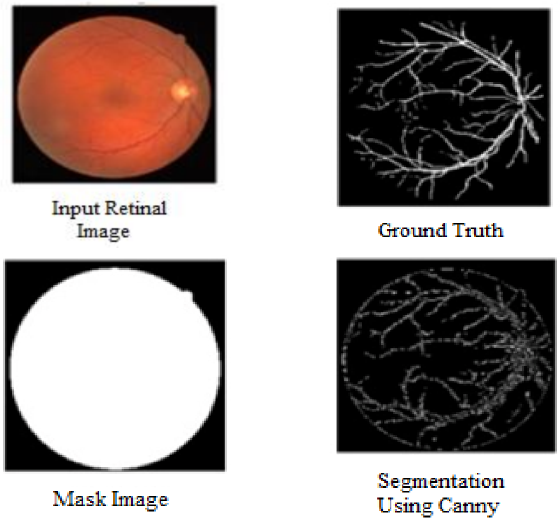

Figure 5. (a) Input retinal image (b) Ground truth (c) Mask image (d) Segmentation result using Canny

In multi-class classification using ensemble learning, the mathematical model extends the weighted averaging approach to handle multiple classes. Let's consider a scenario where we have an ensemble of N individual models for multi-class classification.

The input is fed to the proposed deep ensemble model and output E_c^i is obtained from each deep model ($i \in \{1,2,3...n\}$) for each DR class $c \in C$. The proposed deep ensemble algorithms normalize the output of each DR class. The learning rate of each classifier is 0.0001. Therefore, all the classifier outputs are combined using the SoftMax layer to get the outcome from the proposed deep ensemble model.

$$o_i = F\left(y=c|\ p, s_i\right) = \frac{e^{E_c^i}}{\sum_{c\in C} e^{E_c^i}} \tag{11}$$

$$\text{argmax}\, F(y=c|\ p) = \frac{e^{o_i}}{\sum_{i\in I} e^{o_i}} \tag{12}$$

where, argmax () returns the index corresponding to the highest value in the prediction. Eq. (1) applies the softmax layer across the class value and Eq. (2) applies the highest value of posterior across the final class result to measure and evaluate the finalized class labelling. The model loss is measured using the highest probability of loss function.

The loss function is utilized to enhance the proposed ensemble model training. It speeds up the resolution and lowers back propagation errors. Eq. (3) illustrates the mathematical concept.

$$f^t(W) = -\sum_{i=1}^{n} y_i' \log\left(y_i\right) = -\sum_{i=1}^{n}\left(y_{\max} - y_u\right)\bar{y}_l \log\left(y_i\right) \tag{13}$$

where, y_{max} is the maximum amongst n classes with the actual class. $\bar{y}$ represent the vector of actual classes. y_i' represent the i^{th} coordinate of the vector y_i.

The mathematical model for multi-class classification using ensemble learning involves the aggregation of individual model predictions using weighted averaging, normalization of weights, and selecting the class with the highest probability or score for the final prediction.

3.5.1 InceptionResNetV2

The complex design of Inception-ResNet-V2 is used to extract critical characteristics from diabetic retinopathy images. Standard convolutional layers and a maximum aggregation layer make up the network's first layers. The subsequent step convolutionally combines an entry while applying various filter sizes for each convolution. The network employs dropout layers to keep the filter outputs equal to zero in the next portions of the network, which are repeated 12 or 22 times each, to avoid overfitting [26, 27]. The highest score that Keras has offered for ImageNet is 0.963.

3.5.2 EfficientNetV2

A deep learning network design called EfficientNet was introduced in 2019 [28]. This model demonstrates the connection between three concepts that have a big influence on how well deep network topology functions. They are depth, breadth, and high resolution, respectively. The composite scaling technique is used in this design [29]. This approach starts with the grid search optimizer. The system can correlate between scaled sizes of various sizes to this approach.

3.5.3 Xception

François Chollet expanded the Inception model by introducing the Xception framework [30]. A sequential stacking of depth-wise separated convolutional layers with feedback connections makes up the model design. The depth-wise separated convolution attempts to lower memory and operational costs. With the exception of the initial and final modules, all of the 13 modules in Xception's 36 convolutional have linear residual blocks. The training of channel-wise with space-wise parameters is separated by the separated convolution in Xception. Moreover, the residual connection provides a bypass in the sequential model to address the problems of fading gradations and graphical constraints [31].

3.5.4 DenseNet201

According to recent studies, CNN can get steadily greater in size, much more accurate, and simpler to learn since there are less interactions among each layer nearest to the outcome and the layer nearest to the input. For humans to recognize this reality, the DenseNet, which connects each layer with the following layer in a feed-forward fashion, was presented [32]. Instead of normal CNN with N layers and N interconnections between each layer and its subsequent layer, this suggested network has $N(N+1)/2$deep linkages. Its unique feature representations are utilized as inputs for every level following it, whereas the pertinent feature maps of every level previously it was used as inputs for each of them [33].

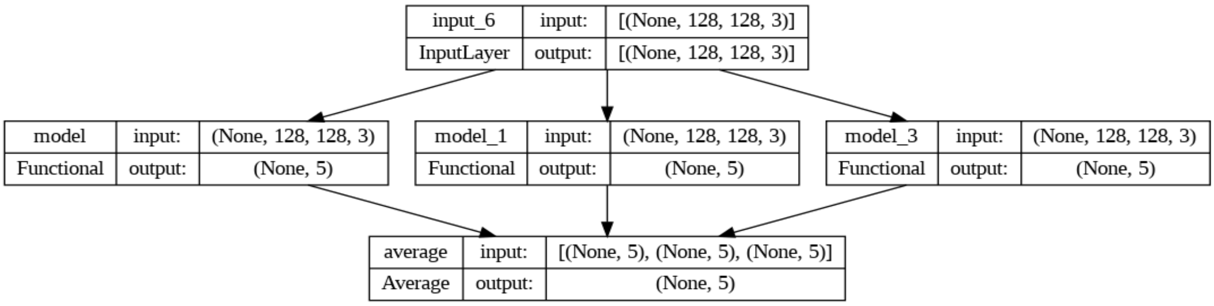

Figure 6. Structure of proposed ensemble model

3.5.5 VGG19

This paper, examines how the convolutional network's depth affects performance in a context where a lot of DR images need to be recognized. The primary contribution is a detailed analysis of networks with increased depth utilizing a model design with relatively tiny (3x3) convolution filters, that demonstrates that extending the depth to 16-19 weighted layers may significantly outperform existing setups [34].

Algorithm: Ensemble Deep Model

Input: DR Images (P, Q); Where $Q = \{q/q \in \{Healthy, Mild\_DR, Moderate\_DR, Proliferate\_DR, Severe\_D$

Output: To trained the Ensemble Deep model ($i \in \{1,2,3...n\}$) that classify the DR images $p \in P$

Step 1: Preprocessing

a. Rescale all DR images using 1/255 Rescale Factor

b. Data Augmentation

c. Normalize each DR image

Step 2: Distribution Dataset: Divide the dataset into Training, Validation, Testing

Step 3: Build Deep Ensemble Model M = $(InceptionResNetV2, EfficientNetV2L, Xception, DenseNet2$ with the 128,128,3-classifier size

for each $\forall h \in M$

Learning Rate $\gamma = 0.0001$

for epochs = 1 to 50 do

for each batch (P, Q) belong to (p_{train}, q_{train}) **do**

Update convolutional Layer of each model h(.) using equation 1 and 2

Calculate Training and Validation Loss of Ensemble Model

if Patience = 3 and Best_weight = True **do**

Early Stopping

End

End

End

for each $p \in p_{test}$ **do**

Ensemble the final output of deep models

End

The above algorithm outlines the process of training an Ensemble Deep Model for classifying diabetic retinopathy (DR) images. In this algorithm used early stoppping criteria. "patience" is a hyperparameter that controls how many epochs the algorithm will wait for validation loss improvement before stopping the training process. It helps ensure that the proposed ensemble model doesn't overfit the training data and can potentially save training time by stopping.

4. RESULT ANALYSIS

4.1 Evaluation parameters

The performance and effectiveness of an ensemble model, used for the identification and classification of diabetic retinopathy, are frequently assessed using the evaluation parameters stated in the equations provided follows. These evaluation parameters offer several perspectives on the Ensemble model's performance and may be used to pinpoint the model's advantages and disadvantages with regard to accurately classifying DR images. Statistical measurements are employed to provide thorough DR images of the ensemble model efficiency based on the prior legislation. Accuracy, precision, recall, and f1-score can be used to assess how well the proposed ensemble model performs. The mathematical formulas used to evaluate the performance of the proposed model are listed below.

A confusion matrix is a table that visualizes the performance of ensemble classifiers tested versus actual classification.

True Positives (T_P): True positive would be when the model correctly identifies a positive class as having diabetic retinopathy.

True Negatives (T_N): True negative would be when the model correctly identifies an instance as not having diabetic retinopathy.

False Positives (F_P): False positive would be when the model incorrectly identifies an instance as having diabetic retinopathy when it does not.

False Negatives (F_N): False negative would be when the model incorrectly identifies an instance as not having diabetic retinopathy when it actually does.

$$Accuracy = \frac{T\_P + T\_N}{T\_P + T\_N + F\_P + F\_N} \tag{14}$$

$$Precision = \frac{T\_P}{T\_P + F\_P} \tag{15}$$

$$Recall = \frac{T\_P}{T\_P + F\_N} \tag{16}$$

$$F1-Score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{17}$$

In this section, presented the ensemble models with deep learning algorithms are given for each problem, using the diabetic retinopathy original dataset for predicting and classification of various grades of DR problems. The System configuration and size of the ensembled models are given in Table 1. The choice of these specific hardware and software configurations is likely based on the need for robust and efficient deep learning model development. These configurations can handle the computational demands of deep learning, accelerate training times with GPU support, and provide a user-friendly environment for building and experimenting with ensemble models.

The ensemble Approach was used to detect and classify multiclass diabetic retinopathy images. The input size of all classifiers used in the ensemble model is the same (128,128,3). Table 2 shows the performance evaluation of Ensemble classifiers (InceptionResNetV2, EfficientNetV2L, Xception, DenseNet201, VGG19) for the detection and classification of diabetic retinopathy. The evaluation of parameters is accuracy, precision, recall, and F-score. It is clearly shown that the proposed ensemble approach achieves a 98.65% accuracy score, precision is 98.73% Recall is 98.62% and F1-score is 97.77%. The proposed ensemble model has a very minimal loss of 0.0469.

Table 1. System configuration

Name	Parameters
Memory	64 GB
Processor	Persistence-M
Server model	Hp z6 g4
Graphics	NVIDIA-SMI 510.47.03
OS	Windows 10
Language	Python 3
Framework	Jupyter Notebook

Table 2. Performance analysis of the proposed ensemble approach for DR detection

Parameters	Predicted Values in %
Accuracy	98.65
Precision	98.73
Recall	98.62
F1-Score	97.77
Loss	0.0469

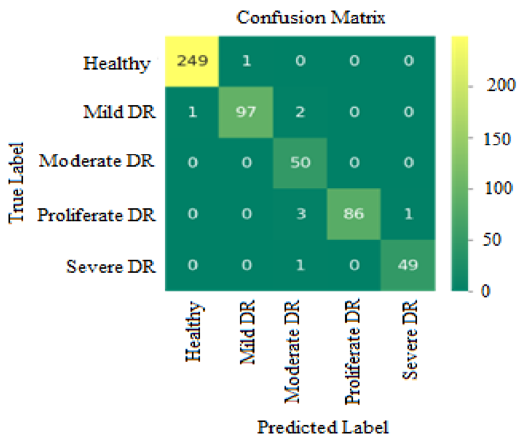

Figure 7. Confusion matrix for five class

Figure 7 shows the confusion matrix with tabular representation of the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions for each class. The matrix allows for an in-depth analysis of the model's accuracy, precision, recall, and F1 score for each class, enabling a better understanding of its strengths and weaknesses. By examining the confusion matrix, valuable insights can be gained into the model's ability to correctly classify DR severity levels, facilitating improvements in the ensemble learning approach and enhancing the overall effectiveness of DR detection and classification.

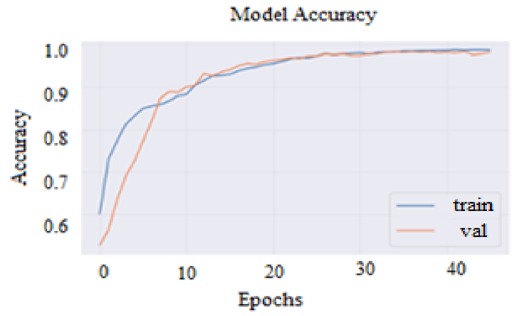

Figure 8. Training and validation of model accuracy

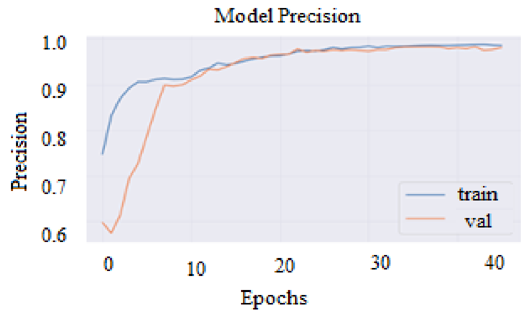

Figure 9. Training and validation of model precision

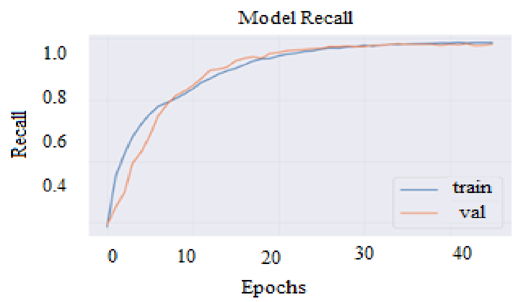

Figure 10. Training and validation of model recall

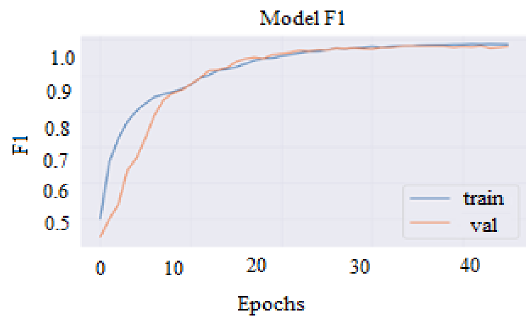

Figure 11. Training and validation of model F1

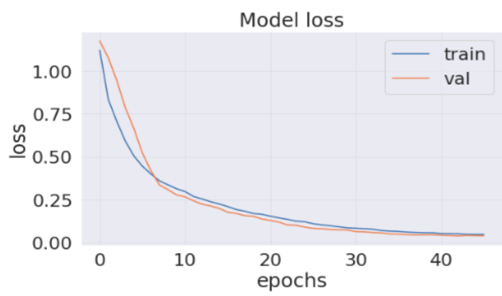

Figure 12. Training and validation of model loss

Figure 8 to Figure 12 show the training and validation score of the proposed model in terms of accuracy, precision, recall, and loss. It is clearly observed that the training and validation accuracy comes out to be 98.15% and 98.00% respectively. Training and validation precision comes out to be 98.14% and 98.00%. Training and validation recall comes out to be 98.14% and 98.00%. Training and validation model loss comes out to be 0.0469% and 0.0391%. Table 3 shows the class-wise result analysis of the proposed ensemble approach. It clearly shows that the accuracy score for each class has reached 99%.

Table 3. Class-wise result analysis of the proposed ensemble approach

Class	N (Truth)	N (Classified)	Accuracy	Precision	Recall	F1 Score
Healthy	250	250	99.63%	1.0	1.0	1.0
Mild_DR	98	100	99.26	0.97	0.99	0.98
Moderate_DR	56	50	98.89	1.0	0.89	0.94
Proliferate_DR	86	90	99.26	0.96	1.0	0.98
Severe_DR	50	50	99.63	0.98	0.98	0.98

5. COMPARATIVE ANALYSIS

The accuracy score and other evaluation parameters obtained by proposed ensembled models are somewhat nearer to those found in previous research as shown in Table 4. This paper might be extended to employ a DR image dataset with more data in order to reach a higher accuracy score. It could also offer a prototype for future studies. DR detection and classification using the proposed ensemble model to detect DR more precisely and quickly. The proposed classifier is computationally cost-effective with a detection accuracy score of 98.65% which is better than other state-of-art methods.

Table 4. Comparative analysis of proposed ensembled model with existing state-of-art methods

State-of-Art Methods	No of Epochs	Training Time in Hr.	Testing Time in Sec	Accuracy	Precision	Recall	F1 Score
Hosseinzadeh et al. [19]	50	3.75	5.86	83.09	87.00	NA	88.24
Kumar et al. [35]	50	14.20	30.21	96.24	96.00	96.00	NA
Sugeno et al. [6]	45	10.45	12.32	96.07	95.00	96.00	NA
Mondal et al. [7]	50	3.45	3.64	86.08	76.00	82.00	80.00
Deshmukh and Roy [36]	60	4.50	3.50	98.00	98.2	94.5	95.00
Deshmukh et al. [37]	60	2.20	4.10	93.86	NA	94.81	93.30
Khare [38]	1000	NA	NA	98.6	NA	NA	NA
Dondeti et al. [39]	NA	NA	NA	77.90	76	77	75
Proposed Deep Ensemble	**50**	**1.40**	**2.35**	**98.65**	**98.73**	**98.62**	**97.77**

6. CONCLUSIONS

Diabetic retinopathy is one of the fastest-growing diseases in the world. The DR has several numbers of grade Mild-to-Severe that may be caused to complete vision loss or blur vision. The traditional approach for detecting the DR is a time-consuming task and error-prone. So, need an automated intelligent system to detect and classify the DR according to their grade. In this paper, we proposed the deep ensemble model to detect and classify the DR grade over the publically available DRIVE dataset. The five deep learning classifiers such as InceptionResNetV2, EfficientNetV2L, Xception, DenseNet201, and VGG19 are used in the ensemble for the optimal topologies for each DR grade for classification. The ensemble model's success in DR detection and classification is a result of its ability to harness the diversity of deep learning architectures, robustly combine their predictions, leverage transfer learning from pre-trained models, adapt to DR-specific features through fine-tuning, and systematically measure its performance using appropriate metrics. These qualities collectively contribute to the ensemble's effectiveness in accurately identifying and classifying diabetic retinopathy. The final ensemble model result shows that each model successfully trained and passes the test with a better accuracy score of 98.65%. The evaluation parameters of this study, the training and testing validation of each parameter are obtained, and a confusion matrix and class-wise classification report are presented in Table 4. The final outcome shows that a significant amount of DR image dataset for the proposed ensemble model improves the learning time and test success, and better-predicted results for classification tasks using the DR image dataset. The main significance of this study is to improve early diagnosis, personalize treatment, reduce healthcare costs, enhance access to care, and ultimately preserve the vision and quality of life for individuals with diabetes. It is an important application of artificial intelligence and deep learning in healthcare that addresses a pressing global health issue. For future enhancement in DR detection and classification, we investigate other pretrained deep learning models like MobileNet, AlexNet, DenseNet169 etc. and also investigate other large DR datasets, may lead to good outcomes. These improvements can lead to higher accuracy, better generalization to diverse patient populations, increased diagnostic precision, and ultimately, improved patient outcomes in the diagnosis and management of diabetic retinopathy.

REFERENCES

[1] Fraser, C.E., D'Amico, D.J., Nathan, D.M., Trobe, J., Mulder, J.E. (2017). Diabetic retinopathy: Classification and clinical features. UpToDate. UpToDate, Waltham, MA.

[2] Griffin, S.J., Rutten, G.E., Khunti, K., Witte, D.R., Lauritzen, T., Sharp, S.J., Dalsgaard, E., Davies, M.J.,

Irving, G.J., Vos, R.C., Webb, D.R., Wareham, N.J., Sandbæk, A. (2019). Long-term effects of intensive multifactorial therapy in individuals with screen-detected type 2 diabetes in primary care: 10-year follow-up of the ADDITION-Europe cluster-randomised trial. The lancet Diabetes & Endocrinology, 7(12): 925-937. https://doi.org/10.1016/S2213-8587(19)30349-3

[3] Shakeel, P.M., Baskar, S., Sampath, R., Jaber, M.M. (2019). Echocardiography image segmentation using feed forward artificial neural network (FFANN) with fuzzy multi-scale edge detection (FMED). International Journal of Signal and Imaging Systems Engineering, 11(5): 270-278. https://doi.org/10.1504/IJSISE.2019.100651

[4] Smith-Morris, C., Bresnick, G.H., Cuadros, J., Bouskill, K.E., Pedersen, E.R. (2020). Diabetic retinopathy and the cascade into vision loss. Medical Anthropology, 39(2): 109-122. https://doi.org/10.1080/01459740.2018.1425839

[5] Jinfeng, G., Qummar, S., Junming, Z., Ruxian, Y., Khan, F.G. (2020). Ensemble framework of deep CNNs for diabetic retinopathy detection. Computational Intelligence and Neuroscience, 2020: 1-11. https://doi.org/10.1155/2020/8864698

[6] Sugeno, A., Ishikawa, Y., Ohshima, T., Muramatsu, R. (2021). Simple methods for the lesion detection and severity grading of diabetic retinopathy by image processing and transfer learning. Computers in Biology and Medicine, 137: 104795. https://doi.org/10.1016/j.compbiomed.2021.104795

[7] Mondal, S.S., Mandal, N., Singh, K.K., Singh, A., Izonin, I. (2022). Edldr: An ensemble deep learning technique for detection and classification of diabetic retinopathy. Diagnostics, 13(1): 124. https://doi.org/10.3390/diagnostics13010124

[8] Shanthi, T., Sabeenian, R.S. (2019). Modified Alexnet architecture for classification of diabetic retinopathy images. Computers & Electrical Engineering, 76: 56-64. https://doi.org/10.1016/j.compeleceng.2019.03.004

[9] Qomariah, D.U.N., Tjandrasa, H., Fatichah, C. (2019). Classification of diabetic retinopathy and normal retinal images using CNN and SVM. In 2019 12th International Conference on Information & Communication Technology and System (ICTS), Surabaya, Indonesia, pp. 152-157. https://doi.org/10.1109/ICTS.2019.8850940

[10] Li, F., Liu, Z., Chen, H., Jiang, M., Zhang, X., Wu, Z. (2019). Automatic detection of diabetic retinopathy in retinal fundus photographs based on deep learning algorithm. Translational Vision Science & Technology, 8(6): 4. https://doi.org/10.1167/tvst.8.6.4

[11] Nielsen, K.B., Lautrup, M. L., Andersen, J.K., Savarimuthu, T.R., Grauslund, J. (2019). Deep learning–based algorithms in screening of diabetic retinopathy: A systematic review of diagnostic performance. Ophthalmology Retina, 3(4): 294-304. https://doi.org/10.1016/j.oret.2018.10.014

[12] Dhoot, D.S., Baker, K., Saroj, N., Vitti, R., Berliner, A.J., Metzig, C., Thompson, D., Singh, R.P. (2018). Baseline factors affecting changes in diabetic retinopathy severity scale score after intravitreal aflibercept or laser for diabetic macular edema: Post Hoc analyses from VISTA and VIVID. Ophthalmology, 125(1): 51-56. https://doi.org/10.1016/j.ophtha.2017.06.029

[13] Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P.C., Mega, J.C., Webster, D.R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA, 316(22): 2402-2410. https://doi.org/10.1001/jama.2016.17216

[14] Gao, J.F., Sehrish, Q., Zhang, J.M., Yao, R.X., Fiaz, G.K. (2020). Ensemble framework of deep CNNs for diabetic retinopathy detection. Computational Intelligence and Neuroscience, 8864698. https://doi.org/10.1155/2020/8864698

[15] Ai, Z, Huang, X., Fan, Y., Feng, J., Zeng F.X., Lu Y.P. (2021) DR-IIXRN: Detection algorithm of diabetic retinopathy based on deep ensemble learning and attention mechanism. Front. Neuroinform. 15: 778552. https://doi.org/10.3389/fninf.2021.778552

[16] Pak, A., Ziyaden, A., Tukeshev, K., Jaxylykova, A., Abdullina, D. (2020) Comparative analysis of deep learning methods of detection of diabetic retinopathy. Cogent Engineering, 7(1): 1805144. https://doi.org/10.1080/23311916.2020.1805144

[17] Johnson, S., JR, L. J., Karthikeyan, G., Vengadapathiraj, M., Sasireka, D. (2022). An ensemble deep learning approach for diabetic retinopathy detection using fundus image. In 2022 6th International Conference on Electronics, Communication and Aerospace Technology, 829-836. https://doi.org/10.1109/ICECA55336.2022.10009304

[18] Asiri, N., Hussain, M., Al Adel, F., Alzaidi, N. (2019). Deep learning based computer-aided diagnosis systems for diabetic retinopathy: A survey. Artificial Intelligence in Medicine, 99: 101701. https://doi.org/10.1016/j.artmed.2019.07.009

[19] Hosseinzadeh K. S., Wesolowski, M. J., Schneider, K. A., Deters, R. (2019). A hybrid deep learning architecture for leukemic B-lymphoblast classification. ARXIV E-PRINTS, ARXIV-1909, pp. 271-276. https://doi.org/10.1109/ICTC46691.2019.8939959

[20] Ishtiaq, U., Abdul Kareem, S., Abdullah, E.R.M.F., Mujtaba, G., Jahangir, R., Ghafoor, H.Y. (2020). Diabetic retinopathy detection through artificial intelligent techniques: A review and open issues. Multimedia Tools and Applications, 79: 15209-15252. https://doi.org/10.1007/s11042-018-7044-8

[21] Alyoubi, W.L., Shalash, W.M., Abulkhair, M.F. (2020). Diabetic retinopathy detection through deep learning techniques: A review. Informatics in Medicine Unlocked, 20: 100377. https://doi.org/10.1016/j.imu.2020.100377

[22] Stolte, S., Fang, R. (2020). A survey on medical image analysis in diabetic retinopathy. Medical Image Analysis, 64: 101742. https://doi.org/10.1016/j.media.2020.101742

[23] Badar, M., Haris, M., Fatima, A. (2020). Application of deep learning for retinal image analysis: A review. Computer Science Review, 35: 100203. https://doi.org/10.1016/j.cosrev.2019.100203

[24] Chu, A., Squirrell, D., Phillips, A.M., Vaghefi, E. (2020). Essentials of a robust deep learning system for diabetic retinopathy screening: A systematic literature review. Journal of Ophthalmology, 2020: 1-11. https://doi.org/10.1155/2020/8841927

[25] Sarki, R., Ahmed, K., Wang, H., Zhang, Y. (2020). Automatic detection of diabetic eye disease through deep

learning using fundus images: A survey. IEEE Access, 8: 151133-151149. https://doi.org/10.1109/ACCESS.2020.3015258

[26] Bhatia, Y., Bajpayee, A., Raghuvanshi, D., Mittal, H. (2019). Image captioning using Google's inception-resnet-v2 and recurrent neural network. In 2019 Twelfth International Conference on Contemporary Computing (IC3), Noida, India, pp. 1-6. https://doi.org/10.1109/IC3.2019.8844921

[27] Villa-Pulgarin, J.P., Ruales-Torres, A.A., Arias-Garzon, D., Bravo-Ortiz, M.A., Arteaga-Arteaga, H.B., Mora-Rubio, A., Alzate-Grisales, J.A., Mercado-Ruiz, E., Hassaballah, M., Orozco-Arias, S., Cardona-Morales, O., Tabares-Soto, R. (2022). Optimized convolutional neural network models for skin lesion classification. Comput. Mater. Contin, 70(2): 2131-2148. https://doi.org/10.32604/cmc.2022.019529

[28] Ulutaş, H., Aslantaş, V. (2023). Design of efficient methods for the detection of tomato leaf disease utilizing proposed ensemble CNN model. Electronics, 12(4), 827. https://doi.org/10.3390/electronics12040827

[29] Tan, M., Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In International Conference on Machine Learning, pp. 6105-6114.

[30] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1251-1258.

[31] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778.

[32] Kobat, S.G., Baygin, N., Yusufoglu, E., Baygin, M., Barua, P.D., Dogan, S., Yaman, O., Celiker, U., Yildirim, H., Tan, R., Tuncer, T., Islam, N., Acharya, U.R. (2022). Automated diabetic retinopathy detection using horizontal and vertical patch division-based pre-trained DenseNET with digital fundus images. Diagnostics, 12(8): 1975. https://doi.org/10.3390/diagnostics12081975

[33] Das, D., Biswas, S.K., Bandyopadhyay, S. (2023). Detection of diabetic retinopathy using convolutional neural networks for feature extraction and classification (DRFEC). Multimedia Tools and Applications, 82(19): 29943-30001. https://doi.org/10.1007/s11042-022-14165-4

[34] Sudha, V., Ganeshbabu, T.R. (2021). A convolutional neural network classifier VGG-19 architecture for lesion detection and grading in diabetic retinopathy based on deep learning. Computers, Materials & Continua, 66(1): 827-842. https://doi.org/10.32604/cmc.2020.012008

[35] Kumar, G., Chatterjee, S., Chattopadhyay, C. (2021). DRISTI: A hybrid deep neural network for diabetic retinopathy diagnosis. Signal, Image and Video Processing, 15(8): 1679-1686. https://doi.org/10.1007/s11760-021-01904-7

[36] Deshmukh, S.V., Roy, A. (2023). Retinal Blood Vessel Segmentation Based on Modified CNN and Analyze the Perceptional Quality of Segmented Images. In: Woungang, I., Dhurandher, S.K., Pattanaik, K.K., Verma, A., Verma, P. (eds) Advanced Network Technologies and Intelligent Computing. ANTIC 2022. Communications in Computer and Information Science, vol 1798. Springer, Cham. https://doi.org/10.1007/978-3-031-28183-9_43

[37] Deshmukh, S.V., Roy, A., Agrawal, P. (2023). Retinal image segmentation for diabetic retinopathy detection using U-Net architecture. International Journal of Image, Graphics and Signal Processing, 13(1): 79-92. https://doi.org/10.5815/ijigsp.2023.01.07

[38] Khare, V. (2023). Automatic detection of exudates in retinal image using statistical techniques. Ingénierie des Systèmes d'Information, 28(1): 255-261. https://doi.org/10.18280/isi.280130

[39] Dondeti, V., Bodapati, J.D., Shareef, S.N., Naralasetti, V. (2020). Deep convolution features in non-linear embedding space for fundus image classification. Revue d'Intelligence Artificielle, 34(3): 307-313. https://doi.org/10.18280/ria.340308

NOMENCLATURE

DR	Diabetic retinopathy
NPDR	Non-Proliferative Diabetic Retinopathy
PDR	Proliferative Diabetic Retinopathy
CNN	Convolutional Neural Networks
DNN	Deep Neural Network
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative