Journal of Software Engineering and Applications, 2025, 18(1), 1-23
https://www.scirp.org/journal/jsea
ISSN Online: 1945-3124
ISSN Print: 1945-3116

Diabetic Retinopathy Severity Classification Using Data Fusion and Ensemble Transfer Learning

Shabib Aftab, Samia Akhtar

Department of Computer Science, Virtual University of Pakistan, Lahore, Pakistan
Email: shabib.aftab@vu.edu.pk, samiaakhtar9898@gmail.com

How to cite this paper: Aftab, S. and Akhtar, S. (2025) Diabetic Retinopathy Severity Classification Using Data Fusion and Ensemble Transfer Learning. *Journal of Software Engineering and Applications*, **18**, 1-23.
https://doi.org/10.4236/jsea.2025.181001

Received: November 26, 2024
Accepted: January 3, 2025
Published: January 6, 2025

Abstract

Diabetic retinopathy is a serious concern for people dealing with diabetes. Detecting diabetic retinopathy poses significant challenges, requiring skilled professionals, extensive manual image processing, and considerable time investment. Fortunately, the integration of deep learning and transfer learning offers invaluable assistance to medical practitioners. This study introduces an ensemble classification framework to detect and grade diabetic retinopathy into 5 classes leveraging the concepts of transfer learning and data fusion. It utilizes three benchmark datasets on diabetic retinopathy: APTOS 2019, IDRiD, and Messidor-2. Initially, these datasets are merged, resulting in a total of 5922 fundus images. Then this fused dataset undergoes pre-processing. Firstly, the images are cropped to remove unwanted regions. Then, Contrast Limited Adaptive Histogram Equalization is applied to improve image quality and fine details. To tackle class imbalance issues, Synthetic Minority Over Sampling technique is employed. Additionally, data augmentation techniques such as flipping, rotation, and zooming are used to increase dataset diversity. The dataset is split into training, validation, and testing sets at a ratio of 70:10:20. For classification, three pre-trained CNN models, EfficientNetB2, DenseNet121, and ResNet50, are fine-tuned. After these models are trained, an ensemble model is constructed by averaging the predictions of each model. Results show that the ensemble model achieved the highest test accuracy of 96.96% in grading diabetic retinopathy into 5 classes outperforming the individual pre-trained models. Furthermore, the ensemble model's performance is compared with previously published approaches where this model demonstrated superior result.

Keywords

Diabetic Retinopathy, Fundus Images, CLAHE, Augmentation, Optimizer

1. Introduction

Diabetes is a chronic health condition which is characterized by high blood sugar levels and can lead to various complications affecting multiple organs in the body when it is left unmanaged [1]. One such complication is diabetic retinopathy (DR), a condition that affects the eyes and is a leading cause of blindness worldwide [2]. DR occurs when prolonged exposure to high blood sugar levels damages the blood vessels in the retina, the light-sensitive tissue at the back of the eye [3]. As these blood vessels weaken and leak fluid or blood, they can cause vision impairment or even total vision loss if not treated on time [4].

The increasing prevalence of diabetes raises concerns about diabetic retinopathy worldwide [5]. According to the International Diabetes Federation (IDF), the number of adults with diabetes is expected to reach 700 million by 2045 [6]. It is estimated that around 27.0% of individuals with diabetes develop diabetic retinopathy globally leading to approximately 0.4 million cases of blindness worldwide [7]. Several factors contribute to the development and progression of diabetic retinopathy. Prolonged exposure to high blood sugar levels is the primary cause, leading to damage to the blood vessels in the retina [8]. Additionally, factors such as hypertension, high cholesterol levels, smoking, and genetics can increase the risk of developing DR.

1.1. Symptoms and Stages of DR

Symptoms of diabetic retinopathy may not be noticeable in the early stages [9], which is why regular eye examinations are essential for early detection and intervention. As the condition progresses, symptoms may include micro aneurysms (small bulges in the blood vessels that leak fluid), hemorrhages (spots or blotches as a result of bleeding from damaged blood vessels), exudates (yellow deposits in the retina as a result of leaking blood vessels), cotton wool spots (areas of swelling in the retina), and neovascularization (abnormal growth of new blood vessels on the surface of the retina). Distinctive features present in fundus images aid medical professionals and physicians in the diagnosis of diabetic retinopathy [10]. Timely detection and intervention plays a pivotal role in preventing the progression of this disease and preserving vision [11]. Regular eye examinations, including dilated eye exams, are crucial for early detection of diabetic retinopathy. These routine examinations are conducted by eye care professionals, such as ophthalmologists or optometrists, and involve a comprehensive evaluation of the retina to detect any signs of diabetic retinopathy. Early intervention, through lifestyle changes, medication, or laser therapy, can help slow or prevent vision loss [12].

Diabetic retinopathy (DR) is divided into two main stages: proliferative and non-proliferative [13]. Proliferative diabetic retinopathy (PDR) is identified by the abnormal growth of blood vessels on the retina, posing a risk of significant vision loss if not addressed. Non-proliferative diabetic retinopathy (NPDR) encompasses initial damage to retinal blood vessels, such as the formation of micro aneurysms, hemorrhages, and lipid deposits. NPDR is further categorized into mild, moderate, and severe stages.

1.2. Role of AI in Diabetic Retinopathy Detection

Advances in imaging technology and artificial intelligence (AI) have enabled more accurate and efficient screening for diabetic retinopathy [14]. Deep learning and transfer learning are increasingly being utilized in the detection and diagnosis of diabetic retinopathy. Deep learning (DL) is a subset of AI that involves training artificial neural networks with large datasets to recognize patterns and make predictions. Transfer learning (TL) is a technique in machine learning where a model trained on one task is adapted or fine-tuned for a related task [15]. In the context of diabetic retinopathy detection, AI algorithms analyze retinal images to identify signs of retinopathy, such as micro aneurysms, hemorrhages, or exudates. Deep learning models can accurately detect and classify retinopathy stages, aiding clinicians in early diagnosis and treatment planning. Transfer learning further enhances the performance of these models by leveraging pre-trained networks and adapting them to specific retinal imaging datasets.

Retinal examinations involve utilizing specialized equipment like fundus cameras to capture detailed images of the retina [16]. During the procedure, the patient's pupils are dilated to ensure a clear view, and the captured images are then evaluated by either ophthalmologists or AI algorithms to detect signs of diabetic retinopathy. Previously, the detection and management of DR relied on manual image analysis methods, which were time-consuming and prone to errors due to subjective interpretation. However, the emergence of deep learning algorithms has revolutionized this process. Deep learning techniques allow for automated analysis of retinal images, enabling the identification of specific features associated with DR [17]. By training deep neural networks on extensive datasets of fundus images, these algorithms can accurately categorize images based on DR severity and predict the risk of disease progression. These advancements not only improve the precision and consistency of DR detection but also decrease the dependency on human involvement, thus streamlining diagnostic procedures [18].

1.3. Research Contributions

This research proposes an ensemble framework for multi-classification of diabetic retinopathy using transfer learning and data fusion. It makes use of three distinctive benchmark datasets. These datasets have fundus images based on severity levels of diabetic retinopathy: 1) DR, 2) Mild DR, 3) Moderate DR, 4) Severe DR and 5) Proliferative DR. Many pre-processing techniques are applied to prepare the dataset for more accurate model training. Data Augmentation is applied to increase the size of dataset. Then a Convolutional Neural Network Model is developed from scratch for classification of DR. The main contributions of this research are:

1) Three distinct DR Datasets are fused in this study. Due to the differences in the datasets, it is challenging to achieve high accuracy on merged diabetic

retinopathy dataset. Images of different resolutions, lighting, and quality levels are included in individual datasets. Various pre-processing methods are employed to address these problems in this research in order to attain a higher accuracy.

2) Contrast Limited Adaptive Histogram Equalization (CLAHE) is utilized to enhance the quality of dull and faded images within the dataset. This technique improved features such as tiny blood vessels and overall image clarity.

3) The fused dataset exhibits an imbalanced class distribution. To counteract this imbalance and prevent bias towards the majority class, Synthetic Minority Over Sampling Technique (SMOTE) is applied. This technique helps balance the representation of different classes in the dataset, thus promoting fair model training and evaluation.

4) Pre-trained CNN models are fine-tuned and trained for classification of diabetic retinopathy. An ensemble model is built by averaging the predictions of these pre-trained models. This model was able to classify images with a high accuracy and performed well when compared with individual pre-trained models and previously published DL techniques.

1.4. Research Organization

This research is structured as follows: Section 2 presents an overview of previous literature regarding the detection and grading of diabetic retinopathy (DR), including various techniques and classifiers utilized. Section 3 offers a detailed explanation of the proposed classification framework. The outcomes of the suggested methodology are analyzed in Section 4, alongside a comparison with previous research using performance evaluation metrics. Finally, Section 5 concludes the study and provides a brief insight into future work.

2. Literature Review

A number of researchers have dedicated their effort in trying a range of methodologies, algorithms, and approaches to identify and classify diabetic retinopathy. Their dedication has resulted in notable advancements in the identification and classification of this disease. A detailed review of modern DR methods is given in this section.

In [19], researchers have introduced a multi-classification system for grading diabetic retinopathy using deep learning techniques. They incorporated semi-supervised learning and a wrapper algorithm in their approach. Initially, the model learns patterns and features from images through semi-supervised learning, followed by transferring this pre-trained model to the target model. Several experiments were conducted, demonstrating that their proposed wrapper algorithm-based technique outperformed previous methods, with an accuracy improvement of approximately 4 to 5%. The study utilized three datasets: EyePACS, DDR, and Messidor-2, all of which underwent preprocessing and cleaning. Images were resized and normalized before classification. The proposed model achieved an accuracy of 86.40% on the EyePACS dataset, 89.62% on DDR, and 90.15% on the

Messidor-2 dataset. Authors of [20] have proposed a binary classification system for distinguishing diabetic retinopathy as either DR or No DR. They utilized Convolutional Neural Network (CNN) based transfer learning concepts for classification purposes. Two datasets are employed: DR1, which comprises 1014 DR images, and Messidor, consisting of 1200 fundus images. Image preprocessing involves cropping followed by resizing. Four pre-trained models are fine-tuned for the task: VggNet-vd-16, Vgg-s, AlexNet, and VggNet-vd-19. Three distinct experiments are conducted: fine-tuning all layers, fine-tuning in a layer-wise manner, and extracting features using CNNs then classifying using SVM. Their proposed framework achieved the highest accuracy of 94.52% on the DR1 dataset and 92.01% on Messidor.

Researchers of [21] have proposed a computer-aided diagnostic (CAD) system for identifying non-proliferative diabetic retinopathy using convolutional neural networks (CNNs), specifically designed for the optical coherence tomography (OCT) imaging modality. They trained the model on an OCT dataset and employed several crucial procedures during preprocessing to extract input retina patches for training the CNN, without resizing the image. Additionally, they utilized transfer learning principles and effective feature combination techniques to enhance performance. The AlexNet CNN was utilized with an input size of 227 × 227. By combining the output features extracted from two independently trained CNNs, the methodology achieved the highest accuracy with the least computational complexity, reaching 94% accuracy, 100% recall, and 88% specificity. In [22], the authors proposed a binary classification framework for detecting diabetic retinopathy using transfer learning. They integrate Pyspark with deep learning, leveraging it as a Big Data tool to process the IDRiD (Indian Diabetic Retinopathy image Dataset) used in the research. As a preprocessing step, the dataset is initially cleaned by removing dark images, followed by cropping to eliminate unwanted spaces and resizing. The classification is executed using a LR classifier, employing DL Pipelines on Apache Spark. The dataset is split into an 80% training set and a 20% testing set. Results reveal that among the three transfer learning models utilized, InceptionV3 demonstrated the best performance, achieving an accuracy of 95%, an AUC of 94.98%, and an F1-Score of 95%.

A binary classification framework is proposed in [23] for grading diabetic retinopathy using the ResNet-50 model. Authors employ transfer learning and modify the ResNet50 model to enhance the accuracy of DR prediction. The dataset utilized is a DR Dataset from Kaggle, containing 35,126 retinal images. Images undergo cropping and removal of black borders, followed by Histogram Equalization. The ResNet-50 model is then utilized for classification, with adaptive learning rate settings and regularization to mitigate overfitting during training. The performance of ResNet-50 is compared with other transfer learning models such as AlexNet, Xception, VggNet-16, etc., and the results indicate that the revised ResNet-50 outperforms the others, achieving an accuracy of 74.32%. In [24], the authors deployed three distinct hybrid models for the classification of DR into five classes, namely Hybrid-a, Hybrid-f, and Hybrid-c. These hybrid models are constructed by integrating five base CNN models: NASNetLarge, EfficientNetB5, EfficientNetB4, InceptionReNetV2, and Xception. Two loss functions are utilized to train the base models, and the outputs of these base models are subsequently employed to train the hybrid models. The researchers conduct experiments with three different datasets: APTOS, EyePACS, and DeepDR. Pre-processing is conducted both before model training and during the training process, primarily involving image enhancement techniques. Among all the hybrid models, the Hybrid-c model achieved the highest accuracy of 86.34%.

Researchers of [25] have proposed a multi-classification method for grading diabetic retinopathy using transfer learning models, namely EfficientNet, ResNet, and VGG. The researchers utilize a Kaggle competition dataset comprising 35,126 fundus images. This dataset is resized and divided into training, validation, and testing sets prior to classification. Various training parameters, types of transfer learning models, and performance metrics are experimented with. Among all the models employed, Efficient_Net_b3_60 achieved the highest accuracy of 87%, followed by a quadratic weighted kappa (QWK) of 0.85, an F1-score of 84%, precision of 85%, and recall of 87%. In [26], researchers have presented a multi-classification framework aimed at detecting and grading diabetic retinopathy into five classes through Transfer Learning. They have employed the Inception-V3 model, pre-trained with ImageNet data. The dataset utilized in their study is EyePACS sourced from Kaggle. Pre-processing of fundus images involves down-sampling, local average color subtraction, and cropping of image borders. Classification is conducted using Inception-V3 after pre-processing. Their proposed model achieved a validation accuracy of 50% and a final test accuracy of 48.2%.

All these previous studies have applied deep learning techniques to detect and classify diabetic retinopathy using various pre-processing strategies and classification models, such as CNNs, hybrid models, and transfer learning. Despite these efforts, many of these approaches still struggle with issues like accuracy, generalization, and class imbalance. In contrast, our research offers a more efficient and robust solution. While many prior methods relied on using individual datasets, we combined three distinct diabetic retinopathy datasets, improving our model's ability to generalize across different types of data. We also implemented advanced preprocessing techniques, such as CLAHE, to enhance image quality and generalization. To tackle class imbalance, we applied SMOTE which helped in reducing bias in classification results. Our framework leverages transfer learning from well-established architectures ensuring faster convergence and more accurate feature extraction. A key innovation in our work is the use of an ensemble method, combining three pre-trained models which significantly boosted both accuracy and efficiency in detecting diabetic retinopathy.

Table 1 summarizes the key aspects of the studies reviewed in this section. It provides an overview of the datasets used, preprocessing techniques, methodologies employed, and the evaluation metrics reported for each study.

Table 1. Summary of the reviewed literature.

Study	Datasets Used	Preprocessing Techniques	Methodology	Evaluation Metrics
[19]	EyePACS, DDR, Messidor-2	Cleaning, cropping, resizing, normalization	Wrapper algorithm based Deep Learning Model	Accuracy, True Positive Rate, True Negative Rate
[20]	DR1, Messidor	Cropping, resizing	VggNet-vd-16, Vgg-s, AlexNet, VggNet-vd-19	Accuracy, Sensitivity, Specificity, AUC
[21]	OCT dataset	Retina segmentation, Patch extraction, Patch Alignment	Transfer Learning based CAD System	Accuracy, Error rate, Precision, Recall
[22]	IDRiD	Cropping, resizing, removal of dark images	3 Pre-trained Models + Logistic Regression	Accuracy, AUC, F1-Score
[23]	Kaggle DR Dataset	Cropping, Histogram Equalization	ResNet-50 (Transfer learning)	Accuracy
[24]	APTOS, EyePACS, DeepDR	Cropping, resizing, data augmentation	Hybrid Models	Accuracy, Sensitivity, Specificity, F1-Score, Precision
[25]	Kaggle DR Dataset	Resizing, dataset split	Pre-trained models (VGG, ResNet, EfficientNet etc.)	Accuracy, F1-Score, Precision, Recall, QWK
[26]	EyePACS	Down-sampling, color subtraction, cropping	Inception-V3	Accuracy

3. Methodology

This study introduces a classification framework aimed at categorizing diabetic retinopathy into five distinct classes. To achieve this classification objective, three benchmark datasets are fused together and an ensemble of three pre-trained CNN models is created. Transfer learning holds significant recognition and application in the fields of image processing and disease detection [27]. The ensemble classification framework proposed in this research is presented in Figure 1. Furthermore, the detail of each step involve in the framework is given below:

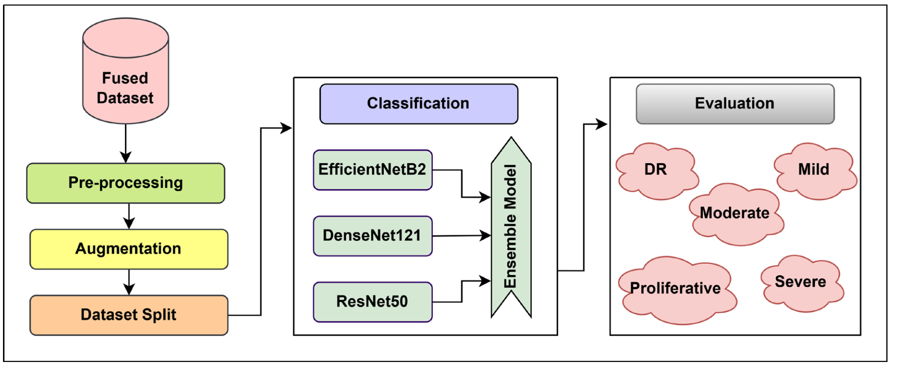


Figure 1. Proposed ensemble classification framework.

3.1. Dataset

This study utilizes three benchmark datasets for the multi-classification of diabetic retinopathy: 1) APTOS (Asia Pacific Tele-Ophthalmology Society) 2019 Blindness Detection, 2) IDRiD (Indian Diabetic Retinopathy Image Dataset), and 3) Messidor-2. The description of each dataset is given below:

3.1.1. APTOS 2019

The APTOS (Asia Pacific Tele-Ophthalmology Society) 2019 Blindness Detection is a benchmark dataset publically available at Kaggle [28]. It consists of 3662 retinal images sourced from the Aravind Eye Hospital in India. These images offer a diverse representation of retinal conditions and are captured under various clinical settings. The hospital's efforts are particularly noteworthy for offering diagnosis and treatment to individuals in rural areas where medical examinations pose challenges. Clinicians at the hospital have carefully graded the severity of diabetic retinopathy in each image on a scale ranging from 0 to 4. This grading system categorizes the disease into five distinct levels of severity: No DR, Mild, Moderate, Severe, and Proliferative DR. Figure 2 displays samples from each severity grade of this dataset.

3.1.2. IDRiD

IDRiD (Indian Diabetic Retinopathy Image Dataset) is the first dataset that represents the Indian population. It is benchmark DR dataset, publically available at IEEE Data Port [29]. The fact that it contains lesions associated with diabetic retinopathy as well as normal retinal features gives it great significance and makes it a valuable addition to datasets. This dataset was also a key part of the "Diabetic Retinopathy: Segmentation and Grading Challenge". The dataset is segmented into three primary categories: Disease Grading, Localization, and Segmentation. Our research specifically targets disease grading, which entails assessing the severity of diabetic retinopathy. For this purpose, the dataset offers 413 images for training and 103 images for testing along with their corresponding labels, resulting in a total of 516 fundus images. Similar to APTOS dataset, IDRiD classifies images into five grades: No DR, Mild, Moderate, Severe, and Proliferative DR. A sample from each grade of this dataset is given in Figure 2.

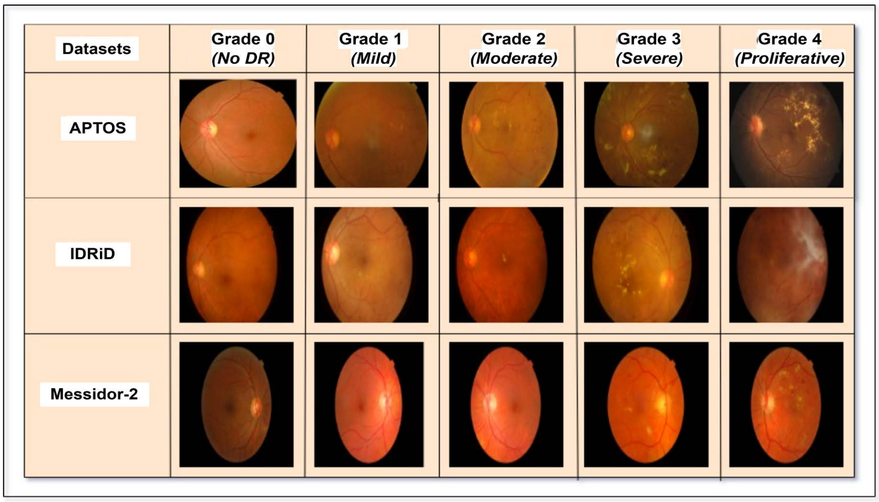


Figure 2. A sample from each grade of three DR datasets.

3.1.3. Messidor-2

Messidor-2 is also a benchmark dataset publically available at ADCIS [30]. It consists of diabetic retinopathy examinations, each containing two fundus images centered on the macula. It comprises a total of 1748 images, with 1744 images labeled by clinicians on a scale of 0 to 4. These images were captured using a fundus camera without pharmacological dilation. **Figure 2** shows a sample of each grade from this dataset.

All these three datasets are fused together resulting in a total of 5922 images. The individual datasets have unbalanced distribution of classes per grade/class. So fused dataset also has unbalanced distribution. **Table 2** presents the distribution of images per grade and the overall number of images for APTOS, IDRiD, Messidor-2 and the fused dataset.

Table 2. Image distribution per grade of individual and fused dataset.

Datasets	Grade					
	No DR (Grade 0)	Mild (Grade 1)	Moderate (Grade 2)	Severe (Grade 3)	Proliferative DR (Grade 4)	Total
APTOS	1805	370	999	193	295	**3662**
IDRiD	168	25	168	93	62	**516**
Messidor-2	1017	270	347	75	35	**1744**
Fused Dataset	2990	665	1514	361	392	**5922**

3.1.4. Benefits and Challenges of Data Fusion Approach

The main benefit of data fusion is that it helps the model generalize better. By exposing the model to more varied data, we can reduce the risk of over fitting to a single dataset [31]. Fusion also increases the overall dataset size and the number of images which provides the model with more examples to learn from [32]. In our case of Diabetic Retinopathy (DR) detection, fusing datasets like APTOS, IDRiD, and Messidor 2 allowed our models to learn from different retinal images. This enhanced their ability to detect DR across various populations and imaging conditions. However, data fusion also presented challenges. The datasets differed in image quality, resolution, and annotation standards which created inconsistencies and impacted performance. Additionally, these datasets have imbalanced data which could have introduced bias. These issues required our careful preprocessing and alignment of the datasets to ensure that the model benefited from the fusion without being negatively affected by these discrepancies.

3.2. Pre-Processing

Data preprocessing involves cleaning, scaling, and encoding data to prepare it for neural network training [33]. It ensures data quality by handling missing values, outliers, and noise. For this research, preprocessing is done in five steps: 1) Image Cropping, 2) Denoising, 3) Contrast Limited Adaptive Histogram Equalization

(CLAHE), 4) Image Resizing, and 5) Synthetic Minority Over-sampling Technique (SMOTE). Since three datasets are merged and each one is taken from a different source, a careful examination of the fused dataset is conducted. These datasets comprise fundus images from various clinics and hospitals, captured at different angles and with diverse cameras. The images vary in resolution, contrast, and background. Some images contain noise and unwanted black areas while others appear dull or lack contrast. This poses challenges for the identification of minute retinal structures such as micro aneurysms and small blood vessels.

3.2.1. Image Cropping

Cropping is the first step of pre-processing. The images in the dataset contain unwanted black background. There may be problems with performance and generalization if a model is given a dataset that represents a wide range of backgrounds [34]. The accuracy of the model will decrease as a result. To get rid of the extraneous background, the retinal images were cropped to the proper size starting from where the eye begins. By removing the undesired areas from the images, this step made sure that each image had the same background.

3.2.2. Denoising

For second step of pre-processing, the noise in the images was reduced using the technique called Image Denoising. Denoising in image preprocessing involves reducing or eliminating noise, which can obscure important features and degrade image quality. The method used in this research to denoise the images is Gaussian Blur. Images were denoised with a kernel size of (3, 3).

3.2.3. Contrast Limited Adaptive Histogram Equalization (CLAHE)

In third step, Contrast Limited Adaptive Histogram Equalization (CLAHE) was applied to the dataset. This technique helped in enhancing the contrast of images and bringing out finer details. This technique addressed the previously identified issues of dullness and low contrast within the dataset.

3.2.4. Image Resizing

Step four includes resizing the images to achieve uniform dimensions of 224 × 224 × 3 pixels. Here, the first two numbers represent the height and width of pixels, respectively, while the third denotes the presence of RGB (red, green, and blue) color channels in the image. As the fused dataset comprises images of varied resolutions and dimensions sourced from different origins, resizing them to a uniform resolution ensured homogeneity in the input data format. The first four preprocessing steps are shown in **Figure 3**.

3.2.5. Synthetic Minority Over-Sampling Technique (SMOTE)

In Step 5, which is the final preprocessing step, the Synthetic Minority Over-sampling Technique (SMOTE) was used to tackle the class imbalance present within the dataset. This technique involves generating synthetic samples for the minority classes to solve the issue of uneven distribution of samples among all the classes

[35]. By creating new instances of the minority classes based on existing ones, SMOTE aims to ensure a more balanced representation across all classes. As obvious from **Table 2**, the fused dataset exhibits an even distribution of images among classes that can bias the model's performance towards predicting the majority class. Therefore, addressing this imbalance is essential for achieving optimal model performance. The objective is to augment the sample size only in the minority classes while keeping the majority class as it is. In our dataset, Grade 0 contains the majority number of images and its count remains unchanged. Meanwhile, the samples in the other classes are increased to a randomized number close to the number of images in Grade 0. The updated image counts per class post-SMOTE implementation are detailed in **Table 3** with the new number of images in minority classes adjusted to 2400 in Grade 1, 2700 in Grade 2, 2600 in Grade 3, and 2400 in Grade 4.

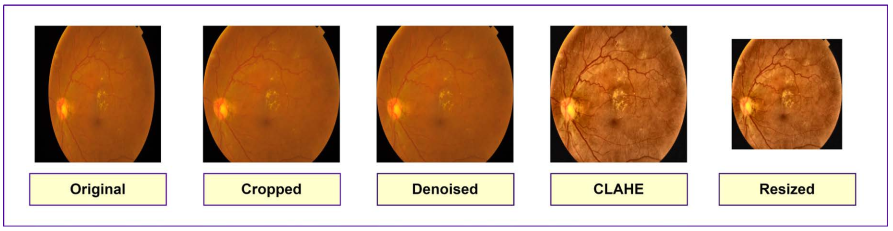


Figure 3. A sample of first four steps of pre-processing.

Table 3. New number of images per grade after applying SMOTE.

Grade	Original Image Count	Image Count after SMOTE
No DR (Grade 0)	2990	2990
Mild (Grade 1)	665	2400
Moderate (Grade 2)	1514	2700
Severe (Grade 3)	361	2600
Proliferative DR (Grade 4)	392	2400
Total	**5922**	**13,090**

3.3. Data Augmentation

After preprocessing, data augmentation is applied on the images using various techniques. This augmentation process enables the model to capture a broader range of patterns and variations within the data, thereby enhancing its generalization capabilities. Data Augmentation means to artificially expand a dataset by applying various transformations to existing data samples [36]. These transformations include rotation, flipping, zooming, scaling, shearing, shifting and many other. The augmentation techniques employed in this research include zooming, rotation, horizontal and vertical flipping. These transformations are randomly applied to the dataset with a zooming range set at 0.2 and a rotation range of 20 degrees. In this research, augmentation process doubles the total number of images in the dataset. After application of SMOTE, the number of images increased to 13,090. Through augmentation, the total number of images became 26,180. Table 4 presents the augmentation techniques utilized in this research.

Table 4. Augmentation techniques used on the dataset.

Flip	Horizontal
Flip	Vertical
Zoom Range	0.2
Rotation Range	10

3.4. Dataset Split

Before proceeding to the model training stage, the augmented dataset undergoes partitioning into training, validation, and testing sets. The dataset is divided according to a ratio of 70:10:20, with 70% designated for the training set, 10% for the validation set, and 20% for the testing set. The training set is utilized to train the models, while the testing set evaluates the performance of these trained models. Meanwhile, the validation set serves the purpose of assessing the model's performance during the training process.

3.5. Proposed Ensemble Model

The ImageNet dataset has been greatly used to train deep CNN models due to its diverse collection of images. ImageNet is a large-scale dataset which contains millions of labeled images across thousands of different categories [37]. Many popular models that are trained using ImageNet include ResNet, Inception, VGG and MobileNet etc. For this study, three pre-trained CNN models are used: EfficientNetB2, DenseNet121 and ResNet50. Prior to training, these models are fine-tuned and modifications are made to both the first and last layers of the models. The input layer was adjusted to accommodate images of size 224 × 224 × 3. Furthermore, a Global Average Pooling (GAP) layer was added and then Output Layer was added where softmax activation function is employed as this research deals with multi-classification. Detailed fine-tuned architectures for these models are illustrated in Figure 4.

Ensemble learning is a powerful approach in deep learning where multiple models are combined to improve overall predictive performance [38]. This method combines a variety of models to reduce the impact of individual deficiencies and improve the overall accuracy. In the context of transfer learning, ensemble involves combining predictions from multiple pre-trained models to create a more robust and accurate predictor. There are multiple techniques through which ensemble learning can be achieved. Common techniques are bagging, boosting, stacking, averaging etc.

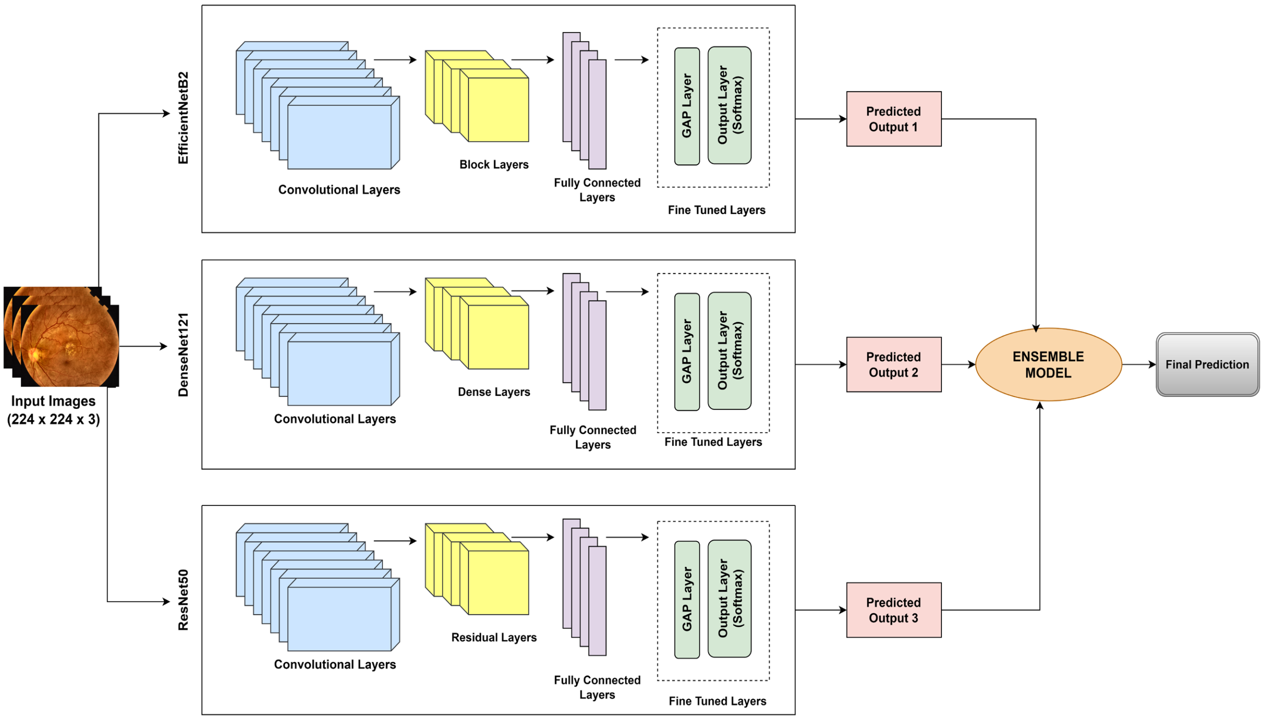


Figure 4. Structure of the used pre-trained models along with fine-tuned layers.

In this study, Ensemble Averaging is the chosen ensemble technique. This method combines predictions from multiple models by averaging them to generate a final prediction. Here, predictions from all three models, EfficientNetB2, DenseNet121 and ResNet50, are combined using averaging. The final prediction for each sample is then determined by selecting the class with the highest average probability across all classes. This process involves summing up the probabilities from each model for each class and dividing by the total number of models. Subsequently, the class with the highest average probability is chosen as the final prediction for each sample. This approach forms the basis of the ensemble model employed in this study.

3.6. Performance Evaluation Metrics

Performance metrics serve as quantitative measures utilized to evaluate the accuracy and efficacy of a model's predictions [39]. They offer valuable insights into the model's performance across different tasks and help researchers in assessing various models. In the evaluation of both the training and testing sets in this study, each data instance can be categorized into one of four outcomes: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). A True Positive indicates the correct prediction of a positive instance by the model, while a True Negative signifies the accurate prediction of a negative instance. Conversely, a False Positive suggests an incorrect prediction of a positive instance, and a False Negative denotes an erroneous prediction of a negative instance. The measures used in this research are as follows

$$\text{Misclassification Rate}(\text{MCR}) = \frac{\text{FP} + \text{FN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

$$\text{Accuracy}(\text{ACC}) = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{2}$$

$$\text{Specificity}(\text{SPC}) = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{3}$$

$$\text{Sensitivity}(\text{SEN}) = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

$$\text{F1-Score}(\text{F1}) = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

$$\text{Positive Prediction Value}(\text{PPV}) = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

$$\text{Negative Prediction Value}(\text{NPV}) = \frac{\text{TN}}{\text{TN} + \text{FN}} \tag{7}$$

$$\text{False Positive Ratio}(\text{FPR}) = 1 - \text{Specificity} \tag{8}$$

$$\text{False Negative Ratio}(\text{FNR}) = 1 - \text{Sensitivity} \tag{9}$$

$$\text{Likelihood Ratio Positive}(\text{LR}+) = \frac{\text{Sensitivity}}{(1 - \text{Specificity})} \tag{10}$$

$$\text{Likelihood Ratio Negative}(\text{LR}-) = \frac{(1 - \text{Sensitivity})}{\text{Specificity}} \tag{11}$$

4. Results and Discussion

Following preprocessing and augmentation, the merged dataset was divided into training, validation and testing sets at a ratio of 70:10:20. Post-augmentation, the dataset comprised a total of 26,180 fundus images. Thus, the distribution across each set is as follows: the training set contains 18,849 images (70%), while the validation and testing sets contain 2095 images (10%) and 5236 images (20%) respectively. For executing our models, we used Python programming language within Kaggle Notebooks. These notebooks offer resources such as a storage capacity of up to 73GB, 29GB of RAM, and a GPU with 15GB memory, capable of supporting heavy deep learning models. Various libraries were used including TensorFlow, OpenCV, Pandas, and Matplotlib.

This study proposed an ensemble model and 3 pre-trained models. Before the training process, the hyper parameters were adjusted. A batch size of 32 was chosen, and the model underwent 40 epochs of training. The Adam optimizer with a learning rate set at 0.001 was set for optimization. To avoid overfitting, regularization techniques such as EarlyStopping and ReduceLROnPlateau were integrated. These mechanisms are designed to halt the training process when the model reaches optimal performance to prevent overfitting. In this study, EarlyStopping monitors validation loss and ceases training if it stabilizes or improves for 10 consecutive epochs. Meanwhile, the ReduceLROnPlateau method adjusts the learning rate by a factor of 0.2 if the loss value remains constant for 2 successive epoch. Table 5 outlines the hyper parameters utilized in the study.

Table 5. Hyper-parameters used for training the models.

Hyper Parameters	Values
Optimizer	Adam
Learning Rate	0.001
Batch Size	32
No. of Epochs	40
Loss Function	Categorical cross entropy
EarlyStopping	patience = 10
ReduceLROnPlateau	factor = 0.2 patience = 2

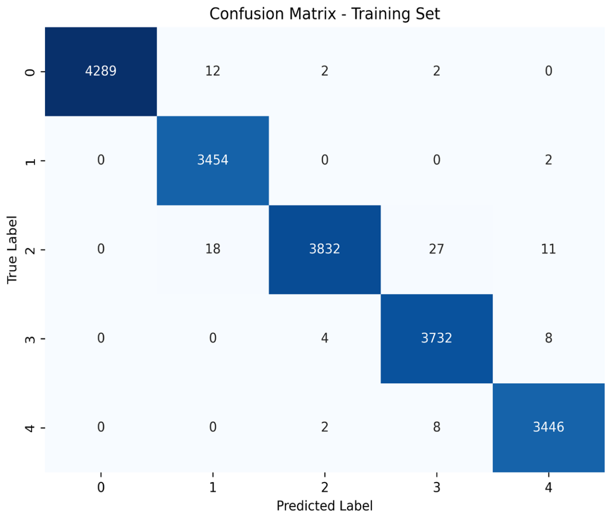
On training the individual models, EfficientNetB2 achieved a training accuracy of 99.49 % with a sensitivity and specificity of 99.50% and 99.87% respectively. When this trained model was evaluated on the test set, EfficientNetB2 gave a test accuracy of 96.27% followed by 96.27% sensitivity and 99.07% specificity.

When DenseNet121 model was trained, it gave a training accuracy of 97.35% with a sensitivity and specificity of 97.29% and 99.34% respectively. Then this trained DenseNet121 model was evaluated on test set. It achieved a test accuracy of 91.21%. Sensitivity and specificity achieved by this model on test set is 91.05% and 97.81% respectively.

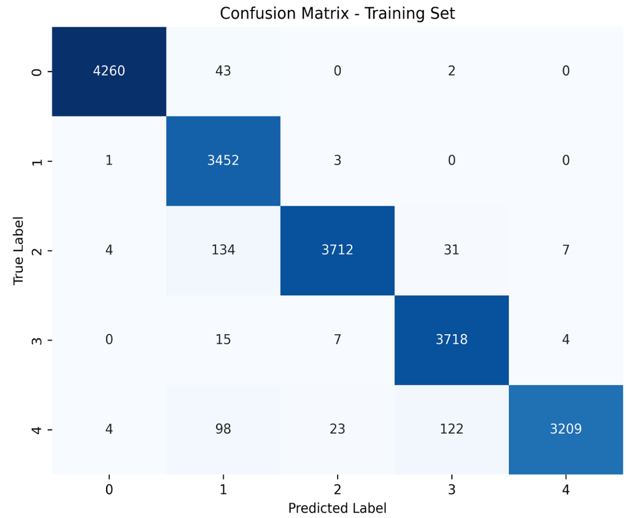
On training the third model that is ResNet50, it achieved a training accuracy of 99.42 % with a sensitivity and specificity of 99.41% and 99.85% respectively. When this trained model was evaluated on the test set, ResNet50 gave a test accuracy of 94.95% followed by 94.90% sensitivity and 98.74% specificity.

After combining the predictions of the above three pre-trained models, the ensemble model was created. It showed the best training accuracy of 99.69%, with 99.68% sensitivity and 99.92% specificity. On the testing set, this ensemble model achieved the highest accuracy of 96.96%, with a sensitivity of 96.93% and a specificity of 99.23%. Figure 5 provides the confusion matrix for both the training and testing sets of all the models.

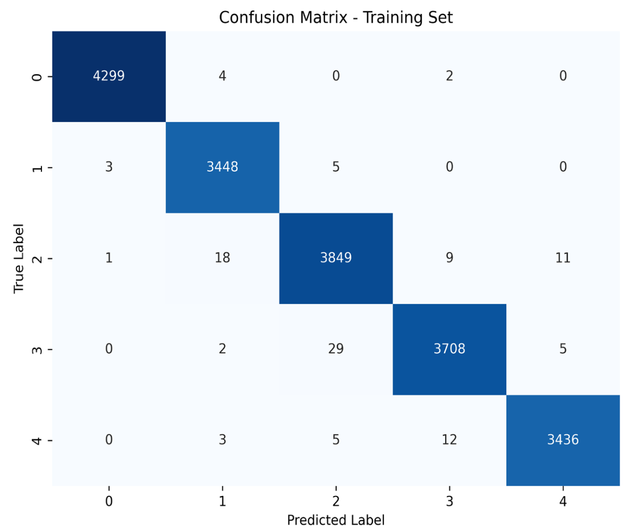
Figure 6 illustrates the training and validation accuracy + loss for all the models graphically. The graphs indicate successful training of these models, as evidenced by their low validation loss. EfficientNetB2 achieved a validation accuracy of 96.80% with a validation loss of 0.1607. DenseNet121 attained a validation accuracy of 91.79% with a validation loss of 0.3550. A validation accuracy of 95.08% was achieved by ResNet50 along a 0.2082 validation loss while the Ensemble model yielded a validation accuracy of 96.99% with a validation loss of 0.1583. These results are visually depicted in the graphs.



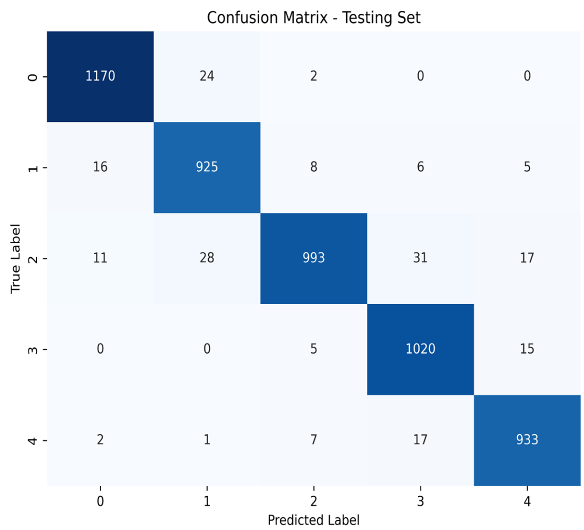
(a) Confusion matrix of training set for EfficientNetB2



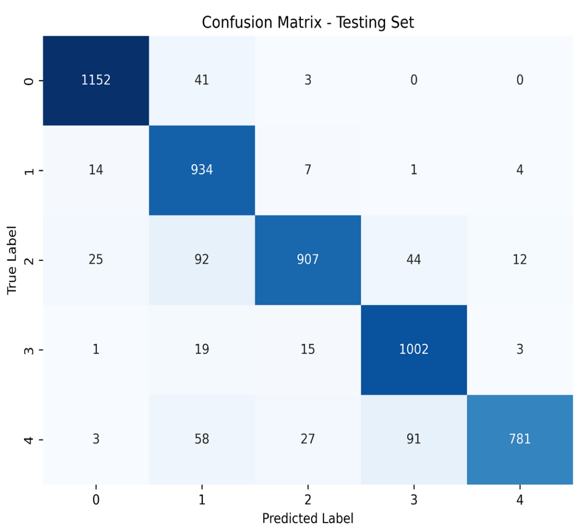
(b) Confusion matrix of training set for DenseNet121



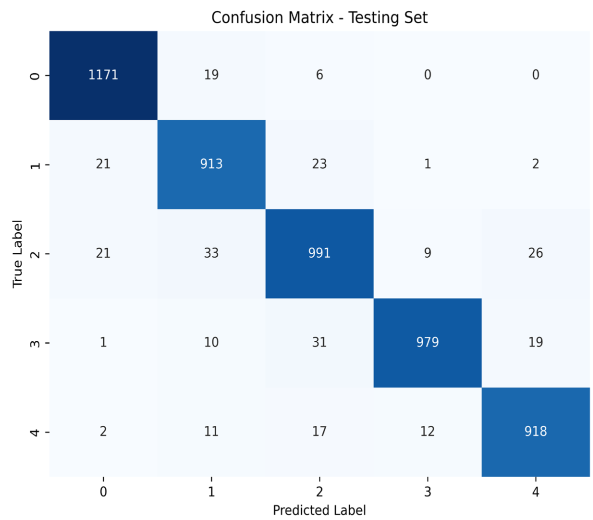
(c) Confusion matrix of training set for ResNet50



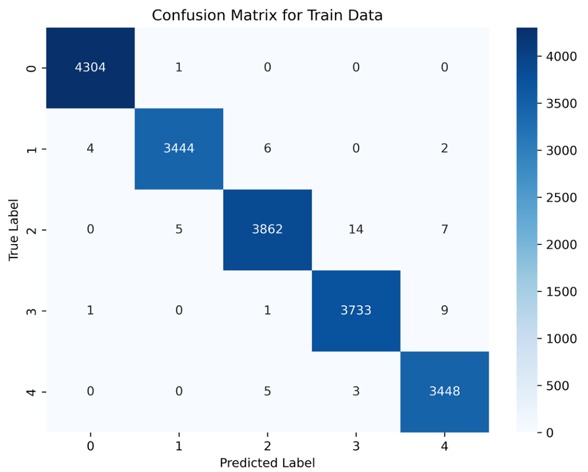
(d) Confusion matrix of testing set for EfficientNetB2



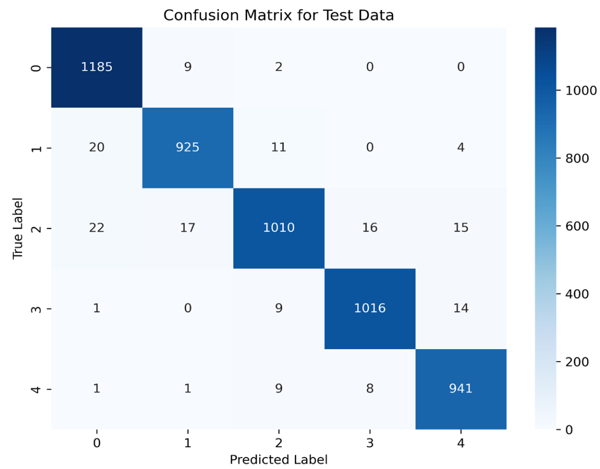
(e) Confusion matrix of testing set for DenseNet121



(f) Confusion matrix of testing set for ResNet50

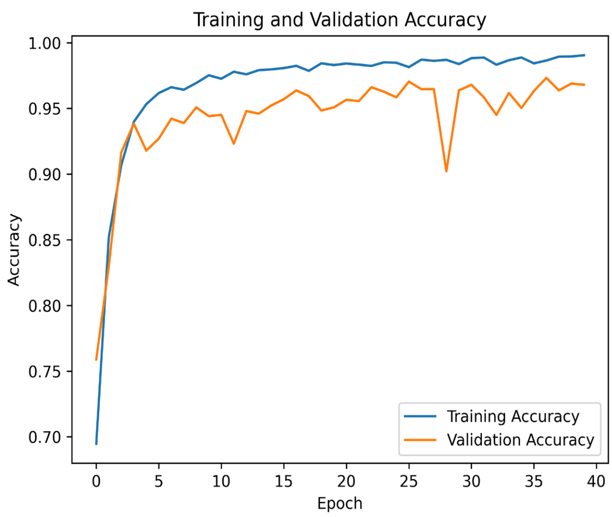


(g) Confusion matrix of training set for ensemble model

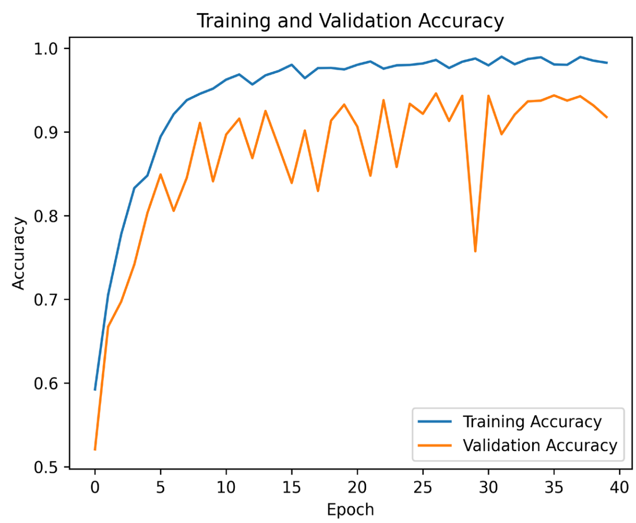


(h) Confusion matrix of testing set for ensemble model

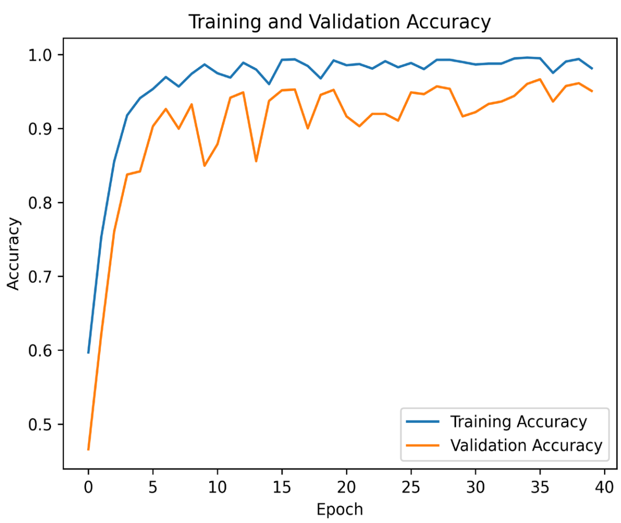
Figure 5. Confusion matrix of individual models and ensemble model.



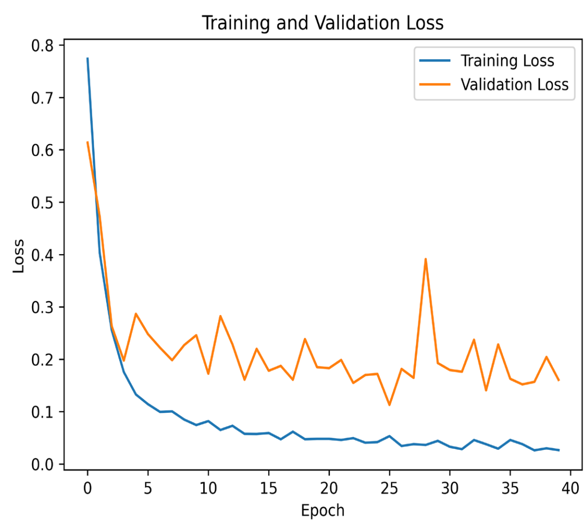
(a) Accuracy graph for EfficientNetB2



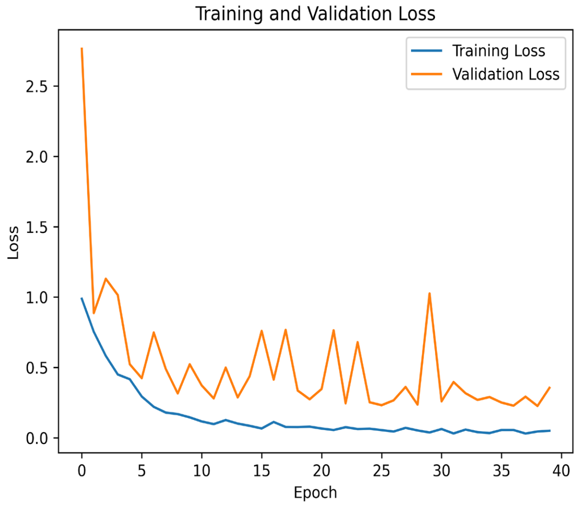
(b) Accuracy graph for DenseNet121



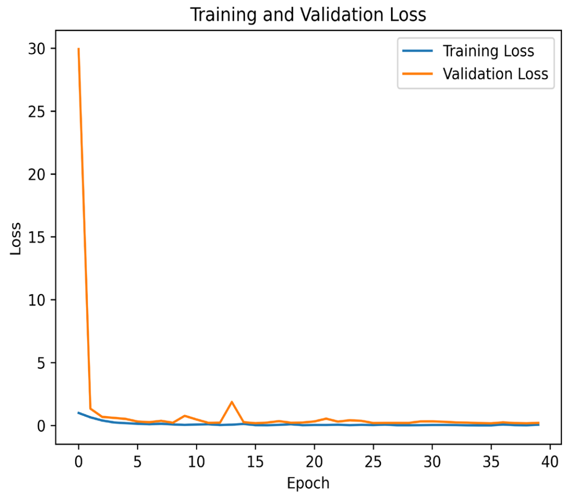
(c) Accuracy graph for ResNet50



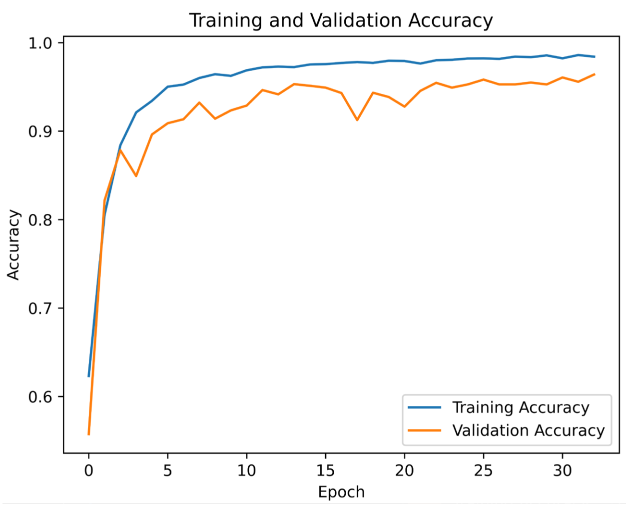
(d) Loss graph for EfficientNetB2



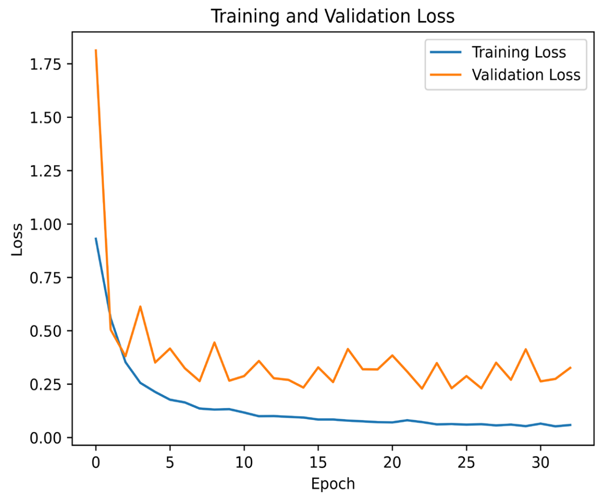
(e) Loss graph for DenseNet121



(f) Loss graph for ResNet50



(g) Accuracy graph for ensemble model



(h) Loss graph for ensemble model

Figure 6. Training and validation accuracy + loss graph for individual models and ensemble model.

Table 6 displays the evaluation metrics calculated for both training and testing of the ensemble model and the individual pre-trained models. The results indicate that the ensemble model surpassed the individual models, enhancing the generalization and classification process. Among the individual models, EfficientNetB2 exhibited impressive performance in grading DR into 5 classes, while ResNet50 and DenseNet121 also played a significant role in classifying and detecting DR. Further details of the results can be seen in **Table 6**.

The performance of our Ensemble model in this study is compared with the previous research in terms of multi-classification and test accuracy. Various researchers have contributed to the detection and grading of diabetic retinopathy (DR). Many datasets such as APTOS, Messidor, EyePACS, and DeepDR, along with diverse preprocessing techniques, data augmentation methods, and dataset split ratios were employed across these studies. Despite these variations, our proposed model demonstrated the highest accuracy by utilizing the proposed pre-processing techniques and data fusion. Table 7 shows a detailed comparison between our Ensemble Model and previous research techniques.

Table 6. Performance evaluation of individual models and ensemble model using various measures.

Performance Metrics	MODELS							
	EfficientNetB2		DenseNet121		ResNet50		Ensemble	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
Accuracy (ACC)	0.9949	0.9627	0.9735	0.9121	0.9942	0.9495	0.9969	0.9696
Misclassification Rate (MCR)	0.0050	0.0372	0.0264	0.0878	0.0057	0.0504	0.0030	0.0303
Specificity (SPC)	0.9987	0.9907	0.9934	0.9781	0.9985	0.9874	0.9992	0.9923
Sensitivity (SEN)	0.9950	0.9627	0.9729	0.9105	0.9941	0.9490	0.9968	0.9693
F1-Score (F1)	0.9950	0.9628	0.9737	0.9129	0.9941	0.9498	0.9968	0.9695
Positive Prediction Value (PPV)	0.9946	0.9623	0.9736	0.9165	0.9940	0.9492	0.9968	0.9698
Negative Prediction Value (NPV)	0.9987	0.9907	0.9934	0.9785	0.9985	0.9874	0.9992	0.9924
False Positive Ratio (FPR)	0.0012	0.0092	0.0065	0.0218	0.0014	0.0125	0.0007	0.0076
False Negative Ratio (FNR)	0.0049	0.0372	0.0270	0.0894	0.0058	0.0509	0.0031	0.0306
Positive Likelihood Ratio (LR+)	∞	115.24	696.21	78.048	1237.2	91.934	1688.44	133.35
Negative Likelihood Ratio (LR−)	0.0049	0.0375	0.0271	0.0905	0.0058	0.0516	0.0031	0.0308

Table 7. Performance evaluation of individual models and ensemble model using various measures.

Reference	Year	Methodology	Accuracy
Alyoubi *et al.* [40]	2021	Model Fusion (CNN512 + YOLOv3)	89%
Ebrahimi *et al.* [41]	2023	CNN (intermediate fusion) Architecture	92.65%
Shaban *et al.* [42]	2020	CNN Model	88%-89%
Mushtaq *et al.* [43]	2021	DenseNet-169	90%
Raja Kumar *et al.* [44]	2021	CNN Model	94.44%
Menaouer *et al.* [45]	2022	Hybrid models (CNN, VGG16 and VGG19)	90.60%
Abbood *et al.* [46]	2022	ResNet-50	93.6%
Current Study	2024	Ensemble Model	96.96%

5. Conclusion and Future Work

This study introduces an ensemble classification framework for grading diabetic retinopathy into 5 classes utilizing data fusion and transfer learning techniques. Diabetic retinopathy stands as a primary cause of blindness among individuals with diabetes, underlining the importance of timely detection and treatment to prevent vision loss.

This research combines three benchmark diabetic retinopathy datasets—APTOS, IDRiD, and Messidor-2—to create a new fused dataset. Initially, dataset underwent cropping to remove unwanted regions, followed by image denoising and Contrast Limited Adaptive Histogram Equalization (CLAHE). Dataset imbalance was addressed using SMOTE. Further dataset diversification was accomplished through augmentation techniques, such as flipping, rotation, and zooming. For classification, three pre-trained models, EfficientNetB2, DenseNet121 and ResNet50, were fine-tuned and trained. An ensemble model was created afterwards by averaging the predictions of these pre-trained models. This ensemble model demonstrated superior performance, achieving a training accuracy of 99.69% and the highest test accuracy of 96.96%. It outperformed the previous published techniques on comparison.

For future work, diabetic retinopathy can be classified by the integration of convolutional neural networks (CNNs) and extending beyond the utilization of transfer learning models. Experimentation with fusing more than three datasets particularly those containing a larger number of fundus images can lead to more accurate training outcomes. Many other pre-processing techniques can be explored that can contribute to achieving higher accuracy levels. Moreover, different techniques of ensemble can be worked with such as bagging, boosting etc.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Jabbar, M.K., Yan, J., Xu, H., Ur Rehman, Z. and Jabbar, A. (2022) Transfer Learning-Based Model for Diabetic Retinopathy Diagnosis Using Retinal Images. *Brain Sciences*, **12**, Article 535. https://doi.org/10.3390/brainsci12050535

[2] Ghosh, S. and Chatterjee, A. (2023) Transfer-Ensemble Learning Based Deep Convolutional Neural Networks for Diabetic Retinopathy Classification. 2023 3*rd International Conference on Advancement in Electronics & Communication Engineering* (*AECE*), Ghaziabad, 23-24 November 2023, 489-493. https://doi.org/10.1109/aece59614.2023.10428233

[3] AbdelMaksoud, E., Barakat, S. and Elmogy, M. (2022) A Computer-Aided Diagnosis System for Detecting Various Diabetic Retinopathy Grades Based on a Hybrid Deep Learning Technique. *Medical & Biological Engineering & Computing*, **60**, 2015-2038. https://doi.org/10.1007/s11517-022-02564-6

[4] Akhtar, S., Aftab, S., Ahmad, M. and Ihnaini, B. (2024. A Classification Framework for Diabetic Retinopathy Detection Using Transfer Learning. 2024 2*nd International Conference on Cyber Resilience* (*ICCR*), Dubai, 26-28 February 2024, 1-5. https://doi.org/10.1109/iccr61006.2024.10533135

[5] Khan, M.A., Elmitwally, N.S., Abbas, S., Aftab, S., Ahmad, M., Fayaz, M., *et al.* (2022) Software Defect Prediction Using Artificial Neural Networks: A Systematic Literature Review. *Scientific Programming*, **2022**, Article ID: 2117339. https://doi.org/10.1155/2022/2117339

[6] Kassani, S.H., Kassani, P.H., Khazaeinezhad, R., Wesolowski, M.J., Schneider, K.A. and Deters, R. (2019) Diabetic Retinopathy Classification Using a Modified Xception Architecture. 2019 *IEEE International Symposium on Signal Processing and Information Technology* (*ISSPIT*), Ajman, 10-12 December 2019, 1-6. https://doi.org/10.1109/isspit47144.2019.9001846

[7] Thota, N.B. and Umma Reddy, D. (2020) Improving the Accuracy of Diabetic Retinopathy Severity Classification with Transfer Learning. 2020 *IEEE* 63*rd International Midwest Symposium on Circuits and Systems* (*MWSCAS*), Springfield, 9-12 August 2020, 1003-1006. https://doi.org/10.1109/mwscas48704.2020.9184473

[8] Gangwar, A.K. and Ravi, V. (2020) Diabetic Retinopathy Detection Using Transfer Learning and Deep Learning. In: Bhateja, V., Peng, S.L., Satapathy, S.C. and Zhang, Y.D., Eds., *Evolution in Computational Intelligence*, Springer Singapore, 679-689. https://doi.org/10.1007/978-981-15-5788-0_64

[9] Hagos, M.T. and Kant, S. (2019) Transfer Learning Based Detection of Diabetic Retinopathy from Small Dataset. arXiv: 1905.07203. https://doi.org/10.48550/arXiv.1905.07203

[10] Bhardwaj, C., Jain, S. and Sood, M. (2021) Transfer Learning Based Robust Automatic Detection System for Diabetic Retinopathy Grading. *Neural Computing and Applications*, **33**, 13999-14019. https://doi.org/10.1007/s00521-021-06042-2

[11] Ihnaini, B., Akhtar, S., Ahmad, M. and Aftab, S. (2024) Data Fusion Based Ensemble Transfer Learning Approach to Detect Diabetic Retinopathy. 2024 2*nd International Conference on Cyber Resilience* (*ICCR*), Dubai, 26-28 February 2024, 1-5. https://doi.org/10.1109/iccr61006.2024.10532998

[12] Akhtar, S., Aftab, S., Ahmad, M. and Ihnaini, B. (2024) A Transfer Learning Based Framework for Diabetic Retinopathy Detection Using Data Fusion. 2024 2*nd International Conference on Cyber Resilience* (*ICCR*), Dubai, 26-28 February 2024, 1-5. https://doi.org/10.1109/iccr61006.2024.10533112

[13] Martinez-Murcia, F.J., Ortiz, A., Ramírez, J., Górriz, J.M. and Cruz, R. (2021) Deep Residual Transfer Learning for Automatic Diagnosis and Grading of Diabetic Retinopathy. *Neurocomputing*, **452**, 424-434. https://doi.org/10.1016/j.neucom.2020.04.148

[14] Le, D., Alam, M., Yao, C.K., Lim, J.I., Hsieh, Y., Chan, R.V.P., *et al.* (2020) Transfer Learning for Automated OCTA Detection of Diabetic Retinopathy. *Translational Vision Science & Technology*, **9**, 35. https://doi.org/10.1167/tvst.9.2.35

[15] Aswathi, T., Swapna, T.R. and Padmavathi, S. (2021) Transfer Learning Approach for Grading of Diabetic Retinopathy. *Journal of Physics*: *Conference Series*, **1767**, Article ID: 012033. https://doi.org/10.1088/1742-6596/1767/1/012033

[16] Vaibhavi, P.M. and Manjesh, R. (2021) Binary Classification of Diabetic Retinopathy Detection and Web Application. *International Journal of Research in Engineering, Science and Management*, **4**, 142-145. https://journal.ijresm.com/index.php/ijresm/article/view/1000

[17] El Houby, E.M.F. (2021) Using Transfer Learning for Diabetic Retinopathy Stage Classification. *Applied Computing and Informatics*. https://doi.org/10.1108/aci-07-2021-0191

[18] Sh. Daoud, M., Fatima, A., Ahmad Khan, W., Adnan Khan, M., Abbas, S., Ihnaini, B., *et al.* (2022) Joint Channel and Multi-User Detection Empowered with Machine Learning. *Computers*, *Materials & Continua*, **70**, 109-121. https://doi.org/10.32604/cmc.2022.019295

[19] Zhang, X., Kim, Y., Chung, Y., Yoon, S., Rhee, S. and Kim, Y.S. (2023) A Wrapped

Approach Using Unlabeled Data for Diabetic Retinopathy Diagnosis. *Applied Sciences*, **13**, Article 1901. https://doi.org/10.3390/app13031901

[20] Li, X., Pang, T., Xiong, B., Liu, W., Liang, P. and Wang, T. (2017) Convolutional Neural Networks Based Transfer Learning for Diabetic Retinopathy Fundus Image Classification. 2017 10*th International Congress on Image and Signal Processing*, *BioMedical Engineering and Informatics* (*CISP-BMEI*), Shanghai, 14-16 October 2017, 1-11. https://doi.org/10.1109/cisp-bmei.2017.8301998

[21] Ghazal, M., Ali, S.S., Mahmoud, A.H., Shalaby, A.M. and El-Baz, A. (2020) Accurate Detection of Non-Proliferative Diabetic Retinopathy in Optical Coherence Tomography Images Using Convolutional Neural Networks. *IEEE Access*, **8**, 34387-34397. https://doi.org/10.1109/access.2020.2974158

[22] Kotiyal, B. and Pathak, H. (2022) Diabetic Retinopathy Binary Image Classification Using Pyspark. *International Journal of Mathematical*, *Engineering and Management Sciences*, **7**, 624-642. https://doi.org/10.33889/ijmems.2022.7.5.041

[23] Lin, C. and Wu, K. (2023) Development of Revised Resnet-50 for Diabetic Retinopathy Detection. *BMC Bioinformatics*, **24**, Article No. 157. https://doi.org/10.1186/s12859-023-05293-1

[24] Liu, H., Yue, K., Cheng, S., Pan, C., Sun, J. and Li, W. (2020) Hybrid Model Structure for Diabetic Retinopathy Classification. *Journal of Healthcare Engineering*, **2020**, Article ID: 8840174. https://doi.org/10.1155/2020/8840174

[25] Chilukoti, S.V., Maida, A.S. and Hei, X. (2022) Diabetic Retinopathy Detection Using Transfer Learning from Pre-Trained Convolutional Neural Network Models. TechRxiv. https://dx.doi.org/10.36227/techrxiv.18515357.v1

[26] Masood, S., Luthra, T., Sundriyal, H. and Ahmed, M. (2017) Identification of Diabetic Retinopathy in Eye Images Using Transfer Learning. 2017 *International Conference on Computing*, *Communication and Automation* (*ICCCA*), Greater Noida, 5-6 May 2017, 1183-1187.

[27] Gao, Z., Pan, X., Shao, J., Jiang, X., Su, Z., Jin, K., *et al*. (2022) Automatic Interpretation and Clinical Evaluation for Fundus Fluorescein Angiography Images of Diabetic Retinopathy Patients by Deep Learning. *British Journal of Ophthalmology*, **107**, 1852-1858. https://doi.org/10.1136/bjo-2022-321472

[28] APTOS 2019 Blindness Detection. https://www.kaggle.com/competitions/aptos2019-blindness-detection

[29] Indian Diabetic Retinopathy Image Dataset (IDRiD). https://ieee-dataport.org/open-access/indian-diabetic-retinopathy-image-dataset-idrid

[30] https://www.adcis.net/en/third-party/messidor2/

[31] Çinarer, G., Kiliç, K. and Parlar, T. (2022) A Deep Transfer Learning Frame-Work for the Staging of Diabetic Retinopathy. *Journal of Scientific Reports-A*, **51**, 106-119.

[32] Sebti, R., Zroug, S., Kahloul, L. and Benharzallah, S. (2022) A Deep Learning Approach for the Diabetic Retinopathy Detection. In: Ben Ahmed, M., Boudhir, A.A., Karaș, İ.R., Jain, V. and Mellouli, S., Eds., *Innovations in Smart Cities Applications Volume* 5, Springer International Publishing, 459-469. https://doi.org/10.1007/978-3-030-94191-8_37

[33] Skouta, A., Elmoufidi, A., Jai-Andaloussi, S. and Ochetto, O. (2020) Automated Binary Classification of Diabetic Retinopathy by Convolutional Neural Networks. In: Saeed, F., Al-Hadhrami, T., Mohammed, F. and Mohammed, E., Eds., *Advances on Smart and Soft Computing*, Springer Singapore, 177-187. https://doi.org/10.1007/978-981-15-6048-4_16

[34] Elsharkawy, M., Sharafeldeen, A., Soliman, A., Khalifa, F., Ghazal, M., El-Daydamony, E., *et al.* (2022) A Novel Computer-Aided Diagnostic System for Early Detection of Diabetic Retinopathy Using 3D-OCT Higher-Order Spatial Appearance Model. *Diagnostics*, **12**, Article 461. https://doi.org/10.3390/diagnostics12020461

[35] Akhtar, S. and Aftab, S. (2024) A Framework for Diabetic Retinopathy Detection Using Transfer Learning and Data Fusion. *International Journal of Information Technology and Computer Science*, **16**, 61-73. https://doi.org/10.5815/ijitcs.2024.06.05

[36] Wang, J., Bai, Y. and Xia, B. (2019) Feasibility of Diagnosing Both Severity and Features of Diabetic Retinopathy in Fundus Photography. *IEEE Access*, **7**, 102589-102597. https://doi.org/10.1109/access.2019.2930941

[37] Salvi, R.S., Labhsetwar, S.R., Kolte, P.A., Venkatesh, V.S. and Baretto, A.M. (2021) Predictive Analysis of Diabetic Retinopathy with Transfer Learning. 2021 4*th Biennial International Conference on Nascent Technologies in Engineering* (*ICNTE*), NaviMumbai, 15-16 January 2021, 1-6. https://doi.org/10.1109/icnte51185.2021.9487789

[38] Bora, A., Balasubramanian, S., Babenko, B., Virmani, S., Venugopalan, S., Mitani, A., *et al.* (2021) Predicting the Risk of Developing Diabetic Retinopathy Using Deep Learning. *The Lancet Digital Health*, **3**, e10-e19. https://doi.org/10.1016/s2589-7500(20)30250-8

[39] Akhtar, S., Aftab, S., Ahmad, M. and Akhtar, A. (2024) Diabetic Retinopathy Severity Grading Usingtransfer Learning Techniques. *International Journal of Engineering and Manufacturing*, **14**, 41-53. https://doi.org/10.5815/ijem.2024.06.04

[40] Alyoubi, W.L., Abulkhair, M.F. and Shalash, W.M. (2021) Diabetic Retinopathy Fundus Image Classification and Lesions Localization System Using Deep Learning. *Sensors*, **21**, Article 3704. https://doi.org/10.3390/s21113704

[41] Ebrahimi, B., Le, D., Abtahi, M., Dadzie, A.K., Lim, J.I., Chan, R.V.P., *et al.* (2023) Optimizing the OCTA Layer Fusion Option for Deep Learning Classification of Diabetic Retinopathy. *Biomedical Optics Express*, **14**, 4713-4724. https://doi.org/10.1364/boe.495999

[42] Shaban, M., Ogur, Z., Mahmoud, A., Switala, A., Shalaby, A., Abu Khalifeh, H., *et al.* (2020) A Convolutional Neural Network for the Screening and Staging of Diabetic Retinopathy. *PLOS ONE*, **15**, e0233514. https://doi.org/10.1371/journal.pone.0233514

[43] Mushtaq, G. and Siddiqui, F. (2021) Detection of Diabetic Retinopathy Using Deep Learning Methodology. *IOP Conference Series*: *Materials Science and Engineering*, **1070**, Article ID: 012049. https://doi.org/10.1088/1757-899x/1070/1/012049

[44] Raja Kumar, R., Pandian, R., Prem Jacob, T., Pravin, A. and Indumathi, P. (2021) Detection of Diabetic Retinopathy Using Deep Convolutional Neural Networks. In: Smys, S., Tavares, J.M.R.S., Bestak, R. and Shi, F., Eds., *Computational Vision and Bio-Inspired Computing*, Springer Singapore, 415-430. https://doi.org/10.1007/978-981-33-6862-0_34

[45] Menaouer, B., Dermane, Z., El Houda Kebir, N. and Matta, N. (2022) Diabetic Retinopathy Classification Using Hybrid Deep Learning Approach. *SN Computer Science*, **3**, Article No. 357. https://doi.org/10.1007/s42979-022-01240-8

[46] Abbood, S.H., Hamed, H.N.A., Rahim, M.S.M., Rehman, A., Saba, T. and Bahaj, S.A. (2022) Hybrid Retinal Image Enhancement Algorithm for Diabetic Retinopathy Diagnostic Using Deep Learning Model. *IEEE Access*, **10**, 73079-73086. https://doi.org/10.1109/access.2022.3189374