

Diabetic Retinopathy Severity Grading Using Transfer Learning Techniques

Samia Akhtar

Department of Computer Science, Virtual University of Pakistan, Lahore, Pakistan

Email: samiaakhtar9898@gmail.com

ORCID iD: <https://orcid.org/0009-0004-4345-793X>

Shabib Aftab*

Department of Computer Science, Virtual University of Pakistan, Lahore, Pakistan

Email: shabib.aftab@vu.edu.pk

ORCID iD: <https://orcid.org/0000-0002-7662-1394>

*Corresponding Author

Munir Ahmad

School of Computer Science, NCBA&E, Lahore, Pakistan

Email: munir@ncbae.edu.pk

ORCID iD: <https://orcid.org/0000-0002-5240-0984>

Asma Akhtar

Department of Computer Science, Virtual University of Pakistan, Lahore, Pakistan

Email: asmaakhtarjanjua@gmail.com

ORCID iD: <https://orcid.org/0009-0008-7289-3203>

Received: 17 February, 2024; Revised: 03 July, 2024; Accepted: 14 August, 2024; Published: 08 December, 2024

Abstract: Diabetic Retinopathy is a severe eye condition originating as a result of long term diabetes mellitus. Timely detection is essential to prevent it from progressing to more advanced stages. Manual detection of DR is labor-intensive and time-consuming, requiring expertise and extensive image analysis. Our research aims to develop a robust and automated deep learning model to assist healthcare professionals by streamlining the detection process and improving diagnostic accuracy. This research proposes a multi-classification framework using Transfer Learning for diabetic retinopathy grading among diabetic patients. An image based dataset, APTOS 2019 Blindness Detection, is utilized for our model training and testing. Our methodology involves three key preprocessing steps: 1) Cropping to remove extraneous background regions, 2) Contrast enhancement using CLAHE (Contrast Limited Adaptive Histogram Equalization) and 3) Resizing to a consistent dimension of 224x224x3. To address class imbalance, we applied SMOTE (Synthetic Minority Over-sampling Technique) for balancing the dataset. Data augmentation techniques such as rotation, zooming, shifting, and brightness adjustment are used to further enhance the model's generalization. The dataset is split to a 70:10:20 ratios for training, validation and testing. For classification, EfficientNetB3 and Xception, two transfer learning models, are used after fine-tuning which includes addition of dense, dropout and fully connected layers. Hyper parameters such as batch size, no. of epochs, optimizer etc were adjusted prior model training. The performance of our model is evaluated using various performance metrics including accuracy, specificity, sensitivity and others. Results reveal the highest test accuracy of 95.16% on the APTOS dataset for grading diabetic retinopathy into five classes using the EfficientNetB3 model followed by a test accuracy of 92.66% using Xception model. Our top-performing model, EfficientNetB3, was compared against various state-of-the-art approaches, including DenseNet-169, hybrid models, and ResNet-50, where our model outperformed all these methodologies.

Index Terms: Diabetic Retinopathy, Transfer Learning, Artificial Intelligence, Deep Learning, Diabetes.

1. Introduction

Diabetic retinopathy (DR) is a prevalent disease that affects millions of people worldwide [1]. It arises as a consequence of diabetes mellitus and can cause vision problems [2]. DR originates when long-term exposure to high blood sugar levels harms the tiny blood vessels in the retina (light-sensitive tissue at the back of the eye), resulting in inadequate oxygen delivery, fluid leakage, total blockage or swelling of the blood vessels [3,4]. Vision gets impaired as a result of these blood vessels getting weakened and leaking blood or fluid into the retina. Individuals who have poorly managed blood sugar levels and have had diabetes for a long time are at a higher risk of developing diabetic retinopathy.

Early detection and treatment of this disease is crucial in order to assist patients restore normal function to their retinas and prevent DR from progressing to more severe stages [5,6]. Regular eye examinations are necessary for monitoring the progression of diabetic retinopathy and prevention of vision loss. Diabetic Retinopathy can be detected in the retinal examination through abnormalities like exudates (EX), micro-aneurysms (MA) and hemorrhages (HM) [7, 8]. Micro aneurysms are an early indication of diabetic retinopathy and appear as tiny bulges within the retinal blood vessel [9]. Exudates are yellow-colored deposits that build up in the retina, usually due to damaged and leaking blood vessels [10]. Hemorrhages appear as dot-like areas ranging from small to large based on the extent of the bleeding from damaged blood vessels into the retinal tissue [11]. These lesions help in detecting the presence of DR in diabetic patients. During the initial stages of diabetic retinopathy, individuals might not notice any symptoms. These can be mild and may include slight blurriness or fluctuation in vision [12].

In more severe stages, there may be blotches in the vision, an abrupt loss of vision, and perhaps partial or complete blindness [13]. The advancement of artificial intelligence (AI), especially in the fields of machine learning (ML) and deep learning (DL), has drastically changed how diabetic retinopathy is identified and treated [14]. The traditional process of detecting DR by manual methods requires a lot of time and effort which frequently results in incorrect classifications as the data involved is pretty large [15]. On the other hand, Computer-assisted tools provide accurate detection of diabetic retinopathy. These advanced methods provide quick and accurate evaluation of retinal photographs, supporting early detection and recognition of DR [16, 17]. Transfer learning (TL) and DL algorithms are gaining a lot of popularity for their remarkable execution in image recognition and classification [18]. Transfer learning as a deep learning technique involves taking a pre-trained model that is trained on a large dataset then fine-tuning it on a smaller dataset specific to the target task [19, 20].

The primary goal of our research is to develop a more accurate and automated deep learning model for the detection and classification of diabetic retinopathy which will ultimately enhance the diagnostic capabilities available to healthcare professionals. The motivation behind our approach arises from the shortcomings of existing models that often overlook the importance of robust preprocessing and struggle with class imbalance which leads to reduced accuracy and generalization. By integrating advanced preprocessing techniques, sophisticated model architecture, and robust data augmentation strategies, we aim to create a system that not only identifies various stages of diabetic retinopathy with a high accuracy but also minimizes the manual effort required for image analysis. Our approach seeks to streamline the diagnostic process allowing for timely intervention and improved patient outcomes.

This research introduces a classification framework for grading diabetic retinopathy into 5 classes by using transfer learning models. Two transfer learning models have been used on a DR Dataset to achieve high test accuracy and a comparison of the results has been made with the previous studies. The dataset used in this research consists of fundus images collected from different clinics. The remaining paper is discussed as: literature review is presented in Section 2; proposed methodology is described in Section 3, results are discussed in detail in Section 4 and Section 5 concludes with insights into future work.

2. Related Work

Many researchers have experimented and made meaningful contributions using numerous techniques, algorithms and methods to detect and classify diabetic retinopathy. Their dedication and efforts have helped improve the detection and classification of DR ultimately leading to better diagnosis of this disease.

The researchers of [21] have proposed a deep learning hybrid methodology for detection and multi-classification of diabetic retinopathy. They have used convolutional neural network (CNN) and 2 VGG models: VGG19 and VGG16. Afterwards they combined their used models to create a hybrid model. They included Image enhancement procedures in pre-processing using OpenCV functions. Datasets used in their research are APTOS 2019, Messidor-2 and a public DR Dataset. Their hybrid approach achieved an accuracy of 90.60%. Authors of [22] have proposed a framework for grading DR into 5 classes by introducing the concepts of automated diagnosis systems in their research. Two convolutional neural network models, CNN512 and YOLOv3 based CNN were fused together for the classification of DR stages and lesion localization. The authors worked with APTOS 2019 and DDR datasets. Techniques like CLAHE, noise removal, cropping, augmentation etc. were applied prior model training. CNN512 was trained for classification purpose whereas YOLOv3 model was used for lesion detection. The fusion of these models improved the accuracy on the datasets to 89% with a sensitivity of 89% and 97.3% specificity.

In [23] a deep learning based methodology is adopted by the researchers for grading diabetic retinopathy. They have used widely known datasets EyePACS and Messidor for evaluation. Researchers have contributed a lot in image pre-processing by presenting an algorithm for enhancing the image quality, contrast improvement and noise reduction by using cropping technique and Gaussian blur. A DL model, ResNet50 architecture is used for extracting features and classification. Their proposed methodology achieved an accuracy of 92% on EyePACS dataset and a highest accuracy of 93.6% on Messidor dataset. Authors in [24] presented a Deep Learning based approach for detection and multi-classification of DR into 5 categories. They proposed a Convolutional Neural Network (CNN) model with 18 convolutional layers along with 3 fully connected layers. The dataset they used for their research is APTOS 2019 from Kaggle. Pre-processing techniques such as image resizing and class specific data augmentation of fundus images was done prior classification. Their CNN approach performed well in grading diabetic retinopathy yielding an accuracy of 88-89%. Sensitivity and specificity were 87-89% and 94-95% respectively.

In [25], the researchers have introduced two techniques for classification of DR. First one is SOP (standard operation procedure) which is used in pre-processing and second one is a revised ResNet-50 architecture. They altered the ResNet-50 model by adding regularization and adjusting the weight of the layers. They used Kaggle dataset for evaluation which first underwent image pre-processing techniques. Their results concluded that revised structure of ResNet-50 performed better than the original one yielding an accuracy of 74.32% and also showed better results as compared to other common CNN models. Researchers in [26] contributed in proposing a deep learning model for early detection of DR. They trained a DenseNet-169 CNN model to classify DR as No DR, Mild, Moderate, Severe and Proliferative DR. They used two Kaggle datasets and pre-processed by using cropping, resizing and Gaussian blur techniques. Augmentation techniques were used to balance the number of images in each class. Researchers also proposed a Regression model along with their main DenseNet-169. Regression model was only able to achieve an accuracy of 78% where as DenseNet169 gave an accuracy of 90%.

In [27], the authors proposed a multi-classification framework for detection and classification of diabetic retinopathy. Dataset taken from Kaggle was resized and cropped to get rid of unwanted regions in the images. Data augmentation techniques like scaling, rotation, translation, flipping etc. were considered to increase the size of dataset. Authors proposed a CNN model using MATLAB for classification purpose. Their proposed model achieved an accuracy of 94.44% with 93.96% sensitivity and 98.54% specificity. Authors of [28] have employed 3 different hybrid models for classification of DR into 5 classes. The three hybrid models are Hybrid-a, Hybrid-f, and Hybrid-c. To form these hybrid models, five (5) base CNN models were used: NASNetLarge, EfficientNetB5, EfficientNetB4, InceptionReNetV2 and Xception. Two loss functions were used to train the base models. The output of these base models was then used to train the hybrid models. Researchers experimented with three different datasets, APTOS, EyePACS and DeepDR. Pre-processing was done before model training and also during the training process which mostly involved image enhancement process. Out of all the hybrid models, Hybrid-c model gave the highest accuracy of 86.34%.

All of the discussed studies have explored deep learning techniques for the detection and classification of diabetic retinopathy. These studies used various pre-processing strategies and classification methodologies using distinct DR Datasets. These approaches incorporated convolutional neural networks (CNNs), hybrid models, and even transfer learning but still faced challenges in terms of accuracy, generalization, and handling class imbalance. Additionally, the complexity of some hybrid and fusion models resulted in higher computational costs, making real-world clinical application less feasible. In contrast, our work addresses these limitations by introducing a more robust and efficient solution. Our approach not only outperforms these state-of-the-art methods in terms of accuracy but also implements advanced preprocessing steps, including CLAHE, to enhance image quality and improve generalization. By applying the SMOTE technique for oversampling minority classes our model effectively handles class imbalance reducing bias in classification results. Furthermore, our framework leverages transfer learning with well-established architectures. This enabled rapid convergence and improved feature extraction which are crucial for accurately distinguishing between various stages of diabetic retinopathy with less computational complexity.

3. Materials and Methods

In this study, a multi-classification framework is proposed to grade diabetic retinopathy. The concepts of Transfer Learning are utilized as two transfer learning models are used for classifying this disease into 5 classes. TL models are well known in image processing and disease detection [29]. Figure 1 shows the proposed framework of this research. All the steps followed by the proposed methodology are presented in order. The detailed explanation of Dataset used in this study and steps adopted is given as follows.

3.1 Dataset

The dataset used in this study is APTOS (Asia Pacific Tele-Ophthalmology Society) 2019 Blindness Detection. It is a widely used dataset available at Kaggle [30]. It contains a total of 3662 retinal images captured through fundus photography under various imaging conditions. Each image is rated on a scale of 0 to 4 by the Clinicians and these ratings represent the severity levels of diabetic retinopathy: **0** for "**No DR**", **1** for "**Mild**", **2** for "**Moderate**", **3** for

"**Severe**" and **4** for "**Proliferative DR**". The images were gathered from multiple clinics using a variety of cameras over an extended period of time. Out of all these 5 levels, a sample from each level of this dataset is represented in Figure 2. There is an uneven distribution of images among each level. The number of images per level is given in Table 1.

Table 1. Image distribution per level/class in APTOS dataset

Level	No. of Images
No DR (Class 0)	1805
Mild (Class 1)	370
Moderate (Class 2)	999
Severe (Class 3)	193
Proliferative DR (Class 4)	295
Total	3662

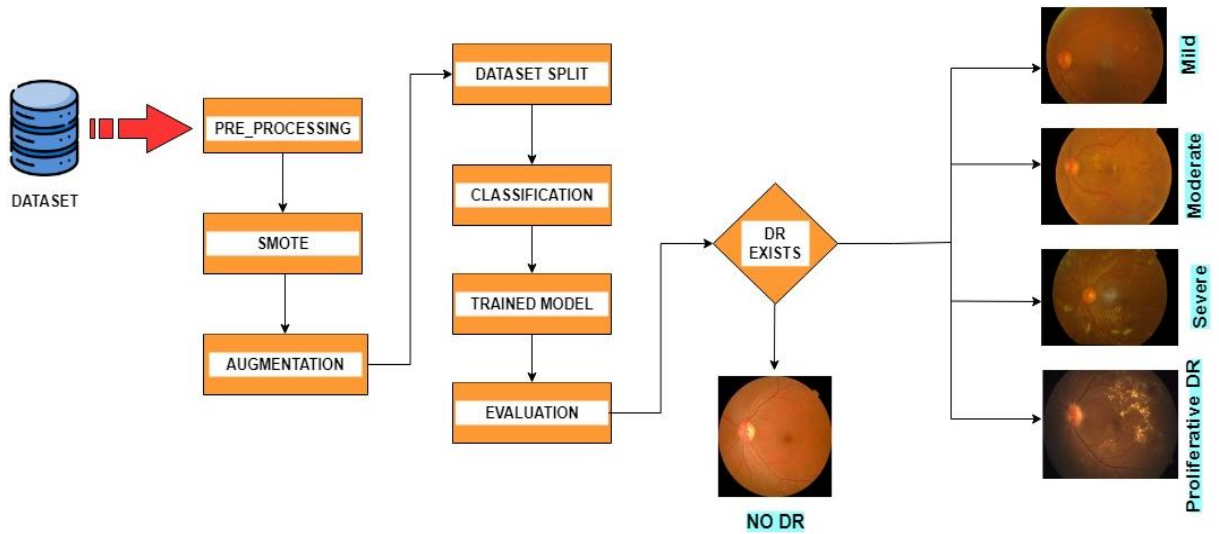


Fig. 1. Proposed Framework for the classification of Diabetic Retinopathy.

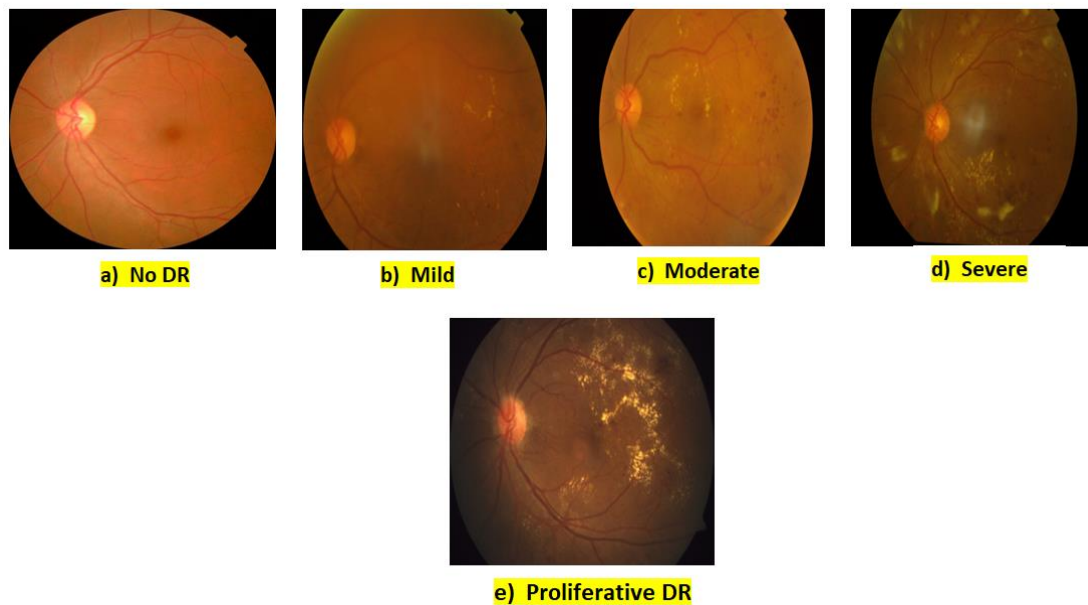


Fig. 2. A sample from each class in APTOS fundus image dataset.

3.2. Pre-Processing

Pre-processing steps are essential before model training to enhance the quality of images and prepare them for accurate classification. There are many techniques that can be adopted for this purpose [31, 32, 33, 34]. In this research, pre-processing is done in three (3) steps: **1) Cropping**, **2) CLAHE**, and **3) Resizing**. The images in dataset were analyzed first. As they have been collected from different clinics and taken from variety of cameras, some images have more noise and extra black back ground than the others. Some of them are way bright with faded details of the features present in the image while others have low contrast.

First step of pre-processing is to crop the images. Some images in the dataset contain unwanted backgrounds, and providing a model with such a dataset that has varying backgrounds can result in performance and generalization problems which reduces the model's accuracy. To address this issue, we cropped the images to the necessary size starting from the where the eye's starts to eliminate the extraneous black background.

Then these cropped images were enhanced using CLAHE (Contrast Limited Adaptive Histogram Equalization) [35]. Our dataset contains images having inconsistencies in brightness and contrast. This step was crucial for revealing finer details and improving the overall contrast and quality while suppressing the noise from the images. We measured its effectiveness by comparing the contrast of images before and after CLAHE application ensuring that improvements were visually and statistically significant.

In last step, we resized these enhanced images. Resizing images to a consistent size ensured uniformity across the dataset which was necessary for efficient model training and to prevent distortions during feature extraction. Images were resized to 224 x 224 x 3 dimension, where first 2 numbers represent pixels in height and width and 3 represents number of color channels in the image i.e. red, green, and blue (RGB). Cropping and Enhancing is shown in Figure 3.

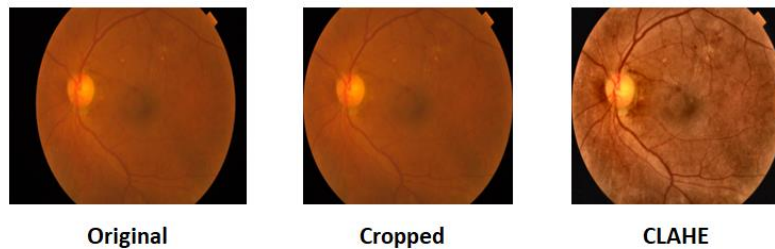


Fig. 3. Original vs. Pre-processed Image

3.3. Balancing Using SMOTE

After the images are pre-processed, they are balanced using an over-sampling technique known as SMOTE (Synthetic Minority Over-sampling Technique). Images in APTOS dataset are unevenly present among each class that can be seen in Table 1. It's an important step to balance the number of images per class so that our TL models won't be biased towards the class with majority samples [36]. The goal here is to increase the number of samples in minority classes only and keep the majority class as it is. In our dataset, Class 0 "NO DR" has the most images. We will increase the samples in rest of the classes to a random number that is close to the number of images in class 0. The new number of images per class after applying SMOTE is given in Table 2.

We conducted experiments with various values for the sampling_strategy and k_neighbors and ended up selecting parameters that yielded optimal performance. We set the sampling_strategy to ensure that the number of images in the minority classes approached somewhere near the number of images of the majority class while k_neighbors was chosen based on the data distribution to enhance the quality of synthetic samples. The impact of SMOTE on model performance was evaluated by analyzing the model's accuracy, sensitivity, specificity etc. This assessment confirmed that this data balancing technique improved classification performance for the minority classes. While SMOTE effectively addresses class imbalance, it is essential to acknowledge its potential downsides. Oversampling can lead to overfitting, as the model may learn from redundant synthetic samples rather than genuine variations in the data. Additionally, there is a risk of reduced data variability, which could negatively affect the model's generalization capabilities. To mitigate these risks, we carefully monitored model performance and applied regularization techniques during training to avoid overfitting in our models.

Table 2. New number of images after SMOTE

	No DR (Class 0)	Mild (Class 1)	Moderate (Class 2)	Severe (Class 3)	Proliferative DR (Class 4)	Total
No. of Original Images	1805	370	999	193	295	3662
No. of Images after SMOTE	1805	1500	1600	1400	1400	7705

3.4 Data Augmentation

The next step is to augment the images in the dataset. The process of data augmentation is widely used in Deep Learning to increase the size of a dataset by applying numerous augmentation techniques [37]. It introduces variation and diversity in data which can prevent over-fitting and help the model generalize the images accurately [38].

For the proposed framework, augmentation techniques used are rotation, zooming, shifting, and brightness and contrast enhancement. These operations are applied randomly to the dataset. After SMOTE the number of images became 7705. Augmentation is applied in such a way that it increases this number twice. Thus after augmentation, the total number of images is 15,410.

3.5 Dataset Split

Before images are fed to the models for training, they are often split to training and testing sets [39, 40]. After augmentation, dataset was split to a 70:10:20 ratio meaning 70% for training set that will be used to train the models, 10% allocated for validation set and 20% for testing set on which the trained model will be evaluated.

3.6 Transfer Learning Models

Compared to training a model from scratch, Transfer Learning accelerates convergence and improves performance, especially when working with limited data. It also reduces the computational cost and training time. The transfer learning models used for this research are EfficientNetB3 and Xception. These models are pre-trained on ImageNet which is a large-scale dataset, containing millions of labeled images across thousand categories [41]. Models trained on this dataset showcase a great performance on tasks related to computer vision [42]. We chose EfficientNetB3 and Xception as the transfer learning models for our study because they offer specific advantages over other well-known architectures like ResNet, AlexNet, and Inception. EfficientNetB3 is particularly notable for its top-notch accuracy which has been proven in various imaging challenges. This makes it a great fit for medical applications that need precise diagnoses. Xception model also helps achieve faster training and better feature extraction which is crucial for accurately identifying the different stages of diabetic retinopathy. Although ResNet and Inception have shown their effectiveness in many tasks, our initial experiments showed that EfficientNetB3 and Xception consistently perform better in medical imaging tasks, including detecting diabetic retinopathy.

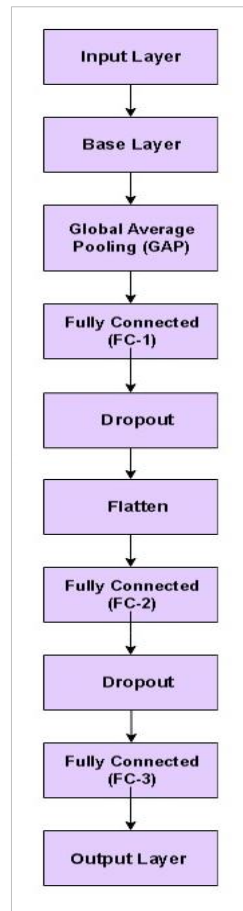


Fig. 4. Fine-tuned Layers of TL Models

Before training, the used models are fine-tuned. The first and last layers are modified for both models with a few additional layers. The input layer is modified and the dimensions were set to 224 x 224. After the base layer, a Global Average Pooling (GAP) layer was added. Rest of the added layers include flatten layer, fully connected layers, dropout layers and output Layer. Activation function used was ReLU. As the proposed framework is for multi-classification, softmax activation function was used in the output layer. Figure 4 represents the fine-tuned architecture used for these models. The details of each layer are shown in Table 3.

Table 3. Detailed structure of the Layers included in TL Models

Layers	Detailed Structure
Input	224 x 224 x 3
Base	TL model
Global Average Pooling	-
Fully Connected (FC-1)	Number of units= 128, activation = ReLu
Dropout	0.1
Flatten	-
Fully Connected (FC-2)	Number of units= 64 activation = ReLu
Dropout	0.1
Fully Connected (FC-3)	Number of units= 32 activation = ReLu
Output	No. of Classes = 5 activation = Softmax

4. Results and Discussion

After preparing the dataset with pre-processing and other techniques, it was used to train the TL models. The data was split in a ratio of 70:10:20. As the total number of images after augmentation were 15, 410, there were a total of 9862 images in training set (70%), 2466 images in validation set (10%) and 3082 images in testing set (20%) after the dataset split. To ensure the reliability of our reported results and mitigate potential biases in data splitting, we employed a robust dataset shuffling technique prior to splitting into training, validation, and test sets. This ensured that each subset was representative of the overall dataset thereby reducing the risk of biased splits or data leakage. The consistent performance metrics achieved across these splits along with the high accuracies demonstrated by our models indicated reliable generalization capabilities. The entire model training and testing process was done using Python programming language and Kaggle Notebooks as a platform to execute the code which provides storage up to 73GB, a RAM of 29GB, and a GPU of 15GB.

In our study, accuracy is our primary performance metric as it provides a straightforward measure of the model's overall correctness in classifying diabetic retinopathy across multiple stages. However, relying solely on accuracy may not provide a complete picture particularly in medical classification tasks where both false positives and false negatives have significant consequences. To ensure a comprehensive evaluation, we have also calculated additional metrics including F1-score, specificity and sensitivity, which are critical for assessing the model's performance. Sensitivity (also known as recall) measures the ability of the model to correctly identify true positive cases. On the other hand, specificity assesses the model's ability to correctly identify true negatives, or healthy patients. To evaluate the performance of the models on both training and testing sets, following measures have been used:

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

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

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

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

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

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

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

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

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

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

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

The transfer learning models used for classification purpose are EfficientNetB3 and Xception. For training the models, different training parameters were experimented with such as batch size, number of epochs, optimizers etc. to improve the training process and analyze which parameters best suit the process. We tested learning rates of 0.0001, 0.001, and 0.01, and found that a learning rate of 0.001 provided the best balance between rapid convergence and training stability. Higher learning rates led to instability, while lower ones significantly slowed down the learning process of our model. For batch size, we explored values of 16, 32, and 64. A batch size of 32 was selected as it offered the best compromise between memory efficiency and model generalization, whereas smaller sizes increased training time, and larger sizes risked overfitting. Initially, we trained the model for up to 100 epochs but settled on 40 epochs after observing that the model converged well by this point. For optimization, we tested different optimizers, including Stochastic Gradient Descent (SGD) and RMSProp, but selected the Adam optimizer due to its adaptive learning rate, which allowed for faster convergence and improved training. Finally, we used the categorical cross-entropy loss function, which is appropriate for multi-class classification tasks and worked effectively in optimizing our model. These hyper-parameters used in the process of training are mentioned in Table 4.

Table 4. Training parameters for both TL models

Input Shape	Optimizer	Learning Rate	Batch Size	Number of Epoch	Loss Function
224 x 224 x 3	Adam	0.001	32	40	Categorical cross-entropy

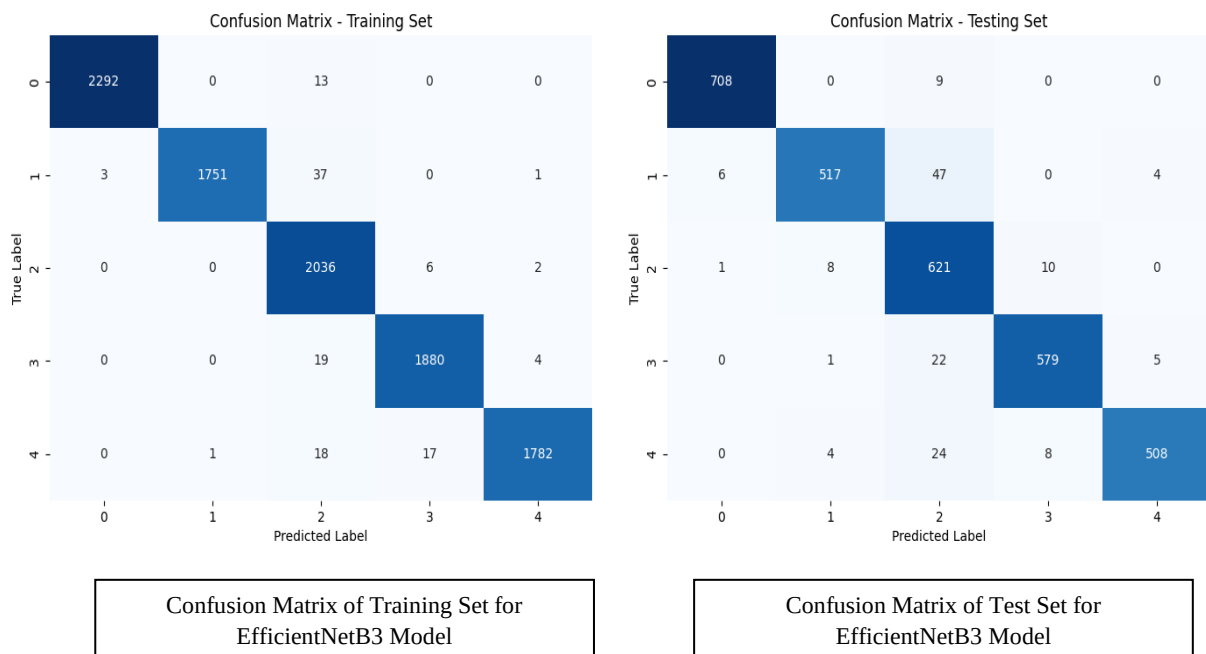


Fig. 5. Confusion matrix for EfficientNetB3 Model

The confusion matrix for both training and testing for EfficientNetB3 model is given in Figure 5. This model achieved a training accuracy of 98.77%, sensitivity of 98.71%, and a specificity of 99.69% on training set. Our model correctly predicted 2292 out of 2305 images of grade 0, 1751 images out of 1792 images of grade 1, 2036 out of 2044 images of grade 2, 1880 out of 1903 images of grade 3 and 1782 images out of 1818 images of grade 4. On the other hand, it achieved a test accuracy of 95.16%, sensitivity of 94.92%, and a specificity of 98.79% on test set where it correctly predicted 708 out of 717 images of grade 0, 517 images out of 574 images of grade 1, 621 out of 640 images of grade 2, 579 out of 607 images of grade 3 and 508 images out of 544 images of grade 4.

When Xception model was trained on the training set, it achieved a training accuracy of 97.09%, sensitivity of 96.98%, and a specificity of 99.27% and a test accuracy of 92.66%, sensitivity of 92.39%, and a specificity of 98.16% on test set. During training model correctly predicted 2300 out of 2305 images of grade 0, 1786 images out of 1792 images of grade 1, 2037 out of 2044 images of grade 2, 1687 out of 1903 images of grade 3 and 1766 images out of 1818 images of grade 4. During Testing Xception model was able to correctly predict 714 out of 717 images of grade 0, 547 images out of 574 images of grade 1, 608 out of 640 images of grade 2, 490 out of 607 images of grade 3 and 497 images out of 544 images of grade 4. Train and Test confusion matrix for this model are shown in Figure 6. It can be clearly seen that EfficientNetB3 outperformed Xception model in classifying the images into 5 classes.

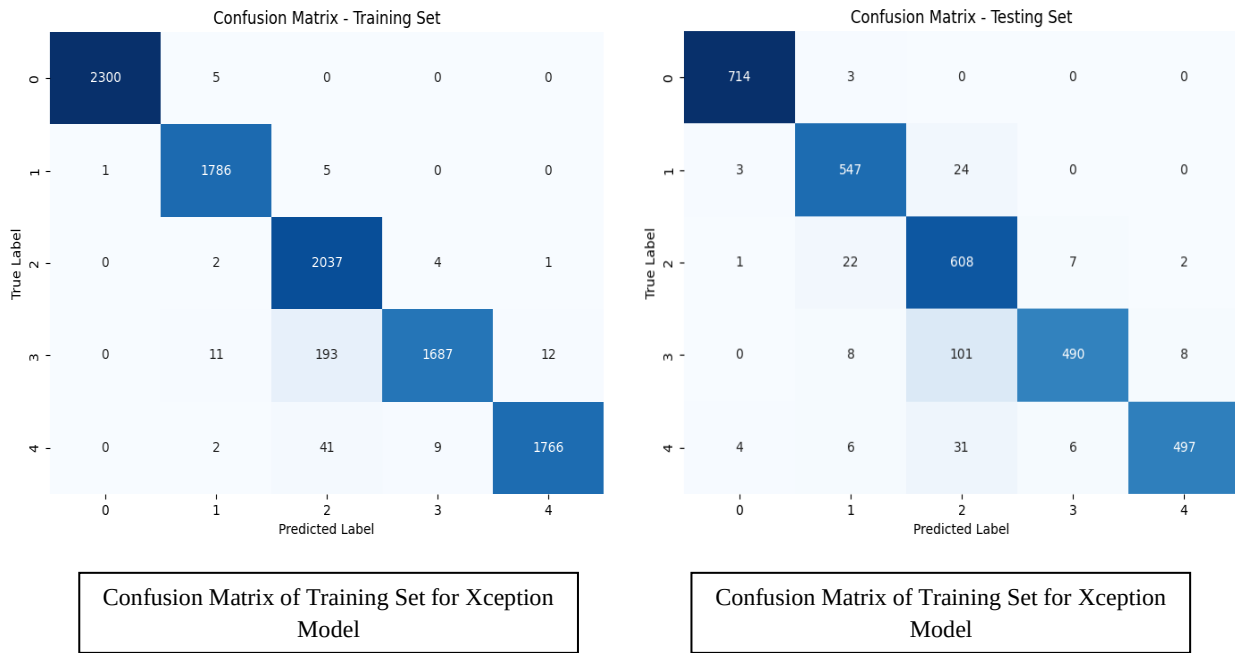


Fig. 6. Confusion matrix for Xception Model

Notably, the EfficientNetB3 model misclassified several images in grade 3 as grade 2, indicating a need for improved feature differentiation between these grades. Similarly, the Xception model demonstrated a higher false negative rate in grade 3 predictions. These misclassifications may be attributed to class imbalance within the dataset, which can lead the model to perform poorly on underrepresented classes. Nevertheless, both models exhibited superior performance overall, achieving high accuracy rates of 95.16% for EfficientNetB3 and 92.66% for Xception. Addressing these issues by incorporating additional data augmentation techniques or balancing the dataset could enhance the models' ability to accurately classify images across all grades, thereby improving overall performance in diabetic retinopathy grading. Table 5 presents the performance evaluation of the two TL models on training and testing set using performance measures.

Table 5. Performance Evaluation of EfficientNetB3 and Xception using various measures

TL Models		Accuracy	Misclassification Rate	Specificity	F1-score	Sensitivity	Positive Prediction Value	Negative Prediction Value	False Positive Ratio	False Negative Ratio	Likelihood Ratio Positive	Likelihood Ratio Negative
EfficientNetB3	Training	0.9877	0.0123	0.9969	0.9879	0.9871	0.9882	0.9969	0.0031	0.0129	2389.6	0.0129
	Testing	0.9516	0.0483	0.9879	0.9524	0.9492	0.9554	0.9881	0.0121	0.0508	185.02	0.0513
Xception	Training	0.9709	0.0290	0.9927	0.9876	0.9698	0.9737	0.9929	0.0073	0.0302	1823.8	0.0303
	Testing	0.9266	0.0733	0.9816	0.9256	0.9239	0.9345	0.9821	0.0184	0.0761	151.22	0.0772

The results of this research are compared with the existing methodologies. Table 6 provides a comparison between the best results of this research and previous studies. The comparison is made in terms of Accuracy and on multi-classification. The proposed methodology performed better than the previous studies. Some researchers used fused models for classification of fundus images in their studies. Others worked with machine learning classifiers and deep learning feature extraction through CNN creating hybrid models for better performance. Some proposed their own convolutional neural network CNN architectures for grading DR. Transfer learning models were also used in studies of a few researchers and they worked with pre-trained models like ResNet, EfficientNet, and GoogleNet. The accuracies these studies achieved are discussed in Table 6. Accuracy achieved by this study using EfficientNetB3 is greater than the previous studies.

Table 6. Comparison of proposed technique with previous studies in terms of Accuracy and multi-classification.

Reference	Methodology	Accuracy
Ebrahimi et al [43]	CNN (intermediate fusion) Architecture	92.65%
Shakibania et al [44]	Deep dual-branch model (transfer learning)	89.60%
Jabbar et al [45]	GoogleNet + ResNet	94%
Singh et al [46]	VGG16 + Logistic Regression	90.4%
Romero et al [47]	FCN + Attention mechanism	83.70%
Mutawa et al [48]	CNN Model	72%
Chilukoti et al [49]	EffNet_b3_60	87%
Boruah et al [50]	ResNet2.0	91%
Current Study	EfficientNetB3	95.16%

5. Conclusions and Future Work

This research presents a multi-classification framework for grading diabetic retinopathy into 5 classes using Transfer Learning. DR is a severe eye condition which causes vision impairment if not treated on time that's why early detection of this disease and regular eye-screening can save diabetic patients from severe vision conditions. For this study, APTOS 2019 dataset is used for DR detection and classification. Fundus images from the dataset were pre-processed, balanced and augmented. Cropping was done to remove extra background and CLAHE was used to enhance the fundus images. Original dataset was unbalanced which was handled using SMOTE. Two TL models were trained on the dataset. These models used after fine-tuning are EfficientNetB3 and Xception. EfficientNetB3 model achieved a training accuracy of 98.77% and scored the highest test accuracy of 95.16% whereas Xception model achieved 97.09% accuracy on training set and 92.66% on test set. EfficientNetB3 model outperformed Xception model with its superior performance. Many other researchers have contributed their results and findings in detecting and classifying diabetic retinopathy. On comparison of this research with the recent existing studies, the methodology of this research performed well. A potential limitation of our research is the class imbalance observed in the dataset, which affected model performance and led to misclassification. To enhance accuracy, future studies would focus incorporating additional data augmentation techniques. Additionally, since this study relied solely on the APTOS 2019 dataset exploring a broader range of distinct diabetic retinopathy datasets, especially those with larger numbers of fundus images, would help improve training robustness. Integrating various deep learning architectures such as convolutional neural networks (CNNs) and pre-trained models along with tailoring pre-processing techniques to specific retinal image requirements will also be essential for achieving higher accuracy in future studies.

References

- [1] Skouta, Ayoub, Abdelali Elmoufidi, Said Jai-Andaloussi, and Ouail Ochetto. "Automated binary classification of diabetic retinopathy by convolutional neural networks." In *Advances on Smart and Soft Computing: Proceedings of ICACI*, Springer Singapore, pp. 177-187, 2021, doi: https://doi.org/10.1007/978-981-15-6048-4_16.
- [2] Ahmed, Usama, Ghassan F. Issa, Muhammad Adnan Khan, Shabib Aftab, Muhammad Farhan Khan, Raed AT Said, Taher M. Ghazal, and Munir Ahmad. "Prediction of diabetes empowered with fused machine learning." *IEEE Access* 10, pp: 8529-8538, 2022, doi: <https://doi.org/10.1109/ACCESS.2022.3142097>
- [3] Kassani, Sara Hosseinzadeh, Peyman Hosseinzadeh Kassani, Reza Khazaeinezhad, Michal J. Wesolowski, Kevin A. Schneider, and Ralph Deters. "Diabetic retinopathy classification using a modified xception architecture." In *2019 IEEE international symposium on signal processing and information technology (ISSPIT)*, IEEE, pp. 1-6, 2019, doi: <https://doi.org/10.1109/ISSPIT47144.2019.9001846>.

- [4] Le, David, Minhaj Alam, Cham K. Yao, Jennifer I. Lim, Yi-Ting Hsieh, Robison VP Chan, Devrim Toslak, and Xincheng Yao. "Transfer learning for automated OCTA detection of diabetic retinopathy." *Translational Vision Science & Technology* 9, no. 2, pp. 30-35, 2020, doi: <https://doi.org/10.1167%2Ftvst.9.2.35>.
- [5] Aswathi, T., T. R. Swapna, and S. Padmavathi. "Transfer learning approach for grading of diabetic retinopathy." *Journal of Physics: Conference Series*, vol. 1767, no. 1, pp. 012033, 2021, doi: [10.1088/1742-6596/1767/1/012033](https://doi.org/10.1088/1742-6596/1767/1/012033).
- [6] Butt, Muhammad Mohsin, DNF Awang Iskandar, Sherif E. Abdelhamid, Ghazanfar Latif, and Runna Alghazo. "Diabetic retinopathy detection from fundus images of the eye using hybrid deep learning features." *Diagnostics* 12, no. 7, pp. 1607, 2022, doi: <https://doi.org/10.3390%2Fdiagnostics12071607>.
- [7] Vaibhavi, P. M., and R. Manjesh. "Binary classification of diabetic retinopathy detection and web application." *International Journal of Research in Engineering, Science and Management* 4, no. 7, pp. 142-145, 2021, doi: <https://journal.ijresm.com/index.php/ijresm/article/view/1000>.
- [8] El Houby, Enas MF. "Using transfer learning for diabetic retinopathy stage classification." *Applied Computing and Informatics*, 2021, doi: <https://doi.org/10.1108/ACI-07-2021-0191>.
- [9] Qiao, Lifeng, Ying Zhu, and Hui Zhou. "Diabetic retinopathy detection using prognosis of microaneurysm and early diagnosis system for non-proliferative diabetic retinopathy based on deep learning algorithms." *IEEE Access* 8, pp. 104292-104302, 2020, doi: <https://doi.org/10.1109/ACCESS.2020.2993937>.
- [10] Sundar, Sumod, and S. Sumathy. "Classification of Diabetic Retinopathy disease levels by extracting topological features using Graph Neural Networks." *IEEE Access* 11, pp. 51435-51444, 2023, doi: <https://doi.org/10.1109/ACCESS.2023.3279393>.
- [11] Gao, Zhiyuan, Xiangji Pan, Ji Shao, Xiaoyu Jiang, Zhaoan Su, Kai Jin, and Juan Ye. "Automatic interpretation and clinical evaluation for fundus fluorescein angiography images of diabetic retinopathy patients by deep learning." *British Journal of Ophthalmology* 107, no. 12, pp. 1852-1858, 2023, doi: <http://dx.doi.org/10.1136/bjo-2022-321472>.
- [12] Salvi, Raj Sunil, Shreyas Rajesh Labhsetwar, Piyush Arvind Kolte, Veerasai Subramaniam Venkatesh, and Alistair Michael Baretto. "Predictive analysis of diabetic retinopathy with transfer learning." *4th Biennial International Conference on Nascent Technologies in Engineering (ICNTE)*, IEEE, pp. 1-6, 2021, doi: <https://doi.org/10.1109/ICNTE51185.2021.9487789>.
- [13] Elsharkawy, Mohamed, Ahmed Sharafeldeen, Ahmed Soliman, Fahmi Khalifa, Mohammed Ghazal, Eman El-Daydamony, Ahmed Atwan, Harpal Singh Sandhu, and Ayman El-Baz. "A novel computer-aided diagnostic system for early detection of diabetic retinopathy using 3D-OCT higher-order spatial appearance model." *Diagnostics* 12, no. 2, p. 461, 2022, doi: <https://doi.org/10.3390/diagnostics12020461>.
- [14] Li, Feng, Yuguang Wang, Tianyi Xu, Lin Dong, Lei Yan, Minshan Jiang, Xuedian Zhang, Hong Jiang, Zhizheng Wu, and Haidong Zou. "Deep learning-based automated detection for diabetic retinopathy and diabetic macular oedema in retinal fundus photographs." *Eye* 36, no. 7, pp. 1433-1441, 2022, doi: <https://doi.org/10.1038/s41433-021-01552-8>.
- [15] Wang, Xiaoling, Zexuan Ji, Xiao Ma, Ziyue Zhang, Zuohuizi Yi, Hongmei Zheng, Wen Fan, and Changzheng Chen. "Automated Grading of Diabetic Retinopathy with Ultra - Widefield Fluorescein Angiography and Deep Learning." *Journal of Diabetes Research*, no. 1, pp. 2611250, 2021, doi: <https://doi.org/10.1155/2021/2611250>.
- [16] Gangwar, Akhilesh Kumar, and Vadmamani Ravi. "Diabetic retinopathy detection using transfer learning and deep learning." In *Evolution in Computational Intelligence: Frontiers in Intelligent Computing: Theory and Applications (FICTA 2020)*, Springer Singapore, Volume 1, pp. 679-689, 2021, doi: https://doi.org/10.1007/978-981-15-5788-0_64.
- [17] Chen, Ping-Nan, Chia-Chiang Lee, Chang-Min Liang, Shu-I. Pao, Ke-Hao Huang, and Ke-Feng Lin. "General deep learning model for detecting diabetic retinopathy." *BMC bioinformatics* 22, pp. 1-15, 2021, doi: <https://doi.org/10.1186/s12859-021-04005-x>.
- [18] Sikder, Niloy, Mehedi Masud, Anupam Kumar Bairagi, Abu Shamim Mohammad Arif, Abdullah-Al Nahid, and Hesham A. Alhumyani. "Severity classification of diabetic retinopathy using an ensemble learning algorithm through analyzing retinal images." *Symmetry* 13, no. 4, p. 670, 2021, doi: <https://doi.org/10.3390/sym13040670>.
- [19] Hagos, Misgina Tsighe, and Shri Kant. "Transfer learning based detection of diabetic retinopathy from small dataset." *arXiv*, pp.1905.07203, 2019, doi: <https://doi.org/10.48550/arXiv.1905.07203>.
- [20] Shi, Danli, Weiye Zhang, Shuang He, Yanxian Chen, Fan Song, Shunming Liu, Ruobing Wang, Yingfeng Zheng, and Mingguang He. "Translation of color fundus photography into fluorescein angiography using deep learning for enhanced diabetic retinopathy screening." *Ophthalmology science* 3, no. 4, 100401, 2023, doi: <https://doi.org/10.1016/j.xops.2023.100401>.
- [21] Menaouer, Brahami, Zoulikha Dermene, Nour El Houda Kebir, and Nada Matta. "Diabetic retinopathy classification using hybrid deep learning approach." *SN Computer Science* 3, no. 5, p. 357, 2022, doi: <https://doi.org/10.1007/s42979-022-01240-8>.
- [22] Alyoubi, Wejdan L., Maysoon F. Abulkhair, and Wafaa M. Shalash. "Diabetic retinopathy fundus image classification and lesions localization system using deep learning." *Sensors* 21, no. 11, p. 3704, 2021, doi: <https://doi.org/10.3390/s21113704>.
- [23] Abboud, Saif Hameed, Haza Nuzly Abdull Hamed, Mohd Shafry Mohd Rahim, Amjad Rehman, Tanzila Saba, and Saeed Ali Bahaj. "Hybrid retinal image enhancement algorithm for diabetic retinopathy diagnostic using deep learning model." *IEEE Access* 10, pp. 73079-73086, 2022, doi: <https://doi.org/10.1109/ACCESS.2022.3189374>.
- [24] Shaban, Mohamed, Zeliha Ogur, Ali Mahmoud, Andrew Switala, Ahmed Shalaby, Hadil Abu Khalifeh, Mohammed Ghazal et al. "A convolutional neural network for the screening and staging of diabetic retinopathy." *Plos one* 15, no. 6, p. e0233514, 2020, doi: <http://dx.doi.org/10.1371/journal.pone.0233514>.
- [25] Lin, Chun-Ling, and Kun-Chi Wu. "Development of revised ResNet-50 for diabetic retinopathy detection." *BMC bioinformatics* 24, no. 1, pp. 157, 2023, doi: <https://doi.org/10.1186/s12859-023-05293-1>.
- [26] Mushtaq, Gazala, and Farheen Siddiqui. "Detection of diabetic retinopathy using deep learning methodology." In *IOP conference series: materials science and engineering*, IOP Publishing, vol. 1070, no. 1, p. 012049, 2021, doi: [10.1088/1757-899X/1070/1/012049](https://doi.org/10.1088/1757-899X/1070/1/012049).
- [27] Raja Kumar, R., R. Pandian, T. Prem Jacob, A. Pravin, and P. Indumathi. "Detection of diabetic retinopathy using deep convolutional neural networks." In *Computational Vision and Bio-Inspired Computing: ICCVBIC*, Springer Singapore, pp. 415-430, 2021, doi: https://doi.org/10.1007/978-981-33-6862-0_34.

- [28] Liu, Hao, Keqiang Yue, Siyi Cheng, Chengming Pan, Jie Sun, and Wenjun Li. "Hybrid model structure for diabetic retinopathy classification." *Journal of Healthcare Engineering*, no. 1, pp. 8840174, 2020, doi: <https://doi.org/10.1155/2020/8840174>.
- [29] Bhardwaj, Charu, Shruti Jain, and Meenakshi Sood. "Transfer learning based robust automatic detection system for diabetic retinopathy grading." *Neural Computing and Applications* 33, no. 20, pp. 13999-14019, 2021, doi: <https://doi.org/10.1007/s00521-021-06042-2>.
- [30] <https://www.kaggle.com/competitions/aptos2019-blindness-detection>.
- [31] Wang, Juan, Yujing Bai, and Bin Xia. "Feasibility of diagnosing both severity and features of diabetic retinopathy in fundus photography." *IEEE access* 7, pp: 102589-102597, 2019, doi: <https://doi.org/10.1109/ACCESS.2019.2930941>.
- [32] Masood, Sarfaraz, Tarun Luthra, Himanshu Sundriyal, and Mumtaz Ahmed. "Identification of diabetic retinopathy in eye images using transfer learning." *International conference on computing, communication and automation (ICCCA)*, IEEE, pp. 1183-1187, 2017, doi: <https://doi.org/10.1016/j.procs.2020.03.400>.
- [33] Li, Xiaogang, Tiantian Pang, Biao Xiong, Weixiang Liu, Ping Liang, and Tianfu Wang. "Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification." *10th international congress on image and signal processing, biomedical engineering and informatics (CISP-BMEI)*, IEEE, pp. 1-11, 2017, doi: <https://doi.org/10.1109/CISP-BMEI.2017.8301998>.
- [34] Thota, Narayana Bhagirath, and Doshna Umma Reddy. "Improving the accuracy of diabetic retinopathy severity classification with transfer learning." *63rd International Midwest Symposium on Circuits and Systems (MWSCAS)*, IEEE, pp. 1003-1006, 2020, doi: <https://doi.org/10.1109/MWSCAS48704.2020.9184473>.
- [35] Kotiyal, Bina, and Heman Pathak. "Diabetic retinopathy binary image classification using PySpark." *International Journal of Mathematical, Engineering and Management Sciences* 7, no. 5, p. 624, 2022, doi: 10.33889/IJMEMS.2022.7.5.041.
- [36] Patel, Rupa, and Anita Chaware. "Transfer learning with fine-tuned MobileNetV2 for diabetic retinopathy." In *2020 international conference for emerging technology (INCET)*, IEEE, pp. 1-4, 2020, doi: <https://doi.org/10.1109/INCET49848.2020.9154014>.
- [37] Bora, Ashish, Siva Balasubramanian, Boris Babenko, Sunny Virmani, Subhashini Venugopalan, Akinori Mitani, Guilherme de Oliveira Marinho et al. "Predicting the risk of developing diabetic retinopathy using deep learning." *The Lancet Digital Health* 3, no. 1, pp: e10-e19, 2021, doi: [https://doi.org/10.1016/S2589-7500\(20\)30250-8](https://doi.org/10.1016/S2589-7500(20)30250-8).
- [38] Martínez-Murcia, Francisco Jesús, Andrés Ortiz-García, Javier Ramírez, Juan Manuel Górriz-Sáez, and Ricardo Cruz. "Deep Residual Transfer Learning for Automatic Diabetic Retinopathy Grading.", 2021, doi: <https://dx.doi.org/10.1016/j.neucom.2020.04.148>.
- [39] Sebt, Riad, Siham Zroug, Laid Kahloul, and Saber Benharzallah. "A deep learning approach for the diabetic retinopathy detection." In *The Proceedings of the International Conference on Smart City Applications*, Cham: Springer International Publishing, pp. 459-469, 2021, doi: https://doi.org/10.1007/978-3-030-94191-8_37.
- [40] Çınar, Gökalep, Kazım Kiliç, and Tuba Parlar. "A Deep Transfer Learning Framework for the Staging of Diabetic Retinopathy." *Journal of Scientific Reports-A* 051, pp. 106-119, 2022.
- [41] AbdelMaksoud, Eman, Sherif Barakat, and Mohammed Elmogy. "A computer-aided diagnosis system for detecting various diabetic retinopathy grades based on a hybrid deep learning technique." *Medical & Biological Engineering & Computing* 60, no. 7, pp: 2015-2038, 2022, doi: <https://doi.org/10.1007/s11517-022-02564-6>.
- [42] Al-Smadi, Mohammed, Mahmoud Hammad, Qanita Bani Baker, and A. Sa'ad. "A transfer learning with deep neural network approach for diabetic retinopathy classification." *International Journal of Electrical and Computer Engineering* 11, no. 4, p. 3492, 2021, doi: 10.11591/ijece.v11i4.pp3492-3501.
- [43] Ebrahimi, Behrouz, David Le, Mansour Abtahi, Albert K. Dadzie, Jennifer I. Lim, RV Paul Chan, and Xincheng Yao. "Optimizing the OCTA layer fusion option for deep learning classification of diabetic retinopathy." *Biomedical Optics Express* 14, no. 9, pp. 4713-4724, 2023, doi: <https://doi.org/10.1364/BOE.495999>.
- [44] Shakibania, Hossein, Sina Raoufi, Behnam Pourafkham, Hassan Khotanlou, and Muharram Mansoorizadeh. "Dual branch deep learning network for detection and stage grading of diabetic retinopathy." *Biomedical Signal Processing and Control* 93, p. 106168, 2024, doi: <https://doi.org/10.1016/j.bspc.2024.106168>.
- [45] Jabbar, Ayesha, Hannan Bin Liaqat, Aftab Akram, Muhammad Usman Sana, Irma Domínguez Azpíroz, Isabel De La Torre Díez, and Imran Ashraf. "A Lesion-Based Diabetic Retinopathy Detection Through Hybrid Deep Learning Model." *IEEE Access*, 2024, doi: <https://doi.org/10.1109/ACCESS.2024.3373467>.
- [46] Singh, Devendra, and Dinesh C. Dobhal. "A Deep Learning-based Transfer Learning Approach Fine-Tuned for Detecting Diabetic Retinopathy." *Procedia Computer Science* 233, pp: 444-453, 2024, doi: <https://doi.org/10.1016/j.procs.2024.03.234>.
- [47] Romero-Ora, Roberto, María Herrero-Tudela, María I. López, Roberto Hornero, and María García. "Attention-based deep learning framework for automatic fundus image processing to aid in diabetic retinopathy grading." *Computer Methods and Programs in Biomedicine* 249, p. 108160, 2024, doi: <https://doi.org/10.1016/j.cmpb.2024.108160>.
- [48] Mutawa, A. M., Khalid Al-Sabti, Seemant Raizada, and Sai Sruthi. "A Deep Learning Model for Detecting Diabetic Retinopathy Stages with Discrete Wavelet Transform." *Applied Sciences* 14, no. 11, p. 4428, 2024, doi: <https://doi.org/10.3390/app14114428>.
- [49] Chilukoti, Sai Venkatesh, Anthony S. Maida, and Xiali Hei. "Diabetic retinopathy detection using transfer learning from pre-trained convolutional neural network models." *IEEE J Biomed Heal Informatics* 20, pp: 1-10, 2022, doi: <https://dx.doi.org/10.36227/techrxiv.18515357.v1>.
- [50] Boruah, Swagata, Archit Dehloo, Prajul Gupta, Manas Ranjan Prusty, and A. Balasundaram. "Gaussian blur masked resnet2.0 architecture for diabetic retinopathy detection." *Computers, Materials and Continua* 75, no. 1, pp: 927-942, 2023, doi: <https://doi.org/10.32604/cmc.2023.035143>.

Authors' Profiles



Samia Akhtar is a student enrolled in M.S. Computer Science program at Virtual University of Pakistan, where she specializes in Software Engineering. Her research interests lie in the fields of Software Engineering and Deep Learning.



Dr. Shabib Aftab (Senior Member, IEEE) earned his Ph.D. in Computer Science from National College of Business Administration and Economics, Lahore. He previously attained his M.S. degree in Computer Science from COMSATS University, Lahore, and his M.Sc. degree in Information Technology from the Punjab University College of Information Technology, Lahore. Currently, he is serving as a Lecturer in the Department of Computer Science at the Virtual University of Pakistan. He has research interests in applied machine learning and software process improvement.



Dr. Munir Ahmad (Senior Member, IEEE) is a distinguished professional with over 16 years of experience. He holds a Ph.D. in computer science from the School of Computer Science, National College of Business Administration and Economics, and a Master of Computer Science degree from the Virtual University of Pakistan. As the Executive Director/CIO at United International Group, Lahore, Pakistan, he has excelled in data management and resource optimization within multinational organizations. Munir Ahmad is renowned for his extensive research in sentiment analysis, AI applications in healthcare and animal facial identification. His expertise lies in data mining, big data, and artificial intelligence.



Asma Akhtar is a researcher who has received her degree in M.S. Computer Science from Virtual University of Pakistan. Her research interests lie in the fields of Software Engineering and Machine learning.

How to cite this paper: Samia Akhtar, Shabib Aftab, Munir Ahmad, Asma Akhtar, "Diabetic Retinopathy Severity Grading Using Transfer Learning Techniques", International Journal of Engineering and Manufacturing (IJEM), Vol.14, No.6, pp. 41-53, 2024. DOI:10.5815/ijem.2024.06.04