

A Framework for Diabetic Retinopathy Detection using Transfer Learning and Data Fusion

Samia Akhtar

Department of Computer Science, Virtual University of Pakistan, Lahore, Pakistan
E-mail: samiaakhtar9898@gmail.com
ORCID iD: <https://orcid.org/0009-0004-4345-793X>

Shabib Aftab*

Department of Computer Science, Virtual University of Pakistan, Lahore, Pakistan
E-mail: shabib.aftab@vu.edu.pk
ORCID iD: <https://orcid.org/0000-0002-7662-1394>

*Corresponding author

Received: 26 June 2024; Revised: 05 September 2024; Accepted: 01 October 2024; Published: 08 December 2024

Abstract: Diabetic retinopathy stands as a significant concern for individuals managing diabetes. It is a severe eye condition that targets the delicate blood vessels within the retina. As it advances, it can inflict severe vision impairment or complete blindness in extreme cases. Regular eye examinations are vital for individuals with diabetes to detect abnormalities early. Detection of diabetic retinopathy is challenging and a time-consuming process, but deep learning and transfer learning techniques offer vital support by automating the process, providing accurate predictions, and simplifying diagnostic procedures for healthcare professionals. This study introduces a multi-classification framework for grading diabetic retinopathy into five classes using Transfer Learning and data fusion. The objective is to develop a robust, automated model for diabetic retinopathy detection to enhance the diagnostic process for healthcare professionals. We fused two distinct datasets, APTOS and IDRiD, which resulted in a total of 4178 fundus images. The merged dataset underwent preprocessing to enhance image quality and to remove unwanted regions, noise and artifacts from the fundus images. The pre-processed dataset is then resized and a balancing technique called SMOTE is applied to it due to uneven class distribution present among classes. To increase diversity and size of the dataset, data augmentation techniques including flipping, brightness adjustment and contrast adjustment are applied. The dataset is split into 80:10:10 ratios for training, validation, and testing. Two pre-trained models, EfficientNetB5 and DenseNet121, are fine-tuned and training parameters like batch size, number of epochs, learning rate etc. are adjusted. The results demonstrate the highest test accuracy of 96.06% is achieved by using EfficientNetB5 model followed by 91.40% test accuracy using DenseNet121 model. The performance of our best model i.e. EfficientNetB5, is compared with several state-of-the-art approaches, including DenseNet-169, Hybrid models and ResNet-50 where our model outperformed these methodologies in terms of test accuracy.

Index Terms: Diabetic Retinopathy, Data Fusion, Transfer Learning, Deep Learning, Augmentation.

1. Introduction

Diabetic retinopathy (DR) is a major health concern that affects millions of people worldwide [1]. It is one of the leading causes of blindness among adults of working age in many countries [2]. DR is a condition caused by long term diabetes mellitus. It occurs as a result of damage to the blood vessels in the retina which is the tissue responsible for vision, present at the back of the eye. This damage can lead to various complications like fluid or blood leakage from the blood vessels into the retina, abnormal growth of new blood vessels or even scarring of the retinal tissue. If not treated on time, these complications might worsen eyesight and result in complete blindness [3]. In order to help patients regain normal function in their retinas, early detection and treatment of this condition is essential [4].

Regular eye examinations are necessary for monitoring the progression of diabetic retinopathy and prevention of vision loss [5]. Diabetic retinopathy can be categorized into two primary stages: proliferative diabetic retinopathy (PDR) and non-proliferative diabetic retinopathy (NPDR) [6, 7]. Non-proliferative Diabetic Retinopathy is the early stage characterized by the presence of micro aneurysms (small bulges in blood vessels), hemorrhages (leakage of blood into the retina), and lipid exudates (yellowish deposits) [8]. Proliferative Diabetic Retinopathy is the advanced stage where

abnormal blood vessels grow on the surface of the retina, leading to severe vision impairment. This is known as Neovascularization. Early treatment assists in preventing the advancement of DR and protecting the vision. This intervention may take the form of medication, having surgery, lifestyle modifications, or laser treatment [9].

Artificial intelligence (AI), especially deep learning (DL) and transfer Learning (TL), has revolutionized the detection and classification of diabetic retinopathy (DR) [10]. Deep learning algorithms can identify intricate patterns and features straight from unprocessed data, including retinal scans, without the requirement for custom features. TL involves pre-training a DL model on a large dataset and then fine-tuning the model on a smaller dataset for classification [11]. The manual image processing approaches used in traditional DR detection methods frequently consisted of time-consuming processes including feature extraction and selection by individuals with expertise [12]. These techniques were challenging, prone to mistakes, and constrained by the experience level of the experts [13]. In comparison, DL and TL algorithms automate the feature extraction process and can learn numerous distinctive features directly from raw image data providing a superior performance. It also makes it possible to create DR detection systems that are more accurate and adaptable, which minimizes the need for human participation [14].

Our research proposes a multi-classification framework to detect and grade DR into five classes using Transfer Learning. The main contribution of this study is the use of data fusion, where we fused together 2 distinct DR Datasets and fine-tuned pre-trained models, which are used to grade DR into 5 classes. All DR Datasets have different image sizes, resolution and quality making it a challenge to attain a highest accuracy when datasets are merged. Training the models become more difficult by the noise and artifacts present in these images. Various pre-processing methods are employed in our study to address these problems in order to attain a higher of accuracy. APTOS and IDRiD are the two datasets that are combined together in this study to form a new dataset. Both of these datasets have DR graded into 5 classes: No DR, Mild, Moderate, Severe, and Proliferative DR. Pre-trained models used here are EfficientNetB5 and DenseNet121. We made use of techniques like CLAHE to improve the quality of images and SMOTE to tackle the class imbalance issues.

The goal of this research is to propose an automated and efficient system to grade diabetic retinopathy with a higher accuracy. Our methodology can help physicians and healthcare professionals by reducing the manual workload with less errors thus helping with the early detection and treatment of DR. Our aim is to reduce model over-fitting and model complexity by introducing various techniques such as fine-tuning, data augmentation and early stopping. The details of each technique used in our methodology is provided in Section 3. This research also aims to provide an efficient system by training models using fused dataset. By merging these datasets, we benefit from a broader and more diverse collection of fundus images, which improves the model's generalization capabilities. The increased data volume and diversity reduce the risk of over fitting which leads to a more robust and adaptable model. Furthermore, the fusion of datasets helps address limitations inherent in individual datasets, such as class imbalance and insufficient variation.

This research is organized into following sections: An over view of previous literature on DR detection is provided in Section 2 along with the methodologies and different classifiers used. Section 3 describes the proposed classification framework of this study in detail. Section 4 discusses the results of the proposed models using numerous performance evaluation metrics and also provides a comparison of these models with previous studies. Section 5 concludes the study with insights into future work.

2. Literature Review

Numerous researchers have devoted their efforts in experimenting with various techniques, algorithms, and methods to achieve the objectives of detecting and classifying diabetic retinopathy. Through their dedication, these researchers have made significant contributions in enhancing the detection of this disease.

In [15], researchers have proposed a multi-classification framework to detect and grade diabetic retinopathy into 5 classes using Transfer Learning. They used Inception-V3 model which is pre-trained using ImageNet. The dataset they used in their study is EyePACS from Kaggle. Fundus images were pre-processed using down sampling, local average color subtraction and cropping of the image border. Inception-V3 was used to for classification after pre-processing. Their proposed model showed a validation accuracy of 50% and the final test accuracy of 48.2%. Researchers of [16] have proposed a binary classification system to detect No-DR and Mild-DR using the concepts of transfer learning. They have used a pre-trained VGG19 model that is fine-tuned first before classification. Fundus image dataset from Kaggle and Messidor are the two datasets that were used in the research. These images were resized, cleaned by removing the dark images and CLAHE technique was applied on them to improve the contrast. Afterwards images were augmented using Augmentation techniques. VGG19 was fine-tuned using fully connected and dropout layers. This model achieved an accuracy of 93.89% and an F1 score of 90% in 20 epochs.

Authors in [17] have introduced a deep transfer learning framework to classify and detect the stages of DR automatically. A DR dataset named APTOS 2019 Blindness Detection is used by the researchers. They split their dataset to train and test where augmentation techniques were applied on the training set. Pre-processing steps were applied on both training and testing sets which was done in 5 steps. Authors used Gaussian Filtering, cropping, rescaling and application of CLAHE (Contrast Limited Adaptive Histogram Equalization) method. Then these pre-processed images were fed to three models: ResNet152, DenseNet201 and VGGNet19. Out of these proposed models, DenseNet201 achieved the highest test accuracy of 82.7% while the highest AUC value of 94.1% was achieved by

ResNet152. In [18], a binary classification system is proposed by the authors to classify diabetic retinopathy as DR/No DR. They have used Convolutional Neural Network (CNN) based transfer learning concepts for classification purpose. Two datasets are used, one is DR1 containing 1014 DR images and other one is Messidor with 1200 fundus images. For pre-processing, images are cropped first, and then resized. Four (4) pre-trained models have been fine-tuned: VggNet-vd-16, Vgg-s, AlexNet, and VggNet-vd-19. Three different experiments were carried out by fine-tuning all layers, fine-tuning in a layer wise manner, and extracting features using CNNs then classifying using SVM. Their proposed framework achieved a highest accuracy of 94.52% on DR1 dataset and 92.01% on Messidor.

Authors of [19] 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%. In [20] a multi-classification methodology is proposed using transfer learning models to grade diabetic retinopathy. The pre-trained transfer learning models used are EfficientNet, ResNet and VGG. Researchers used Kaggle competition dataset consisting of 35, 126 fundus images. This dataset is resized and split to training, validation and testing set prior classification. Different training parameters are experimented with along with different types of TL models and various performance metrics. Among all the models used, the Efficient_Net_b3_60 achieved the highest accuracy of 87 % followed by 0.85 quadratic weighted kappa (QWK), 84% F1-score, 85% Precision and 87% Recall.

A Binary classification framework is proposed in [21] using Inception-V3 transfer learning model to detect and classify DR into 2 classes. Authors chose a smaller version of EyePACS dataset available at Kaggle containing 2500 retinal images and divided them in healthy and unhealthy class. Images were pre-processed by getting cropped first and then resized to 300x300 dimensions. Afterwards Inception-V3 was trained using these pre-processed images for feature extraction and classification using SGD optimizer (Stochastic Gradient Descent) and softmax function. On testing the model, it gave a test accuracy of 90.9% and a test loss of 3.94%. Authors of [22] proposed deep learning algorithm for automatic detection of diabetic retinopathy. They used Inception-V3 model for classification purpose. They have worked on privately collected dataset from different clinics along with Messidor-2 dataset as the test set for performance evaluation of the trained model. Numerous pre-processing techniques were applied to the fundus images including, cropping, rescaling, CLAHE to improve contrast etc. The proposed method achieved a classification accuracy of 93.49% along with 96.93% sensitivity and 93.45% specificity. AUC achieved was 0.9905.

In contrast to these previous studies that relied on single datasets or limited image sources, the use of data fusion in our study aims to mitigate dataset-specific biases and improve robustness. We fused two benchmark datasets, APTOS and IDRiD, both containing 5 grades of DR. Moreover, this study enhances data quality through advanced preprocessing techniques, including CLAHE, and applies SMOTE to address class imbalance. The use of extensive data augmentation further diversifies the training set, contributing to improved model performance. In comparison to existing models, such as those using DenseNet-169 and various hybrid approaches, our framework demonstrates superior accuracy, achieving a test accuracy of 96.06% with EfficientNetB5. This surpasses the performance of many state-of-the-art methods, highlighting the effectiveness of data fusion and refined model tuning in our work.

3. Methodology

This research introduces a multi-classification framework using concepts of transfer learning to grade diabetic retinopathy into 5 classes. TL models are well known in image processing and disease detection [23]. Figure 1 presents the proposed classification framework outlining all steps involved in the methodology. A comprehensive explanation of the dataset utilized in this study and the corresponding steps implemented is provided below.

3.1. Used Datasets

There are two benchmark datasets used in this research: 1) APTOS (Asia Pacific Tele-Ophthalmology Society) 2019 Blindness Detection and 2) IDRiD (Indian Diabetic Retinopathy Image Dataset). The APTOS and IDRiD datasets were chosen for their comprehensive coverage of DR stages, diversity in imaging conditions and ability to provide a balanced and enriched training set. The description of each dataset is as under:

APTOS 2019 Blindness Detection is a widely utilized dataset available at Kaggle [24]. This collection comprises 3662 retinal images obtained from the Aravind Eye Hospital in India, captured through fundus photography under various image settings. This eye hospital played a role to identify and treat this illness in those individuals who reside in rural places where it is challenging to perform medical examinations. The severity levels/grades of diabetic retinopathy have been assigned to each image by the clinicians using a scale of 0 to 4, where 0 means No DR, 1 means Mild, 2 shows Moderate, 3 means Severe, and 4 represents Proliferative DR. A sample from each grade is shown in figure 2.

IDRiD (Indian Diabetic Retinopathy Image Dataset) is the first Indian population-representative database [25]. Furthermore, it is the unique dataset that includes both normal retinal features that have been labeled and diabetic retinopathy lesions. For every image, this dataset offers details on the severity of diabetic macular edema and diabetic

retinopathy. Since it can detect diabetic retinopathy early, this makes it ideal for developing and testing image processing techniques. This dataset was a part of the "Diabetic Retinopathy: Segmentation and Grading Challenge". The dataset is divided into three sections: 1) Disease Grading, 2) Localization, and 3) Segmentation. As the focus of this research is on grading of diabetic retinopathy, we only worked with Disease Grading using this dataset. For disease grading the dataset is divided into 413 images for training set and 103 images for testing set along with their labels giving a total of 516 fundus images. This dataset is also graded into 5 classes of No DR, Mild, Moderate, Severe and Proliferative DR. A sample from each grade of this dataset is shown in figure 2.

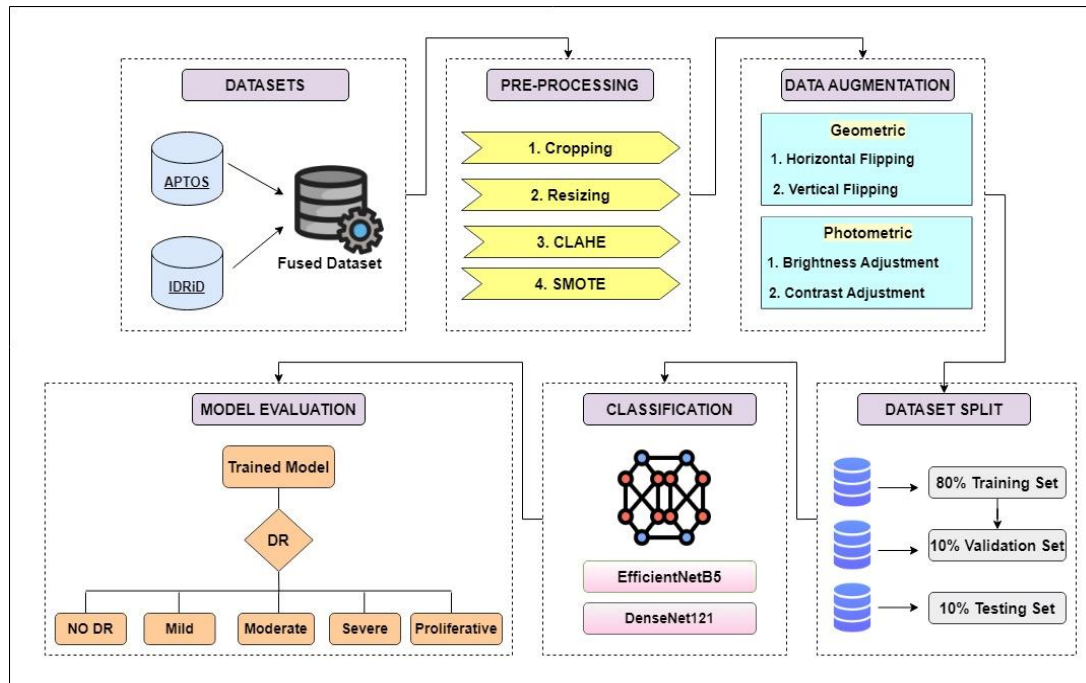


Fig.1. Proposed multi-classification framework

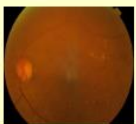
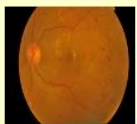
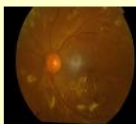
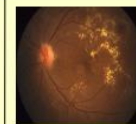
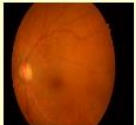
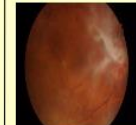
Datasets	No DR	Mild	Moderate	Severe	Proliferative
APTOS					
IDRiD					

Fig.2. A sample of fundus image from each grade of APTOS and IDRiD dataset

Both APTOS and IDRiD datasets have unbalanced distribution of images per grade/label. For this research the images from these datasets are fused together to increase the size of dataset. Both datasets have 5 grades/classes. The images of grade 0 in APTOS dataset are combined with the images of grade 0 in IDRiD and all the images of grades 1, 2, 3 and 4 are merged the same way. This way we combined all the images from individual datasets and formed a new one containing all the fundus images from both APTOS and IDRiD dataset. After fusion, the new number of fundus images becomes 4178. The image distribution per grade and the total number of images for APTOS, IDRiD and fused dataset is given in Table 1. The reason for fusing these two datasets in our study is that fusion mitigates the risk of over fitting to the specifics of a single dataset. It provides a more comprehensive training set, which improves the model's performance and its ability to generalize across different scenarios. The IDRiD dataset complements the APTOS dataset by providing additional examples of DR stages and features not covered extensively in APTOS. This diversity helps in improving the model's ability to detect and classify DR more accurately.

Table 1. Image distribution in individual and fused dataset

Grade	Number of Images		
	APTOS	IDRiD	Fused Dataset
No DR (Grade 0)	1805	168	1973
Mild (Grade 1)	370	25	395
Moderate (Grade 2)	999	168	1167
Severe (Grade 3)	193	93	286
Proliferative DR (Grade 4)	295	62	357
Total	3662	516	4178

3.2. Data Pre-processing

Data preprocessing involves various techniques to prepare raw image data for model training. These techniques enhance the quality of the data, eliminate the noise, standardize the format, and help in achieving optimal accuracy for classification. Fundus image datasets available to the public vary in resolutions and compression formats. The transfer learning models require clean, enhanced and quality data for successful grading of DR which makes pre-processing an essential step to be carried out before classification [26].

In this research, pre-processing is done in four (4) steps: 1) Cropping, 2) Resizing, 3) CLAHE and 4) SMOTE. Dataset analysis was done first. As we have merged two different datasets, the combined images are from different clinics and hospitals, photographed from different cameras and angles. Some images have noise and unwanted, extra black region as compared to the others. Similarly, some images are too dull with faded details of retinal features while others have a low contrast to see blood vessels and other abnormalities clearly.

First step of pre-processing adopted in our proposed methodology is cropping. Unwanted background can be seen in some of the images of the dataset. If such dataset with variation of backgrounds is given to a model for classification, it may lead to performance and generalization issues. That will lower the accuracy of the model. Thus, the images were cropped to the required size from where the eye starts in order to remove the unwanted black background. We used the cv2.boundingRect function from the OpenCV library to calculate the bounding box of the largest contour, which was then used to crop the image.

Then for second step, these cropped images were resized to bring all the images to a same dimension of 180 x 180 x 3. Here, first 2 numbers represent pixels in height and width whereas 3 represents number of color channels in the image i.e. red, green, and blue (RGB). We standardized the image dimensions to ensure uniform input size for the model. For this purpose, cv2.resize function was used.

In third step, Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied to the resized images. This technique improved the contrast and quality of the images by adding more color to the faded dull regions, showing finer details and improving the overall visibility of features in both bright and dark regions of the images. We used python's cv2.createCLAHE function with a clip limit of 4 and tileGridSize of (8, 8). Figure 3 shows the first three steps of pre-processing



Fig.3. A sample of resultant image after first three steps of pre-processing

After all these steps, the final step is to balance the dataset using an over-sampling technique known as SMOTE (Synthetic Minority Over-sampling Technique). The distribution of images across classes in the merged dataset is uneven, as shown in Table 1. It's a crucial step to address this class imbalance to prevent biasness in the transfer learning models, particularly towards the class with the majority number of samples and to improve the overall performance. The objective is to augment the number of samples only in the minority classes while maintaining the majority class unchanged. Among our dataset, Grade 0 "NO DR" contains the highest number of images. Our approach involves increasing the sample count in the remaining classes to a randomized figure close to the number of images in class 0. Table 2 provides the new number of images after applying SMOTE. As this technique generates synthetic samples by interpolating between existing samples, it can lead to the creation of samples that may not fully capture the variability present in real-world data, potentially introducing artifacts or unrealistic features that could affect model performance. Thus, careful inspection is required for model training.

Table 2. Number of image distribution per class after SMOTE

	No DR (Grade 0)	Mild (Grade 1)	Moderate (Grade 2)	Severe (Grade 3)	Proliferative DR (Grade 4)	Total
Original Images	1973	395	1167	286	357	4178
Balanced Images	1973	1880	1650	1750	1650	8903

3.3. Data Augmentation

After Pre-processing, the images are augmented using data augmentation techniques. The aim is to increase the size of dataset as a larger dataset allows the model to capture a broader range of patterns and variations in the data, which can lead to better generalization. Data augmentation is the process of expanding the size of the dataset by applying a variety of transformations to the existing data samples [27]. These transformations mainly include techniques like rotation, translation, scaling, flipping, cropping and more.

Data augmentation techniques enhance model performance by introducing variability and diversity into the training dataset, which helps the model generalize better to unseen data. The objective of data augmentation is to increase the diversity of the dataset and bring variation. There are many forms of augmentation but we used a mix of geometric and photometric techniques. Geometric augmentation involves applying transformations that alter the spatial properties of the image, such as its size, orientation, and position. Photometric augmentation involves altering the pixel values or color properties of the image.

The geometric augmentation techniques used in this research are horizontal and vertical flipping while photometric ones are brightness adjustment and contrast adjustment. These techniques are randomly applied on the dataset. Augmentation is done in such a way that it increases the total number of images twice. The total number of images after balancing using SMOTE was 8903. Thus, after augmentation it became 17,806 images. Data augmentation was applied using the ImageDataGenerator class from Keras. Table 3 reflects the used techniques in this research. Figure 4 shows the applied augmentation techniques.

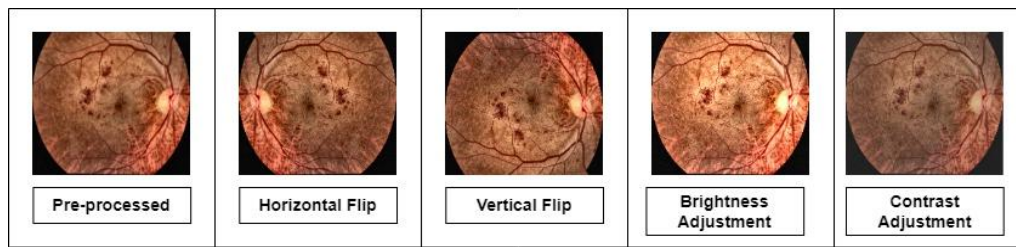


Fig.4. A sample of each technique applied for augmentation

Table 3. Augmentation techniques used and their corresponding values

Technique	Value
Flip	Vertical
Flip	Horizontal
Brightness Adjustment	10
Contrast Adjustment	10

3.4. Dataset Split

Images are split to training and testing sets before model training. After augmentation, we split the dataset to an 80:10:10 as in 80% for training set, 10% allocated for validation set and 10% for testing set. Training set will be used to train the transfer learning models and then these models will be used to evaluate the test set. Validation set is used to evaluate the performance of the TL models during training process.

3.5. Classification using Transfer Learning

Transfer Learning models are pre-trained on ImageNet which is a large-scale dataset, containing millions of labeled images across thousand categories [28]. Numerous deep learning architectures have undergone training using the ImageNet dataset. Among the most popular ones are models such as AlexNet, VGG, GoogLeNet, ResNet, and more. These models have excelled in the ImageNet challenge, showcasing their capability to accurately identify objects within images with remarkable precision. 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 two transfer learning models for this study are EfficientNetB5 and DenseNet121. These models are fine-tuned before training. The first and last layers are modified for both models with a few additional layers. The input layer was altered and the input shape was set to 180 x 180 x 3. The additional layers added after the base TL layer include a Global Average Pooling (GAP) layer, dropout layers, flatten layer, fully connected (FC) layers and an

output layer. Activation function used in hidden layers is ReLU (Rectified Linear Unit). Softmax activation function is used in the output layer as the proposed framework is for multi-classification. Figure 5 provides the detailed fine-tuned architecture used for these models.

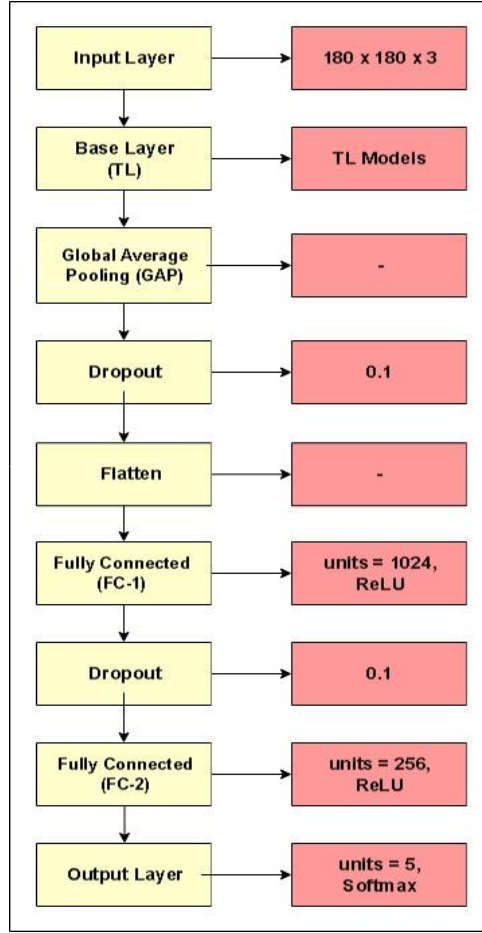


Fig.5. Layer wise architecture of fine-tuned models

3.6. Performance Evaluation Metrics

Evaluation metrics are used to assess the performance of transfer learning models. These metrics provide insights into how well a model is performing and help in comparing different models or tuning their parameters. Each data instance can be categorized into one of four potential outcomes, True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) [29]. A true positive occurs when the model correctly predicts a positive instance whereas true negative occurs when the model correctly predicts a negative instance. False positive implies that the model has incorrectly predicted a positive instance. False negative means the opposite of false positive i.e. when a model incorrectly predicts a negative instance. The metrics used to evaluate the training and testing set for this research are as follow

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

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

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

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

$$\text{Positive Prediction Value (PPV)} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{Negative Prediction Value (NPV)} = \frac{TN}{TN + FN} \quad (6)$$

$$\text{False Positive Ratio (FPR)} = 1 - \text{Specificity} \quad (7)$$

$$\text{False Negative Ratio (FNR)} = 1 - \text{Sensitivity} \quad (8)$$

$$\text{Likelihood Ratio Positive (LR+)} = \frac{\text{Sensitivity}}{(1 - \text{Specificity})} \quad (9)$$

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

4. Results and Discussion

When the dataset is prepared and pre-processed, it is split to a ratio of 80:10:10 and training set is used to train the TL models. The total number of images after augmentation was 17,806. After the split, the training set has a total of 12,820 images (80%), validation set has 3205 images (10%) and testing set has 1781 images (10%).

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. EfficientNetB5 and DenseNet121 are the models employed in this research. Batch size for training was set to 32 and total number of epochs was 40. The Adam algorithm was used as an optimizer along with a learning rate of 0.001.

Initially we experimented with various learning rates (0.001, 0.0001, 0.01) and optimizers such as RMSProp, SGD (Stochastic Gradient Descent) and Adam, then we picked the one which gave the best performance i.e. Adam Optimizer with a learning rate of 0.001. Similarly, we carried out our experimentations with and without the regularization techniques and further experimented with other hyper-parameters like batch size (64, 32) and No. of epochs (40, 60, 80). Without the regularization techniques, the model would perform exceptionally well only on the training set but would not give much accurate results on validation or test set because it would essentially memorize the training data rather than generalizing from it. So, to address the issue of over fitting, this study employs regularization techniques such as EarlyStopping and ReduceLROnPlateau methods.

These methods are configured to halt training when the model achieves its best performance before overfitting occurs. EarlyStopping is utilized to monitor the validation accuracy. It halts training if the validation accuracy remains stable or increases for five (5) consecutive epochs. The ReduceLROnPlateau method is employed to adjust the learning rate. If the loss value does not change for 3 successive epochs, it decreases the learning rate by a factor of 0.2. .. In conclusions we chose all those parameters which boosted the performance of our models and helped with achieving higher accuracy in correctly detecting the DR stages. The training parameters used for this study are given in Table 4.

Table 4. Hyper-parameters for training the TL models

Training Parameters	
Input Shape	180 x 180 x 3
Optimizer	Adam
Learning Rate	0.001
No. of Epochs	40
Batch Size	32
Loss Function	Categorical cross entropy

On training EfficientNetB5, it gave an accuracy of 99.08% followed by 99.18% precision, 98.96% recall and 99.07% F1-score. On evaluating this model on the test set, it achieved the best accuracy of 96.06% followed by 96.22% precision, 95.84% recall and 96.05% F1-score. Confusion matrix of both training and testing set for EfficientNetB5 is given in figure 6.

Training DenseNet121 gave an accuracy of 96.42%, a precision, recall and F1-score of 96.58%, 96.16% and 96.37% respectively. When it was evaluated using test set, it achieved an accuracy of 91.40% followed by 91.85% precision, 91.24% recall and 91.56% F1-score. Confusion matrix of both training and testing set for this model is given in figure 7. The accuracy achieved by this model is a little less than EfficientNetB5.

The evaluation metrics for both of the TL models are given in Table 5. Training and testing, both metrics are evaluated for each model. It can be seen from the results given in the table that EfficientNetB5 performed well as compared to DenseNet121. It gave a higher training and testing accuracy along with other metrics. DenseNet121 model also performed well in classifying the DR into 5 classes.



Fig.6. Confusion matrix for EfficientNetB5 model

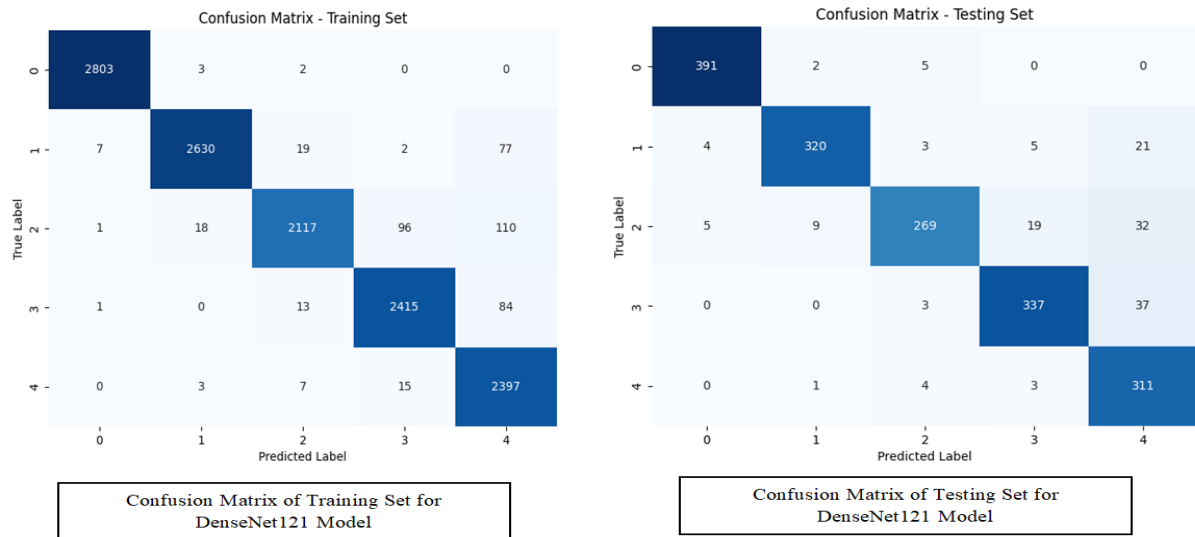


Fig.7. Confusion matrix for DenseNet121 model

Table 5. Performance Evaluation of EfficientNetB5 and DenseNet121 using various measures

TL Models		Accuracy	Misclassification Rate	Specificity	Sensitivity	Positive Prediction Value	Negative Prediction Value	False Positive Ratio	False Negative Ratio	Likelihood Ratio Positive	Likelihood Ratio Negative
EfficientNetB5	Training	0.9908	0.0091	0.9977	0.9905	0.9907	0.9977	0.0022	0.0094	450.60	0.0094
	Testing	0.9606	0.0393	0.9902	0.9596	0.9597	0.9902	0.0097	0.0403	118.83	0.0408
DenseNet121	Training	0.9642	0.0357	0.9911	0.9628	0.9644	0.9912	0.0088	0.0371	374.23	0.0373
	Testing	0.9140	0.0859	0.9787	0.9126	0.9179	0.9789	0.0212	0.0873	79.770	0.0887

Accuracy and Loss of training and validation for these models is presented in figure 8 in the form of a graph. It shows both of these models were trained well with minimum validation loss. EfficientNetB5 achieved the validation accuracy of 96.41% with 0.1454 validation loss. Validation accuracy for DenseNet121 is 89.67% with 0.4238 validation loss. This can be visualized from the graphs.

The best results of this research are compared with the previous studies in terms of test accuracy, best performance and multi-classification. Many researchers have worked on grading DR. Table 6 provides a comparison of the previous methodologies and our framework. Table shows that the researchers, Liu et al and Menaouer et al, have used hybrid models after combining deep learning models to grade diabetic retinopathy and a model fusion is done by Alyoubi et al after combining two deep learning architectures. Shaban et al and Raja Kumar et al have proposed their own deep learning convolutional neural network architectures in their research. Remaining authors have worked with pre-trained models (transfer learning) to achieve a better accuracy. All these researchers have worked with different datasets such

as APTOS, EyePACS, Messidor, DeepDR etc. for training and testing their models along with distinct pre-processing techniques. Our proposed model EfficientNetB5 showed the best accuracy as compared to these studies with the proposed pre-processing techniques and data fusion. Although these previous studies worked with more than one dataset, but they used individual datasets for training/testing and evaluated results for these datasets individual rather than using the concept of data fusion. Some of these techniques lack the use of regularization techniques or proper pre-processing techniques due to which their models were unable to achieve higher accuracy. The details of the comparison are provided in Table 6.

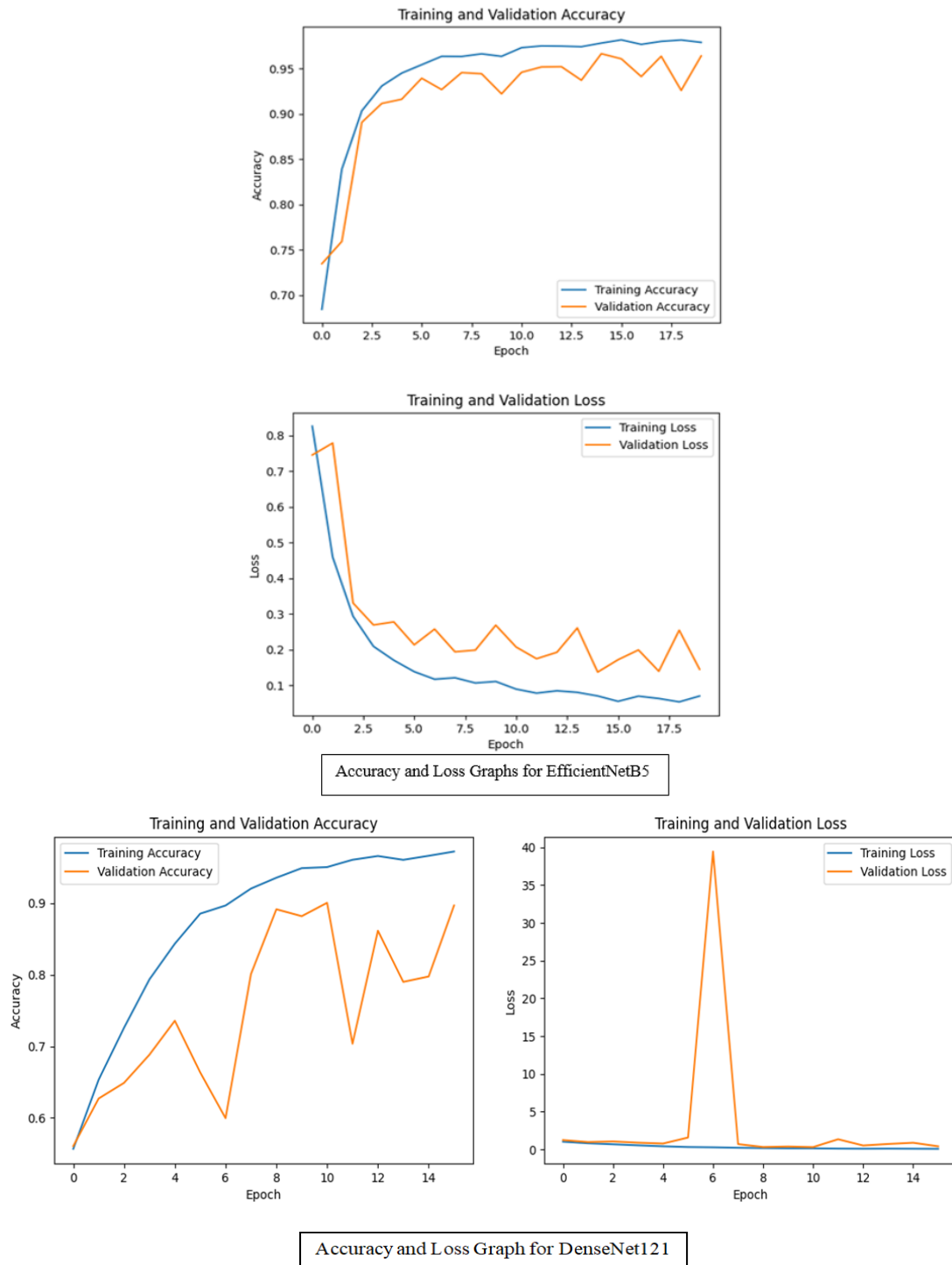


Fig.8. Graphs for training and validation accuracy + training and validation loss for both TL models

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

Reference	Year	Datasets Used	Methodology	Accuracy
Mushtaq et al [30]	2021	DR detection 2015 and APTOS2019	DenseNet-169	90%
Shaban et al [31]	2020	APTOS2019	CNN Model	88-89%
Menaouer et al [32]	2022	APTOS2019	Hybrid models (CNN, VGG16 and VGG19)	90.60%.
Lin et al [33]	2023	EyePACS	Revised ResNet-50	74.32%
Alyoubi et al [34]	2021	DDR and APTOS2019	Model Fusion (CNN512 + YOLOv3)	89%
Raja Kumar et al [35]	2021	Kaggle Dataset	CNN Model	94.44%
Abbood et al [36]	2022	Messidor	ResNet-50	93.6%
Current Study	2024	(APTOS2019 + IDRiD) fused	EfficientNetB5	96.06%

The analysis of our results highlights the superior performance of the EfficientNetB5 model compared to previous studies, with a notable test accuracy of 96.06%. This superior performance can be attributed to several key factors. Unlike earlier studies that relied on single datasets or limited model combinations, our approach utilized a fusion of two distinct datasets (APTOS2019 and IDRiD), which enhanced the diversity and comprehensiveness of the training data. This data fusion approach allowed our model to learn from a broader range of variations in diabetic retinopathy, thus improving its generalization capabilities. The EfficientNetB5 model's architecture, with its efficient use of depthwise separable convolutions, further enabled it to capture intricate features of the retinal images more effectively. In contrast, methods like hybrid models or individual deep learning architectures faced limitations due to their reliance on specific datasets or less robust preprocessing. Our framework addressed these issues through rigorous data fusion and advanced preprocessing techniques.

5. Conclusions and Future Works

This research introduced a robust multi-classification framework for detecting and grading diabetic retinopathy into 5 classes using data fusion and transfer learning. Diabetic Retinopathy is a complication of diabetes that affects the blood vessels in the retina of the eye. This damage can lead to vision problems or even blindness if left untreated. For this study, two datasets, APTOS 2019 and IDRiD, are fused together for detection and classification of diabetic retinopathy. By merging the datasets, we addressed the challenge of dataset variability and class imbalance, significantly enhancing model performance.

Our approach utilized advanced techniques, including CLAHE for contrast enhancement, SMOTE for balancing class distribution, and extensive data augmentation to increase dataset diversity. Images were cropped to remove unwanted regions and CLAHE was utilized to enhance the contrast. Dataset was originally unbalanced which was balanced using SMOTE method. Augmentation techniques including photometric and geometric techniques were applied to increase diversity. EfficientNetB5 and DenseNet121 were used for classification after fine-tuning. EfficientNetB5 model achieved a training accuracy of 99.08% and scored the highest test accuracy of 96.06% whereas DenseNet121 model achieved 96.42% accuracy on training set and 91.40% on test set. Many researchers have worked to detect and grade DR using numerous methodologies and algorithms. The best methodology of this research i.e. EfficientNetB5 model was compared with the previous studies and it performed well in comparison. This framework not only improves diagnostic accuracy but also reduces manual workload, supporting early detection and treatment of diabetic retinopathy.

Future work could focus on integrating more advanced deep learning architectures, such as more sophisticated convolutional neural networks, and exploring the benefits of fusing other publically available distinct DR Datasets. Additionally, experimenting with ensemble techniques could further refine and enhance the system's accuracy

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] 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>.
- [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.
- [6] Ahmad, Munir, Majed Alfayad, Shabib Aftab, Muhammad Adnan Khan, Areej Fatima, Bilal Shoaib, Mohammad Sh Daoud, and Nough Sabri Elmitwal. "Data and Machine Learning Fusion Architecture for Cardiovascular Disease Prediction." *Computers, Materials & Continua* 69, no. 2, 2021. DOI: <https://doi.org/10.32604/cmc.2021.019013>.
- [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 Houbby, 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 Sharafeldein, 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, pp. 461, 2022. DOI: <https://doi.org/10.3390/diagnostics12020461>.
- [14] 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>.
- [15] 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>.
- [16] Ebin, P. M., and P. Ranjana. "A Variant Binary Classification Model for No-DR Mild-DR Detection using CLAHE Images with Transfer Learning." *International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS)*, IEEE, pp. 1-5, 2022. DOI: <https://doi.org/10.1109/IC3SIS54991.2022.9885665>.
- [17] Çinarer, 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.
- [18] 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>.
- [19] 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>.
- [20] Chilukoti, Sai Venkatesh, Anthony S. Maida, and Xiali Hei. "Diabetic retinopathy detection using transfer learning from pre-trained convolutional neural network models." *J Biomed Heal Informatics* 20, IEEE, pp. 1-10, 2022. DOI: <https://doi.org/10.36227/techrxiv.18515357.v1>.
- [21] 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>.
- [22] Li, Feng, Zheng Liu, Hua Chen, Minshan Jiang, Xuedian Zhang, and Zhizheng Wu. "Automatic detection of diabetic retinopathy in retinal fundus photographs based on deep learning algorithm." *Translational vision science & technology* 8, no. 6, pp. 4-4, 2019. DOI: <https://doi.org/10.1167/tvst.8.6.4>.
- [23] 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>.
- [24] <https://www.kaggle.com/competitions/aptos2019-blindness-detection>.
- [25] <https://ieee-dataport.org/open-access/indian-diabetic-retinopathy-image-dataset-idrid>
- [26] 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>.
- [27] Shi, Danli, Weiyi 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>.
- [28] Luo, Xiaoling, Wei Wang, Yong Xu, Zhihui Lai, Xiaopeng Jin, Bob Zhang, and David Zhang. "A deep convolutional neural network for diabetic retinopathy detection via mining local and long-range dependence." *CAAI Transactions on Intelligence Technology* 9, no. 1, pp. 153-166, 2024. DOI: <https://doi.org/10.1049/cit.2.12155>.
- [29] 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>.

- [30] 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.
- [31] 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>.
- [32] Menaouer, Brahami, Zoulikha Dermane, 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>.
- [33] 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>.
- [34] 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>.
- [35] 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.
- [36] 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>.

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 received 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.

How to cite this paper: Samia Akhtar, Shabib Aftab, "A Framework for Diabetic Retinopathy Detection using Transfer Learning and Data Fusion", International Journal of Information Technology and Computer Science(IJITCS), Vol.16, No.6, pp.61-73, 2024. DOI:10.5815/ijitcs.2024.06.05