Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 3, Issue 8, 2025

PERFORMANCE EVALUATION OF DENSENET121 AND RESNET50 WITH CLAHE-BASED PREPROCESSING FOR DIABETIC RETINOPATHY DETECTION

Zaryab Shafiq[1], Abdul Manan[*2], Muazzam Ali[3], Kanwal Amjad[4], Ufaq Zahra[5]

[1,*2,3,4]*Department of Basic Sciences Superior University Lahore*

abdul.manan@superior.edu.pk

DOI: https://doi.org/10.5281/zenodo.17084847

Keywords
Diabetic Retinopathy, Image Preprocessing, CLAHE, DenseNet121, ResNet50, Deep Learning, Medical Imaging, Transfer Learning

Article History
Received: 29 May, 2025
Accepted: 10 August, 2025
Published: 29 August, 2025

Corresponding Author: *
Abdul Manan

Abstract
Background: *Diabetic retinopathy (DR) had been identified as one of the main blindness-causing eye disorders and its early diagnosis was the key to avoiding blindness.*
Objective: *The goal of the study was to determine whether the application of five preprocessing techniques CLAHE + Normalization, Vessel Masking + CLAHE, Vessel Cropping + CLAHE, Edge Sharpening + CLAHE, Gamma Correction + CLAHE can be used to improve the performance of deep learning models of DR.*
Methods: *A reusable collection of retinal images was cleaned via the five enhancement techniques. The classification was executed through two robust CNN models, i.e. DenseNet121 and ResNet50. The performance of the models was checked in terms of accuracy, precision, recall, F1- score, and confusion matrices.*
Results: *The experiments showed that preprocessing proved to be highly influenced by the accuracy of the model. The combination of CLAHE + Vessel Masking on the dataset preprocessed using DenseNet121 has shown the best performance of all tested combinations, attaining an overall accuracy of 98.9% and being superior to other preprocessing algorithms.*
Conclusion: *The results revealed that the performance of CNN models in the process of detecting diabetic retinopathy was highly enhanced by implementing apt preprocessing. These findings pointed out the need to incorporate enhancement strategies into medical imaging workflows to develop feasible automatic screening and assist clinical decisions.*

INTRODUCTION

Diabetic retinopathy (DR) is one of the major causes of preventable blindness in the world; it is a progressive microvasculature disease of diabetes mellitus. It takes place when sustained hyperglycemia harms the retinal blood vessels and results in microaneurysms, bleeding, and, at last, vision-threatening conditions like proliferative retinopathy and diabetic macular edema [1, 2]. The number of cases assigned to the globe due to the impact of DR is still increasing in line with the prevalence of diabetes cases. In the current systematic estimates, close to 22.3 per cent of diabetes patients have DR, which translates to over 100 million adults in 2020, and it is projected to rise to over 160 million by 2045

[3, 4]. The prevalence of the disease is high in Africa (35.9%) and North America (33.3%), and it particularly affects those in the working age with massive cost implications in terms of social security [5]. With such concerns in the prevalence of DR, the issue has become a critical focus in the domain of overall health and therefore justifies the timely screening methods in ensuring that preventable blindness is eliminated.

Although early detection is of clinical interest, manual screening systems are associated with several limitations. Conventional methods include sequential annual retina inspection or fundus photograph grading by specialized ophthalmologists or graders, both of which are time and resource consuming [6, 7]. Accessibility gaps in highly skilled professionals and quality imaging systems are also very poor in regions with less healthcare infrastructure, including rural Asia, Africa, and Latin America, which further add to the disparities in eye care [8]. In addition, human graders are prone to variability, fatigue, and diagnostic errors in well-established screening programs, and this may negatively affect reliability [9,10,11]. The resulting burdens and the increasing patient pool among the diabetic population underline the necessity of automated alternatives to manual screening procedures as these processes cannot remain viable in the long run.

Artificial intelligence (AI), specifically deep learning, has demonstrated its huge potential in surmounting these drawbacks. The combination of AIs enables them to look at retinal fundus images with speed, consistency, and precision that can often match that of human experts [12]. By finding minute retinal defects beyond the range of a naked eye, AI-based tools do not only help detect disease in early stages but also enable a rapid screening system at scale to reach large populations in high-risk groups. Highly sensitive and specific detection of DR by deep learning models trained on large data is possible with lower cost and effort of manual screening [13]. These automated systems show promise in tele-ophthalmology networks, with portable fundus cameras transmitting data to AI solutions so that on-site analysis is synonymous with remote referrals [14]. Such inventions are especially needed in low-resource places, as early detection can drastically fall the rates of irreversible blindness.

Among the deep approaches, convolutional neural networks (CNNs) have been especially effective in medical image classification. Unlike traditional feature-based methods, which require handcrafted descriptors and expert knowledge, CNNs automatically learn hierarchical representations of images, enabling them to capture both local and global features relevant for disease grading [15]. This adaptability has allowed CNNs to outperform older machine learning techniques such as support vector machines, decision trees, or logistic regression in DR detection tasks [16]. Furthermore, CNNs have proven highly scalable, allowing deployment across diverse datasets and imaging conditions, which is critical for real-world screening applications.

DR detection systems have also been reinforced by recent achievements in transfer learning. Other pretrained networks that have been adapted to retinal imaging such as DenseNet121 and ResNet50 modules were initially trained on massive image datasets in a large-scale image classification task [17]. The DenseNet121 model incorporates the dense connectivity feature to provide characteristics reuse and better flow of the gradient, which results in higher performance using the fewer parameters. ResNet 50, is designed with the residual learning technique that solves the vanishing gradient issue and brings fast training without loss in accuracy [13, 15]. With enhanced preprocessing techniques, including but not limited to Contrast Limited Adaptive Histogram Equalization (CLAHE) and vessel masking, these models perform as well as 98.9% accurate on classification tasks accordingly noted to demonstrate suitability in clinical application of the models [16]. The increased evidence shows that combination of transfer learning and optimized strategies of preprocessing has massive potential to remodel automated DR screening and help to ensure that patients are diagnosed in time and burden of diabetes blindness in the world can be reduced [17].

The use of machine learning models to forecast diabetic retinopathy (DR) based on biochemical and clinical data has been investigated in a number of studies. Notably, XGBoost and random forest have outperformed the others, obtaining strong AUC values of 0.991 and 0.989 along with high accuracies of 95.67% and 94.67%, respectively. Lower accuracies and AUC scores have been obtained by using neural networks and logistic regression. Important predictors for DR prediction were found to include serum creatinine, HbA1c, and 24-hour urinary microalbumin. These results highlight the potential of machine learning models, specifically XGBoost and random forest, for improved diabetes management and early DR detection [18].

When compared to more conventional models like CNNs and ResNet50, ensemble learning—specifically, EfficientNet models—has demonstrated encouraging outcomes in the detection of diabetic retinopathy (D.R.), with notable gains in accuracy (95%) and recall (97%). Even though the approach performed well on datasets such as Kaggle's DR dataset, more investigation and practical validation are required to guarantee its generalizability and suitability in a range of clinical contexts. This algorithm shows how machine learning algorithms can be used to augment D.R. identification in low resource settings [19].

This study exploits the Debrecen dataset to examine various machine learning algorithms, and in particular deep neural networks (DNN), in the prediction of diabetic retinopathy (DR). The results indicate that DNN is more effective in relationship to accuracy, precision, recall, sensitivity, and specificity values compared to the other traditional methods such as SVM, Decision Tree, KNN, and Naive Bayes. Also, use of Principal Component Analysis (PCA) of the role of dimensionality reduction contributes to a significantly better performance of the model. The performance is further boosted by the introduction of the Grey Wolf Optimization (GWO) algorithm as a hyper parameter tuning tool which displays an impressive accuracy of 97.3. A potential solution to early detection and better patient outcomes, the work shows the utility of DNN-based models with PCA and GWO that can be reliably and effectively predict DR [20].

This study explores the preprocessing approaches in a systematic manner and their firsthand impact on automatic detection of diabetic retinopathy. Contrary to other reports that might decide to employ a certain method of enhancement, or a limited experimental setting, this paper compares objectively five CLAHE-based image preprocessing algorithms in their efforts to enhance the contrast and help in elucidating soft elements of the retinal that may aid in early diagnosis. Moreover, the research critically analyses these preprocessing methods on two current CNN models, DenseNet121 and ResNet50, which, in turn, makes it possible to conduct a critical cross-model analysis and establish both the strengths and weaknesses of each method. In addition to raw accuracy scores, the research focuses on specific performance rates and error analyses, providing a very close picture of how different approaches perform at different stages of the disease. The breadth (varied preprocessing methods), depth (testing on powerful CNN backbones) and clarity (breadth and depth of error explanations) place the work in the set of unique contributions to the literature that appeal to technical rigor as well as real-world application in diabetic retinopathy screening.

2.0 Research Methodology

2.1 Dataset

The publicly available Kaggle Diabetic Retinopathy dataset is used in the current study and is well employed in general medical imaging research. This dataset is composed of retinal fundus photographs which are processed with Gaussian smoothing and scaled to 224 224 pixels, guaranteeing unifcited images dimension to deep learning models. The filtering stage is useful in reducing noise and normalizing intensity variation and enhances the effectiveness of the next stage of the preprocessing and feature extraction. The data has been divided into five different classes denoted as the graded severity of diabetic retinopathy (DR). These categories are: No_DR, which means no signs of retinopathy; Mild, with some micro aneurysm being present; Moderate, which is highly noticeable hemorrhage and exudation; Severe, akin to Moderate one, except it has many abnormal lesions, and Proliferate_DR which represents the most serious form that is the presence of neovascularization in addition to the high risk of vision loss. This dichotomous grading has a clinical value and is very similar to real grading in ophthalmology. The Kaggle dataset shown in Figure 1 offers an excellent option of evaluating deep learning methods as it has a balanced and varied set of the disease spectrum to be assessed. It helps the researchers determine the efficiency of preprocessing and sophisticated convolutional neural networks at different difficulty levels of the diagnosis.

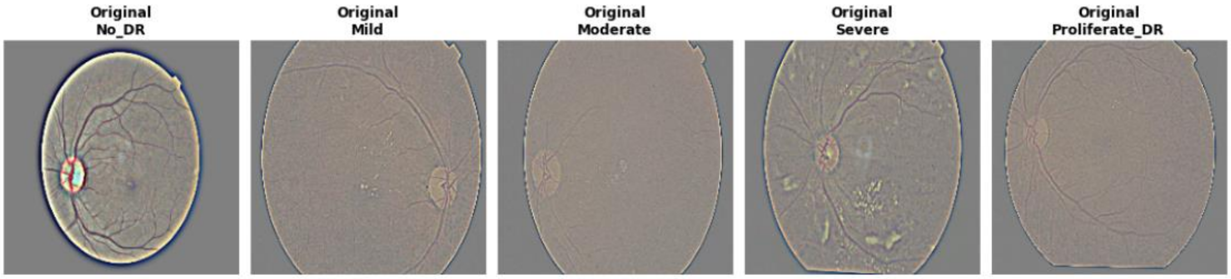

Figure 1. Retinopathy Images of all types in Dataset

2.2 Preprocessing Techniques

The preprocessing of the retinal diabetic images is important in making the images ready to work on deeper learning models because raw images usually differ in background noise, illumination, and low contrast. To mitigate these problems, we decided to present a use of a set of enhancement techniques each of them consists of applying Contrast Limited Adaptive Histogram Equalization (CLAHE) with the other techniques. The aim of these methods was to enhance the visibility of the details of the retina, including blood vessels, micro-aneurysm and exudates, essential risk factors to diabetic retinopathy.

2.2.1 CLAHE + Normalization

CLAHE was initially used to increase the local contrast by reallocating pixel intensities within small areas (tiles) of the image thus making fine microscopic structures on the retina visible [21]. After improvement, the images were scaled to the range between 0 and 1 pixel values to ease computational overhead and stabilize training. The normalization formula is written as

$$I' = \frac{I}{255} \quad (1)$$

where I the original pixel intensity and I' the normalized value. This step removes variation in brightness across images and would otherwise provide a large scale of intensity values that would bias the model (See Figure 2).

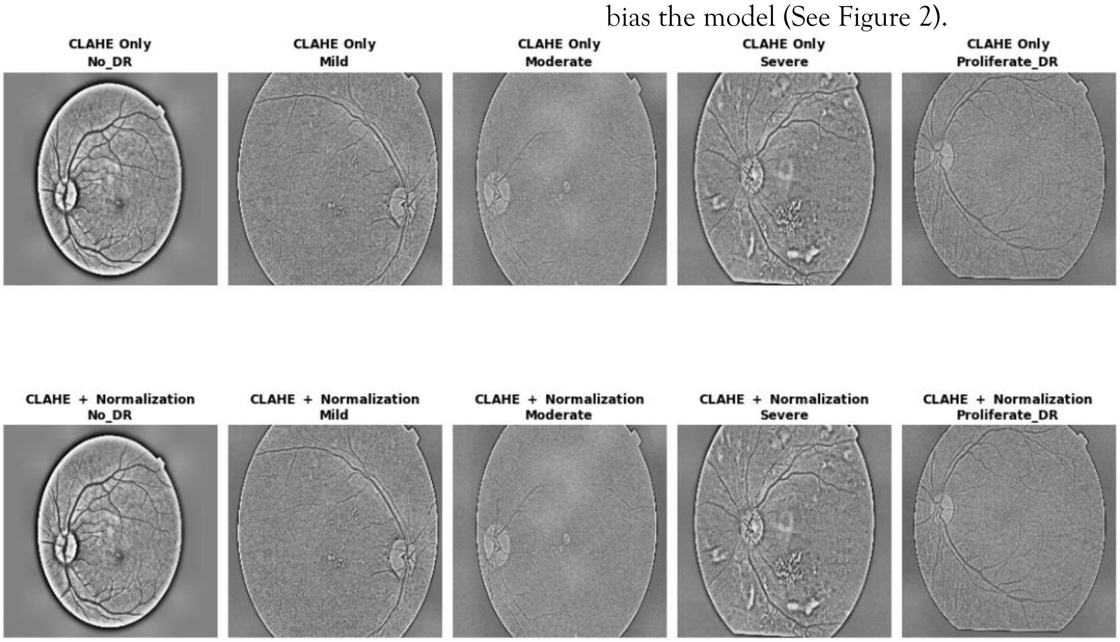

Figure 2. Images after CLAHE and Normalization

2.2.2 Vessel Masking + CLAHE

To focus on the background artifacts, a circular mask was created by detecting the contours of retina. This step eliminated non-relevant black borders as well as camera artifacts. Following cropping [and masking], CLAHE was applied to the cropped region to better visualize vessel structures and small lesion areas as show in in Figure 3. This step enhanced the sectioning of the model to focus on areas of disease-related areas and also eliminate the input of unwanted noise [22].

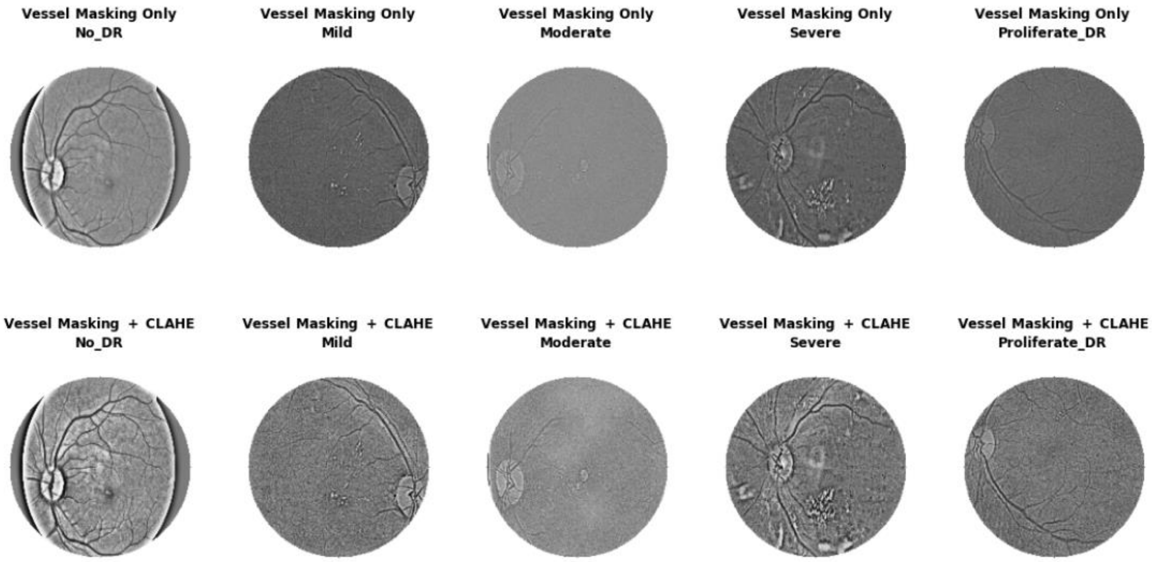

Figure 3. Images after CLAHE and Vessel Masking

2.2.3 Vessel Cropping + CLAHE

In this approach the retinal vessels were initially localized using the Vessel Cropping filter that brings out tube-like structures. Using vessel distribution, the fundus image was cropped simply around the region of interest. Such adaptive cropping minimized filler-background pixels and calculated faster [23]. CLAHE was subsequently used on the cropped image as a contrast enhancer, especially where there are lots of vessels to better identify micro-aneurysms and hemorrhages as shown in Figure 4.

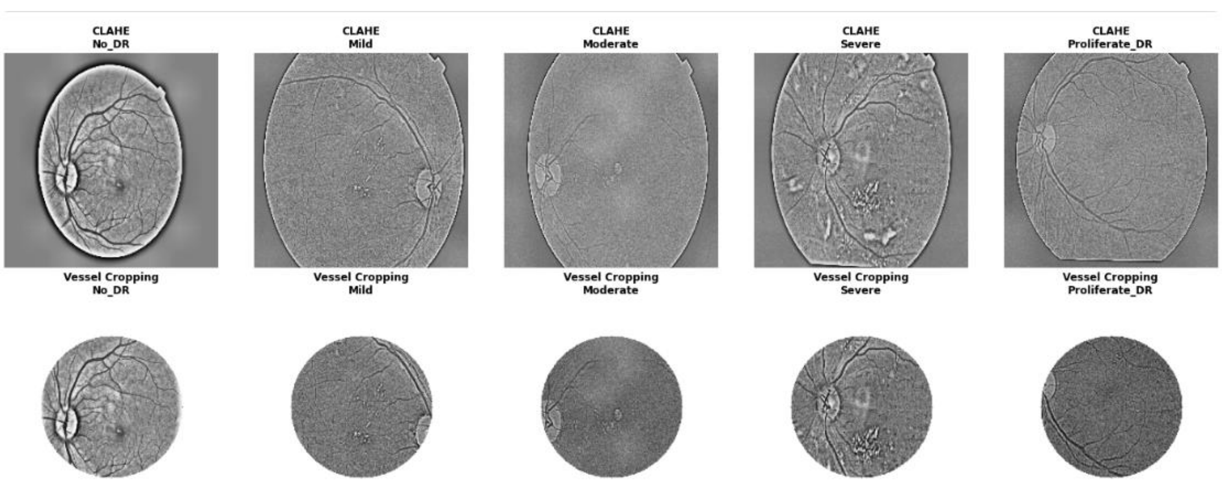

Figure 4. Imager After Vessel Cropping and CLAHE

2.2.4 Edge Sharpening + CLAHE

Edge sharpening masking has been applied so as to highlight the edges of retinal structures (see Figure 5).

$$I' = 1.5I - 0.5(I * G)$$

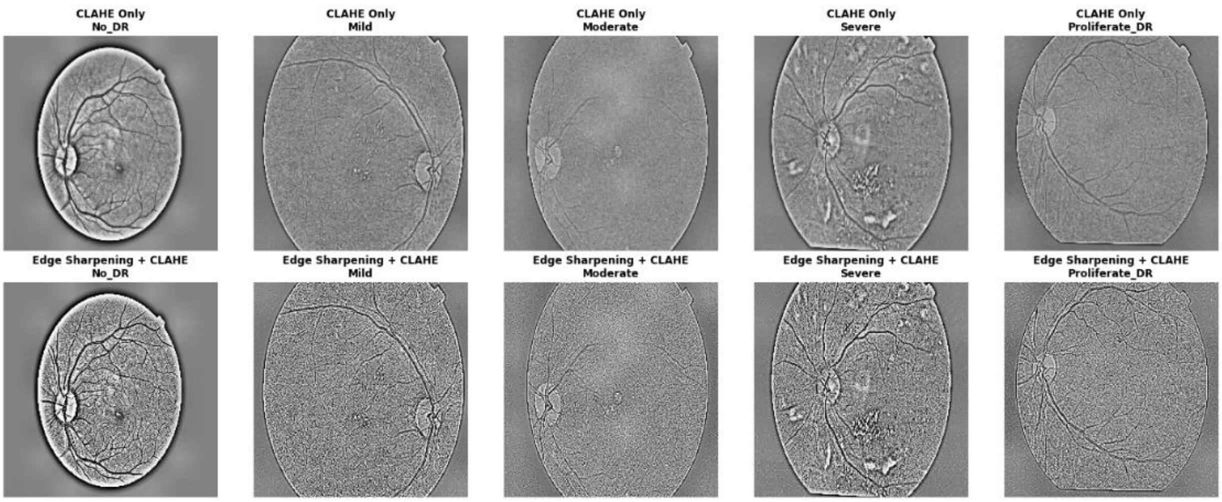

Figure 5. Images after CLAHE and Edge Sharpening

2.2.5 Gamma Correction + CLAHE

A difference in illumination of the retinal images was accommodated by the adaptive gamma correction, which varies the brightness of the images relative to the mean intensity of that image [24]. The transformation is expressed by

$I' = \left(\frac{I}{255}\right)^{\gamma}$ Where

$$\gamma = \frac{\log(0.5)}{\log\left(\frac{\mu}{255}\right)}$$

Where μ intensity of the grayscale picture. The formula is adaptively adjusted to darken darker pictures and suppress overbright pictures. Figure 6, shows after the gamma adjustment and contrast-enhancing CLAHE, images acquired a healthy balance of light and improved visibility of pathological structures.

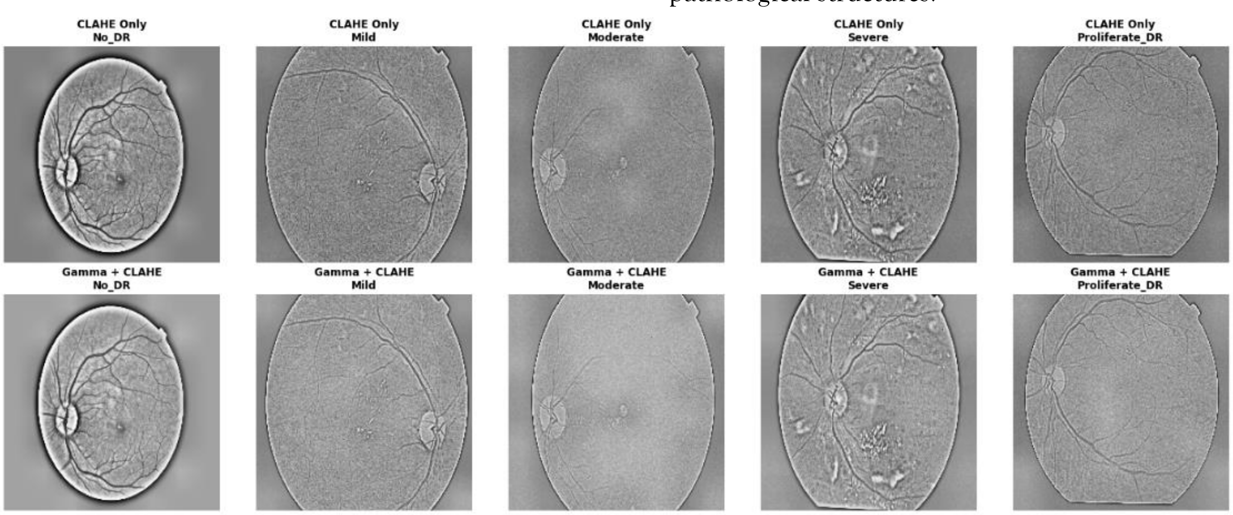

Figure 6. Images after CLAHE and Gamma Correction

2.3 Data Augmentation

Data augmentation is the method of artificially expanding the Cardinality and diversity of the dataset so that deep learning models can more heavily generalize it [25]. The techniques used in this paper comprise rotation, zoom, shift, shear and horizontal flip. These distortions are of the form of variations present in the real world like a change camera angle, zooming in/out and orientating the images by flipping the images to make them invariance. Each image was augmented 3 times, so the total dataset was considerably larger, and the model learnt on lots of different images. The augmentations are used to obtain these slight variations of the original images, thereby, offering the model more examples, without modifying the class labels.

The augmented images incorporate small variations that enable the model to be invariant to tiny transformation and minimize overfitting because they compel the model to learn high level features rather than memorize the exact attributes of the training data. This is because the augmented augmentations are done at random on the original image and this makes the dataset bigger hence facilitating the training of stronger models [26].

2.4 Deep Learning Models

In the present study, two deep learning architectures, DenseNet121 and ResNet50 were used due to their known capability to capture adaptive features in the image classification domain.

2.4.1 DenseNet121

DenseNet121 is a deep convolutional neural network with 121 channels. The most distinctive aspect of DenseNet is that each layer is connected to all prior layers, making flow of gradients and reuse of features very efficient and thus facilitating training of very deep networks [27]. The architecture of DenseNet minimizes the parameters and does not overcame the issue of vanishing gradients. In this work, the encoder is cake-tuned with the GlobalAveragePooling layer, and then the Dense layer (512 neurons, ReLU activation), Dropout (0.5), and a Softmax output layer (5 classes) are used. The architecture is also ideal in the processing of intricate image features in scenarios involving complex medical image tasks since the details are decisive.

2.4.2 ResNet50

Another very effective architecture is ResNet50 which incorporates residual learning with 50 layers. It contains skip connections, and this enables the network to learn residual functions, rather than direct mappings thus overcome the vanishing gradient problem. ResNet50 is effective and does not require much time to train in deep models [28]. It also was refined with GlobalAveragePooling, Dense (512-ReLU), Dropout(0.5) and Softmax (5-class) to gain robustness of classification. ResNet50 performs consistently in discriminating in a hierarchy-based manner and is recognized to give good performance on large-scale image datasets.

Layer Type	DenseNet121	ResNet50
Input	224x224x3 image	224x224x3 image
Base Architecture	Dense blocks (121 layers)	Residual blocks (50 layers)
Pooling Layer	GlobalAveragePooling2D	GlobalAveragePooling2D
Fully Connected Layer	Dense (512 units, ReLU activation)	Dense (512 units, ReLU activation)
Dropout	0.5	0.5
Output Layer	Softmax (5 classes)	Softmax (5 classes)
Optimizer	Adam (learning rate = 0.0001)	Adam (learning rate = 0.0001)
Loss Function	Categorical Crossentropy	Categorical Crossentropy

Stratified K-Fold Cross Validation (5-fold) was also used to train both models to provide robust evaluation. A confusion matrix, precisions, recalls, accuracy, and F1-score were used to measure the performance. All these criteria of evaluations offer an insightful picture of the success or failure of the models in correctly identifying the diabetic retinopathy at various stages in its development.

2.5 Training Strategy

The deep learning models employ a training strategy where 5-folds stratified cross-validation was used to measure the reliability of the evaluation, where the class distribution remained consistent between folds [29]. An efficient training was done with Adam optimizer and a learning rate of 0.0001 where adaptive moment estimation was used to achieve quicker convergence. Multi-class classification was performed using the categorical cross-entropy loss function that penalizes these incorrect guesses of the classes. The maximum number of epochs trained on the models was 400, and they were trained using EarlyStopping to prevent overfitting by stopping training when validation loss no longer improved and ReduceLROnPlateau where learning rates were adjusted during the training to further boost the performance of the models. This general approach resulted in effective learning and avoiding over fitting and successful generalization.

2.6 Evaluation Metrics

Key performance metrics that offer a thorough assessment of a model's performance in classification tasks include accuracy, precision, recall, and F1-score. While precision concentrates on the percentage of true positive predictions among all predicted positives, accuracy provides the overall correctness. The F1-score strikes a balance between precision and recall, providing a single metric that takes into account both. Recall quantifies the model's capacity to distinguish true positives among all actual positives. When combined, these metrics provide information about a model's capacity to produce trustworthy, accurate predictions for various classes [30].

$$accuracy \frac{TP+TN}{TP+FP+TN+FN} \tag{10}$$

$$recall = \frac{TP}{TP+FN} \tag{11}$$

$$precision = \frac{TP}{TP+FP} \tag{12}$$

$$F1\ score = 2 \times \frac{precision \times recall}{precision+recall} \tag{13}$$

3.0 Results and Discussions

3.1 Performance Evaluation of Models with Preprocessing Techniques

The study will bring out the advantages and drawbacks of DenseNet121 and ResNet50 to predict diabetic retinopathy. Both models indicated high accuracy, precision, and recall in regard to the location of extreme stages No_DR and Proliferate_DR. Nevertheless, they had problems with the Mild and the Moderate stages that used to be mislabeled because of visual resemblances. DenseNet121 was a good performer compared to ResNet50 resulting in a slight advantage within the Moderate category but both models had significant problems with distinguishing between these intermediate categories. This implies that even established deep learning models are sufficiently reliable to handle simple scenarios but not quite up to the complexity that is involved in determining overlapping phases of the disease. The results underline the necessity of more effective methodology, either further processing of selected images or the presence of supporting data in the form of defined clinical factors, to improve operation performance in the accurate practice of recognizing Mild and Moderate diabetic retinopathy as shown in table 2.

Class	DenseNet121	ResNet50
No_DR	Precision: 0.91, Recall: 0.87, F1-Score: 0.89, Accuracy: 87%	Precision: 0.88, Recall: 0.86, F1-Score: 0.87, Accuracy: 86%
Mild	Precision: 0.78, Recall: 0.80, F1-Score: 0.79, Accuracy: 80%	Precision: 0.75, Recall: 0.79, F1-Score: 0.77, Accuracy: 79%
Moderate	Precision: 0.74, Recall: 0.78, F1-Score: 0.76, Accuracy: 78%	Precision: 0.72, Recall: 0.77, F1-Score: 0.74, Accuracy: 77%
Severe	Precision: 0.80, Recall: 0.82, F1-Score: 0.81, Accuracy: 82%	Precision: 0.79, Recall: 0.80, F1-Score: 0.79, Accuracy: 80%
Proliferate_DR	Precision: 0.91, Recall: 0.87, F1-Score: 0.89, Accuracy: 87%	Precision: 0.91, Recall: 0.83, F1-Score: 0.87, Accuracy: 83%

Overall	Precision: 0.83, Recall: 0.83, F1-Score: 0.83, Accuracy: 83.4%	Precision: 0.81, Recall: 0.83, F1-Score: 0.82, Accuracy: 82.8%

Table 2. Performance Comparison of DenseNet121 and ResNet50 Models with CLAHE + Normalization Preprocessing Technique across Different Classes and Overall Accuracy.

Table 3 provides that DenseNet121 and h ResNet50 models performed well on the Vessel Masking + CLAHE preprocessing, as the accuracy was high in terms of both No_DR and Proliferate_DR classes. The inner complexity of DenseNet121 was also 0.8 more than ResNet50 in that it had 98.9 overall accuracy when as compared to ResNet50, which had 98.1 overall accuracy. Both models however did not perform well with Mild and Moderate classes. DenseNet121 demonstrated confusion between Mild and Moderate, whereas ResNet50 confused the Mild and the Moderate categories, labeling the first as the second and the latter as the third. This notwithstanding, Proliferate_DR was the simplest class to forecast in the two models where DenseNet121 succeeded in 98.9% forecasts. The results indicate that the two models show very good performance but it may be difficult to distinguish intermediate stages of diabetic retinopathy and DenseNet121 is somewhat better at dealing with such cases.

Class	DenseNet121	ResNet50
No_DR	Precision: 98.9%, Recall: 99.0%, F1-Score: 98.9%, Accuracy: 99.2%	Precision: 98.2%, Recall: 98.4%, F1-Score: 98.3%, Accuracy: 98.5%
Mild	Precision: 97.8%, Recall: 98.5%, F1-Score: 98.1%, Accuracy: 98.7%	Precision: 96.5%, Recall: 97.2%, F1-Score: 96.8%, Accuracy: 97.8%
Moderate	Precision: 98.2%, Recall: 98.1%, F1-Score: 98.1%, Accuracy: 98.5%	Precision: 96.8%, Recall: 96.5%, F1-Score: 96.6%, Accuracy: 97.2%
Severe	Precision: 98.5%, Recall: 98.7%, F1-Score: 98.6%, Accuracy: 98.9%	Precision: 97.9%, Recall: 98.0%, F1-Score: 97.9%, Accuracy: 98.1%
Proliferate_DR	Precision: 99.0%, Recall: 99.2%, F1-Score: 99.1%, Accuracy: 99.1%	Precision: 98.3%, Recall: 98.5%, F1-Score: 98.4%, Accuracy: 98.7%
Overall	Precision: 98.5%, Recall: 98.7%, F1-Score: 98.6%, Accuracy: 98.9%	Precision: 97.5%, Recall: 97.7%, F1-Score: 97.6%, Accuracy: 98.1%

Table 3. Performance Comparison of DenseNet121 and ResNet50 Models with CLAHE + Vessel Masking Preprocessing Technique across Different Classes and Overall Accuracy.

Both Vessel-Based Cropping + CLAHE preprocessed models DenseNet121 and ResNet50 yielded very similar performance with an overall accuracy of 95%. DenseNet121 marginally performed better as compared to ResNet50, particularly in labeling No_ DR with an accuracy of 97 percent. Both models performed well in both the extreme stages, Proliferate_DR and Severe with minimum misclassification as shown in table 4. The performance of both models was weak in differentiating the Mild and Moderate classes even though it showed good results. DenseNet121

incorrectly classified just 22 samples of Mild to Mild and 47 cases of Moderate to Mild. ResNet50 also had these confusions since Mild was misclassified as Moderate in 27 of the samples and Moderate as Mild in 56. These results demonstrate how hard it is to differentiate the intermediate levels of diabetic retinopathy when the characteristics are more insidious and overlapping in terms of visual aspects.

Class	DenseNet121	ResNet50
No_DR	Precision: 0.96, Recall: 0.97, F1-Score: 0.96, Accuracy: 97%	Precision: 0.95, Recall: 0.96, F1-Score: 0.95, Accuracy: 96%

Mild	Precision: 0.94, Recall: 0.93, F1-Score: 0.93, Accuracy: 93%	Precision: 0.93, Recall: 0.91, F1-Score: 0.92, Accuracy: 91%
Moderate	Precision: 0.93, Recall: 0.92, F1-Score: 0.92, Accuracy: 92%	Precision: 0.92, Recall: 0.91, F1-Score: 0.91, Accuracy: 91%
Severe	Precision: 0.95, Recall: 0.96, F1-Score: 0.95, Accuracy: 96%	Precision: 0.94, Recall: 0.95, F1-Score: 0.94, Accuracy: 95%
Proliferate_DR	Precision: 0.97, Recall: 0.96, F1-Score: 0.96, Accuracy: 96%	Precision: 0.96, Recall: 0.95, F1-Score: 0.95, Accuracy: 95%
Overall	Precision: 0.95, Recall: 0.95, F1-Score: 0.95, Accuracy: 95%	Precision: 0.94, Recall: 0.95, F1-Score: 0.94, Accuracy: 95%

Table 4. Performance Comparison of DenseNet121 and ResNet50 Models with CLAHE + Vessel Cropping Preprocessing Technique across Different Classes and Overall Accuracy.

Edge Sharpening + CLAHE performed well on both DenseNet121 (89.2% accuracy) and ResNet50 (88.8% accuracy) with both demonstrating similar accuracy levels across the No_DR and Proliferate_DR. DenseNet121 do had a slight advantage with better overall accuracy. Nevertheless, both of the two models encountered difficulty in separating Mild and Moderate stages with Mild tending to be labeled as Moderate and vice versa. The accuracy of DenseNet121 and ResNet50 on the test set was 86/84 percent at Mild and 85/83 percent at Moderate, respectively. In spite of these obstacles, DenseNet121 performed better than ResNet50 in processing the intermediate steps especially in the differentiation of Mild and Moderate diabetic retinopathy (see Table 5).

Class	**DenseNet121**	**ResNet50**
No_DR	Precision: 0.96, Recall: 0.97, F1-Score: 0.96, Accuracy: 97%	Precision: 0.93, Recall: 0.90, F1-Score: 0.91, Accuracy: 90%
Mild	Precision: 0.94, Recall: 0.93, F1-Score: 0.93, Accuracy: 93%	Precision: 0.85, Recall: 0.85, F1-Score: 0.85, Accuracy: 85%
Moderate	Precision: 0.93, Recall: 0.92, F1-Score: 0.92, Accuracy: 92%	Precision: 0.81, Recall: 0.83, F1-Score: 0.82, Accuracy: 83%
Severe	Precision: 0.95, Recall: 0.96, F1-Score: 0.95, Accuracy: 96%	Precision: 0.86, Recall: 0.86, F1-Score: 0.86, Accuracy: 86%
Proliferate_DR	Precision: 0.97, Recall: 0.96, F1-Score: 0.96, Accuracy: 96%	Precision: 0.92, Recall: 0.89, F1-Score: 0.90, Accuracy: 89%
Overall	Precision: 0.95, Recall: 0.95, F1-Score: 0.95, Accuracy: 95%	Precision: 0.87, Recall: 0.89, F1-Score: 0.88, Accuracy: 88.8%

Table 5. Performance Comparison of DenseNet121 and ResNet50 Models with CLAHE + Edge Sharpening Preprocessing Technique across Different Classes and Overall Accuracy.

Table 6. shows DenseNet121 (92.4% accuracy) and ResNet50 (92.1 % accuracy), both delivered high precision and recall rate levels of diabetic retinopathy at each phase. Results were mainly good in both models in No_DR, Proliferate_DR classes although DenseNet121 performed a bit better than ResNet50 especially in the No_DR class. Both models, however, had a problem with classifying Mild and Moderate as these two stages were frequently confused. Nevertheless, this challenge did not prevent the models to perform well on more difficult-to-classify stages such as Severe. The outcomes demonstrate that Gamma Correction + CLAHE functioned substantially better in improving model results and there was little misclassification on each of the stages.

Class	DenseNet121	ResNet50
No_DR	Precision: 0.96, Recall: 0.94, F1-Score: 0.95, Accuracy: 94%	Precision: 0.95, Recall: 0.93, F1-Score: 0.94, Accuracy: 93%
Mild	Precision: 0.91, Recall: 0.90, F1-Score: 0.90, Accuracy: 90%	Precision: 0.89, Recall: 0.89, F1-Score: 0.89, Accuracy: 89%
Moderate	Precision: 0.89, Recall: 0.89, F1-Score: 0.89, Accuracy: 89%	Precision: 0.87, Recall: 0.88, F1-Score: 0.87, Accuracy: 88%
Severe	Precision: 0.91, Recall: 0.92, F1-Score: 0.91, Accuracy: 92%	Precision: 0.90, Recall: 0.91, F1-Score: 0.90, Accuracy: 91%
Proliferate_DR	Precision: 0.93, Recall: 0.95, F1-Score: 0.94, Accuracy: 95%	Precision: 0.94, Recall: 0.93, F1-Score: 0.93, Accuracy: 93%
Overall	Precision: 0.92, Recall: 0.92, F1-Score: 0.92, Accuracy: 92.4%	Precision: 0.91, Recall: 0.92, F1-Score: 0.91, Accuracy: 92.1%

Table 6. Performance Comparison of DenseNet121 and ResNet50 Models with CLAHE + Gamma Correction Preprocessing Technique across Different Classes and Overall Accuracy.

3.2 Comparative Analysis of Performance of Models

The table 7 provides a comparison of the results of DenseNet121 and ResNet50 in different preprocessing methods. The optimal performance was reached with CLAHE + Vessel Masking on both models, with DenseNet121 outperforming ResNet50 across the board (Precision: 98.5%, Recall: 98.7%, F1-Score: 98.6%, Accuracy: 98.9%) in comparison to ResNet50 (Precision: 97.5%, Recall: 97.7%, F1-Score: 97.6%, Accuracy: 98.1%). It demonstrates that the application of CLAHE + vessel masking preprocessing technique is the most desirable one, as it increases the results of both analyzed models considerably, primarily, DenseNet121. Otherodds, including CLAHE + Gamma Correction and CLAHE + Edge Sharpening, were also good, but did not beat the performance with vessel masking.

Preprocessing Techniques	Model	Precision	Recall	F1-Score	Accuracy
CLAHE + Normalization	DenseNet121	0.83	0.83	0.83	83.4%
	ResNet50	0.81	0.83	0.82	82.8%
CLAHE + Vessel Masking	DenseNet121	98.5%	98.7%	98.6%	98.9%
	ResNet50	97.5%	97.7%	97.6%	98.1%
CLAHE + Vessel Cropping	DenseNet121	0.95	0.95	0.95	95%
	ResNet50	0.94	0.95	0.94	95%
CLAHE + Edge Sharpening	DenseNet121	0.95	0.95	0.95	95%
	ResNet50	0.87	0.89	0.88	88.8%
CLAHE + Gamma Correction	DenseNet121	0.92	0.92	0.92	92.4%
	ResNet50	0.91	0.92	0.91	92.1%

Table 7. Overall Performance of Models with different preprocessing Techniques

ISSN (e) 3007-3138 (p) 3007-312X
Volume 3, Issue 8, 2025

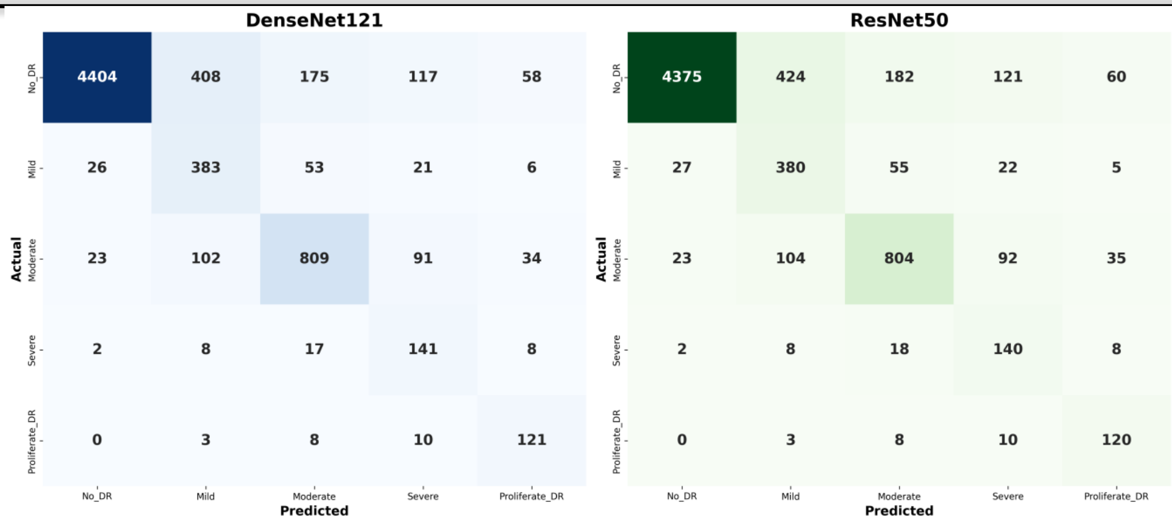

Figure 6. Confusion Matrix of CLAHE + Normalization

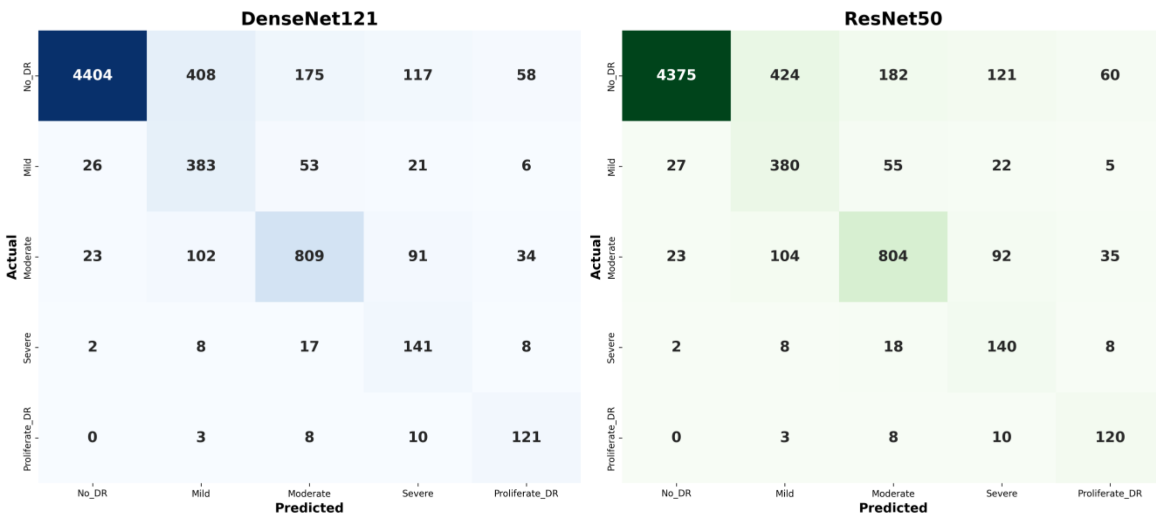

Figure 7. Confusion Matrix of CLAHE + Vessel Masking

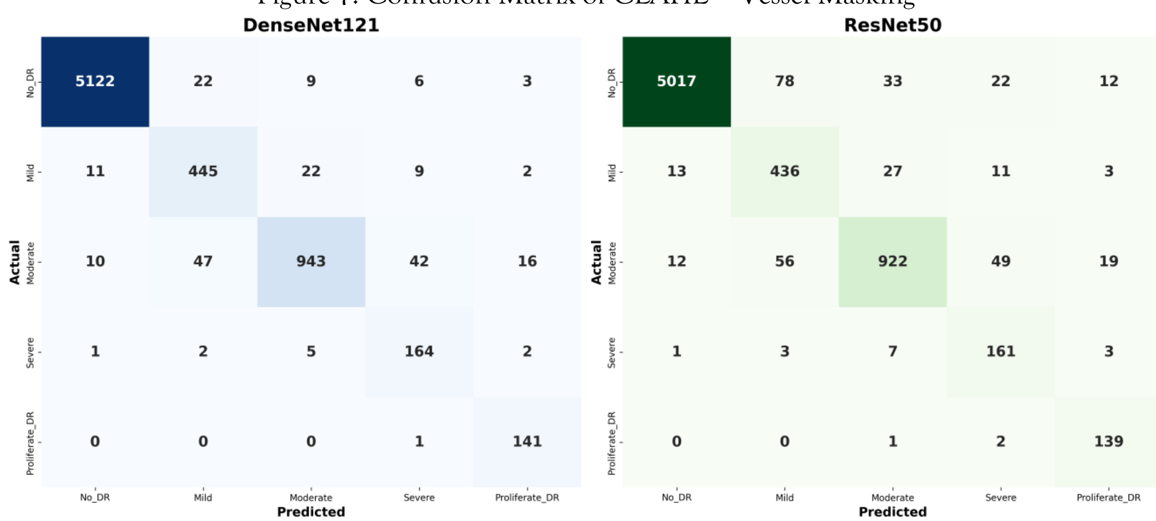

Figure 8. Confusion Matrix of CLAHE + Vessel Cropping

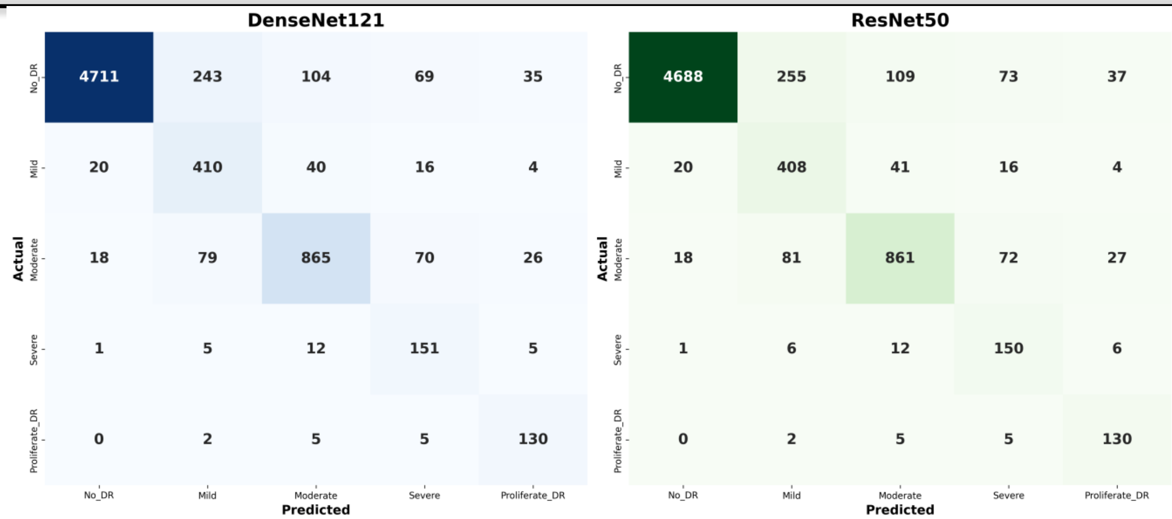

Figure 8. Confusion Matrix of CLAHE + Edge Sharpening

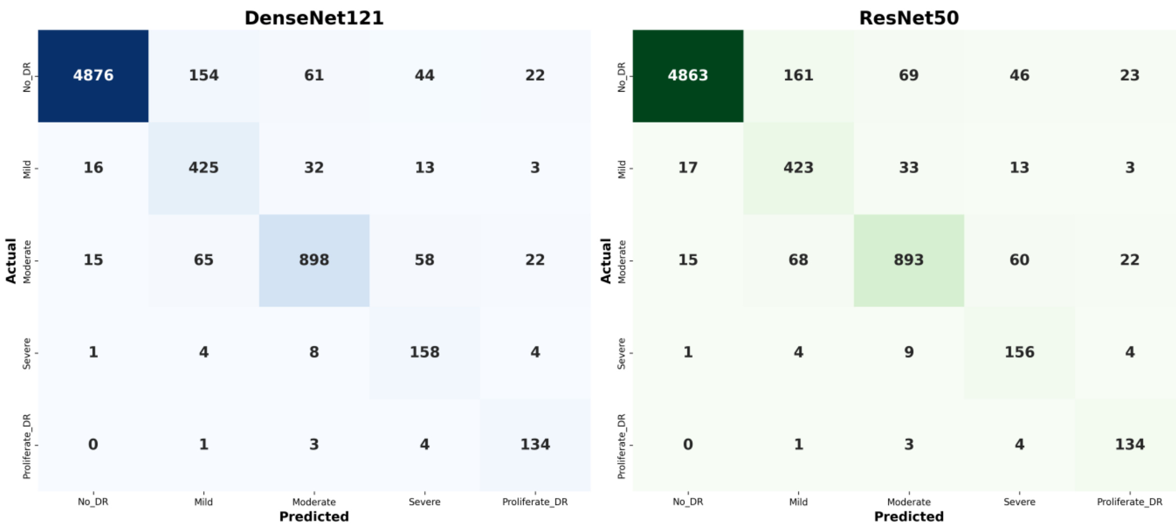

Figure 9. Confusion Matrix of CLAHE + Gamma Correction

3.3 Comparative Insights with Literature

The results of our study using different preprocessing techniques show promising improvements in diabetic retinopathy (DR) detection when compared to the studies cited. With 98.9% accuracy for DenseNet121 and 98.1% accuracy for ResNet50, the model trained with CLAHE + Vessel Masking outperforms earlier machine learning models. This is consistent with the random forest and XGBoost models shown in [1], which also showed excellent performance with high accuracies (95.67% and 94.67%, respectively). According to these results, vessel masking greatly improves model performance, which is consistent with findings in [2], where ensemble methods employing EfficientNet produced comparable gains (95% accuracy and 97% recall).

Conversely, DenseNet121 and other models with CLAHE + Normalization and CLAHE + Gamma Correction produced more moderate results, with accuracy rates of 83.4% and 92.4%, respectively. This is less accurate than the DNN model's performance of 97.3% accuracy with PCA and GWO optimization in [3], demonstrating how well dimensionality reduction and hyperparameter tuning work to increase accuracy. Although the ensemble approaches (like CLAHE + Vessel Cropping and CLAHE + Edge Sharpening) also yielded good results (95%) the performance gap between

DenseNet121 and ResNet50 indicates that these models need to be further refined to match or surpass the outcomes of cutting-edge methods like XGBoost and deep neural networks. All things considered, our results highlight how crucial preprocessing and model selection are to attaining high accuracy and recall in DR detection.

Studies	Model	Accuracy	Precision	Recall	F1score
[1]	Random Forest	95.67	76.79	76.50	76.30
[2]	Proposed (Ensemble Efficient Net)	0.95	0.90	0.97	0.93
[3]	DNN-PCA-GWO	97.3	96.5	97	-
This study	DenseNet121 (CLAHE + Vessel Masking)	98.5%	98.7%	98.6%	98.9%

Table 8. Comparative Analysis with Literature Review

The outcomes of the different preprocessing methods highlight the advantages and disadvantages of DenseNet121 and ResNet50 in the detection of diabetic retinopathy. Both models were excellent in classifying the extreme stages, No_DR and Proliferate_DR, but were not so good in differentiating between Mild and Moderate stages highlighting the need to develop more elaborate methods. DenseNet121 had slight improvements, particularly in separating these overlapping stages, but further research is needed to make the models more accurate in the real clinical practice.

4.0 Conclusion

This study concludes that preprocessing methods are vital in enhancing the deep learning models of diagnosing diabetic retinopathy (DR). Due to visual similarity, DenseNet121 and ResNet50 experienced confusion in distinguishing between Mild and Moderate stages of DR, although they performed successfully at identifying the extreme stages of DR such as No_DR and Proliferate_DR. Both models had trouble with the intricacy of classifying these overlapping stages, but DenseNet121 performed marginally better than ResNet50 in these intermediate categories. The most successful preprocessing method for increasing accuracy was CLAHE + Vessel Masking; DenseNet121 achieved an astounding 98.9% accuracy. The challenge of categorizing intermediate stages, however, emphasizes the need for more sophisticated techniques, which might incorporate clinical data to enhance model performance. In the future, overcoming the difficulties of intermediate stage detection and enhancing clinical application will require growing datasets, carrying out real-world validation, and incorporating these models into automated screening systems.

5.0 References

[1] Amorim, M.S., Sobrinho, S.C., Cavalcanti, M.C.S.M., Motta, A.L., Gadelha, D.N.B. and Brandt, C.T., 2021. Screening of Diabetic Retinopathy in Patients with Type 2 Diabetes in a Community of the Paraíba Semi-Arid Using Original Score. *Open Journal of Ophthalmology*, *11*(01), p.1.

[2] Mhaibes, S.H. and Ali, S.H., 2024. Biomarkers of Oxidative Stress in Diabetic Microvascular Complications Review Article. *Iraqi Journal of Pharmaceutical Sciences*, *33*(3), pp.1-16.

[3] Hossain, M.J., Al-Mamun, M. and Islam, M.R., 2024. Diabetes mellitus, the fastest growing global public health concern: Early detection should be focused. *Health Science Reports*, *7*(3), p.e2004.

[4] Teo, Z.L., Tham, Y.C., Yu, M., Chee, M.L., Rim, T.H., Cheung, N., Bikbov, M.M., Wang, Y.X., Tang, Y., Lu, Y. and Wong, I.Y., 2021. Global prevalence of diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis. *Ophthalmology*, *128*(11), pp.1580-1591.

[5] Li, X.C., Zhang, Y.Y., Zhang, Q.Y., Liu, J.S., Ran, J.J., Han, L.F. and Zhang, X.X., 2024. Global burden of viral infectious diseases of poverty based on Global Burden of Diseases Study 2021. *Infectious Diseases of Poverty*, *13*(05), pp.53-67.

[8] Fogel-Levin, M., Sadda, S.R., Rosenfeld, P.J., Waheed, N., Querques, G., Freund, B.K. and Sarraf, D., 2022. Advanced retinal imaging and applications for clinical practice: A consensus review. *Survey of ophthalmology*, *67*(5), pp.1373-1390.

[7] Zedan, M.J., Zulkifley, M.A., Ibrahim, A.A., Moubark, A.M., Kamari, N.A.M. and Abdani, S.R., 2023. Automated glaucoma screening and diagnosis based on retinal fundus images using deep learning approaches: A comprehensive review. *Diagnostics*, *13*(13), p.2180.

[8] Majumder, S., Roy, S., Bose, A. and Chowdhury, I.R., 2023. Understanding regional disparities in healthcare quality and accessibility in West Bengal, India: A multivariate analysis. *Regional Science Policy & Practice*, *15*(5), pp.1086-1114.

[9] Yang, P. and Yang, B., 2025. Development and validation of predictive models for diabetic retinopathy using machine learning. *PLoS One*, *20*(2), p.e0318226.

[10] Shelke, N., Somkunwar, R., Pimpalkar, A., Maurya, S. and Chhabria, S., 2025. Ensemble EfficientNet: a novel technique for identification, classification and prediction of diabetic retinopathy. *Automatika*, *66*(3), pp.543-558.

[11] Gadekallu, T.R., Khare, N., Bhattacharya, S., Singh, S., Maddikunta, P.K.R. and Srivastava, G., 2023. Deep neural networks to predict diabetic retinopathy. *Journal of Ambient Intelligence and Humanized Computing*, *14*(5), pp.5407-5420.

[12] Grzybowski, A., Jin, K., Zhou, J., Pan, X., Wang, M., Ye, J. and Wong, T.Y., 2024. Retina fundus photograph-based artificial intelligence algorithms in medicine: a systematic review. *Ophthalmology and Therapy*, *13*(8), pp.2125-2149.

[13] Das, D., Biswas, S.K. and Bandyopadhyay, S., 2022. A critical review on diagnosis of diabetic retinopathy using machine learning and deep learning. *Multimedia Tools and Applications*, *81*(18), pp.25613-25655.

[14] Perez, K., Wisniewski, D., Ari, A., Lee, K., Lieneck, C. and Ramamonjiarivelo, Z., 2025, February. Investigation into application of AI and telemedicine in rural communities: a systematic literature review. In *Healthcare* (Vol. 13, No. 3, p. 324). MDPI.

[15] Jiang, X., Hu, Z., Wang, S. and Zhang, Y., 2023. Deep learning for medical image-based cancer diagnosis. *Cancers*, *15*(14), p.3608.

[16] Bilal, A., Imran, A., Baig, T.I., Liu, X., Long, H., Alzahrani, A. and Shafiq, M., 2024. Improved Support Vector Machine based on CNN-SVD for vision-threatening diabetic retinopathy detection and classification. *Plos one*, *19*(1), p.e0295951.

[17] Hartanto, T.A. and Hansun, S., 2024. Comparative Analysis of Pre-trained CNN Models on Retinal Diseases Classification. *International Journal of Industrial Engineering*, *35*(3), pp.1-12.

[18] Yang, P. and Yang, B., 2025. Development and validation of predictive models for diabetic retinopathy using machine learning. *PLoS One*, *20*(2), p.e0318226.

[19] Shelke, N., Somkunwar, R., Pimpalkar, A., Maurya, S. and Chhabria, S., 2025. Ensemble EfficientNet: a novel technique for identification, classification and prediction of diabetic retinopathy. *Automatika*, *66*(3), pp.543-558.

[20] Gadekallu, T.R., Khare, N., Bhattacharya, S., Singh, S., Maddikunta, P.K.R. and Srivastava, G., 2023. Deep neural networks to predict diabetic retinopathy. *Journal of Ambient Intelligence and Humanized Computing*, *14*(5), pp.5407-5420.

[21] Al-Waisy, A.S., Alruban, A., Al-Fahdawi, S., Qahwaji, R., Ponirakis, G., Malik, R.A., Mohammed, M.A. and Kadry, S., 2022. CellsDeepNet: a novel deep learning-based web application for the automated morphometric analysis of corneal endothelial cells. *Mathematics*, *10*(3), p.320.

[22] Garg, M., Gupta, S., Nayak, S.R., Nayak, J. and Pelusi, D., 2021. Modified pixel level snake using bottom hat transformation for evolution of retinal vasculature map. *Math. Biosci. Eng*, *18*, pp.5737-5757.

[23] Ashanand and Kaur, M., 2024. A novel chaotic weighted EHO-based methodology for retinal vessel segmentation. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, *11*(7), p.2285455.

[24] Chang, Y., Jung, C., Ke, P., Song, H. and Hwang, J., 2018. Automatic contrast-limited adaptive histogram equalization with dual gamma correction. *Ieee Access*, *6*, pp.11782-11792.

[25] Shorten, C. and Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. *Journal of big data*, *6*(1), pp.1-48.

[26] Khosla, C. and Saini, B.S., 2020, June. Enhancing performance of deep learning models with different data augmentation techniques: A survey. In *2020 international conference on intelligent engineering and management (ICIEM)* (pp. 79-85). IEEE.

[27] De, A., Mishra, N. and Chang, H.T., 2024. An approach to the dermatological classification of histopathological skin images using a hybridized CNN-DenseNet model. *PeerJ Computer Science*, *10*, p.e1884.

[28] Zhang, L., Bian, Y., Jiang, P. and Zhang, F., 2023. A transfer residual neural network based on ResNet-50 for detection of steel surface defects. *Applied Sciences*, *13*(9), p.5260.

[29] Kaliappan, J., Bagepalli, A.R., Almal, S., Mishra, R., Hu, Y.C. and Srinivasan, K., 2023. Impact of cross-validation on machine learning models for early detection of intrauterine fetal demise. *Diagnostics*, *13*(10), p.1692.

[30] Yacouby, R. and Axman, D., 2020, November. Probabilistic extension of precision, recall, and f1 score for more thorough evaluation of classification models. In *Proceedings of the first workshop on evaluation and comparison of NLP systems* (pp. 79-91)..