

Early Detection and Severity Classification of Diabetic Retinopathy Through Image Processing and Deep Learning

Nelson Nishio

*BASIS Independent Silicon
Valley, San Jose, USA*

Email: nelsonknishio@gmail.com

Jason Nishio

*BASIS Independent Silicon
Valley, San Jose, USA*

jasonnishioxd@gmail.com

Alexander Nishio

*Department of Computer Science
Purdue University
West Lafayette, USA
anishio@purdue.edu*

Abstract—Diabetic retinopathy (DR) remains a leading cause of blindness among individuals with diabetes, making early detection crucial to mitigate vision deterioration and prevent vision loss. However, the current manual DR diagnosis by ophthalmologists in clinics is labor-intensive, particularly due to challenges in identifying subtle retinal lesions. Hence, there is a need to develop an automated detection framework. This study leveraged the two pre-trained ResNet50 and VGG19 models for detecting DR in retinal fundus images from the APTOS 2019 dataset. The advanced image preprocessing techniques such as the Graham method, Contrast Limited Adaptive Histogram Equalization (CLAHE), and Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) were explored. The results demonstrate that combining these preprocessing methods with optimized parameters significantly enhances performance. For binary classification, the pre-trained VGG19 model, combined with the proposed CLAHE + ESRGAN preprocessing method, achieved an accuracy of 98.91%. For multi-class classification, the pre-trained ResNet50 model, combined with the proposed Graham20 + CLAHE and ESRGAN + CLAHE preprocessing methods, attained accuracies of 86.07% and 84.15%, respectively. This study is the first systematic study of deep learning architectures with various combinations of DR image preprocessing methods, offering a significant step forward in automated DR detection and severity classification.

Keywords—Diabetic Retinopathy, Deep Learning, Image Processing, CLAHE, Gaussian Blur, ESRGAN, ResNet50, VGG19

I. INTRODUCTION

Diabetic retinopathy (DR) is a leading cause of blindness among working-aged adults, affecting millions of people worldwide. Approximately 537 million people globally live with diabetes [1], and around 103 million (22%) of them suffer from diabetic retinopathy. In the United States, according to the Centers for Disease Control and Prevention (CDC), one in ten people have diabetes [2], and one in five who have diabetes are unaware of their condition. Among those over 40, one in three exhibit signs of DR. DR is developed when a patient has diabetes for at least 10 years without diagnosis and is unaware of it. This trend is getting worse each year.

Clinical data has shown that diabetes can lead to numerous severe complications, including as DR, visual impairment, cardiovascular disease, kidney disease, and strokes. Elevated blood glucose levels cause inflammation and leakage within retinal blood vessels, leading to macular swelling or thickening, which hinders blood flow. Additionally, abnormal new blood vessel growth may occur on the retina [3]. Based on the type and quantity of lesions in fundus images, DR can be classified into five stages: 0 (No DR), 1 (mild DR), 2 (moderate DR), 3 (severe DR), and 4

(proliferative DR), as used in the Asia Pacific Tele-Ophthalmology Society (APTOS) 2019 Blindness Detection dataset [4]. Stage 1, 2, and 3 are non-proliferative. The primary pathological DR lesions include hemorrhages, microaneurysms, exudates, and retinal neovascularization [5]. Fig. 1(a) shows sample retinal lesions of a retinal fundus image in a non-proliferative stage, with red dot-shaped microaneurysms as the first visible indications of mild DR. The microaneurysms are small circular red dots with distinct margins, typically no larger than a blood vessel's width. Hemorrhages, in contrast, appear as larger, darker red spots, indicating bleeding within the retina, often resulting from ruptured microaneurysms or increased blood leakage from capillaries. Damage to nerve fibers leads to soft exudates resembling yellowish-white fluffy specks or cotton wool spots. Hard exudates are white-yellowish cholesterol deposits that usually originate from leaking microaneurysms. A higher prevalence of these lesions across all four eye quadrants indicates more severe gradings of non-proliferative DR. Fig. 1(b) and 1(c) demonstrate the sample lesions of two retinal fundus images in proliferative DR stages, where neovascularization—the development of new retinal vessels near the optic disc due to retinal malnourishment—is the key indicator. Following neovascularization, fibrous proliferation occurs as white fibrous tissues grow around the new vessels to provide protective support. As these new blood vessels grow, they can increase pressure in the eyeball, disrupting fluid flow and causing the retina to detach from the back of the eye. Additionally, blood may leak into the vitreous, the jelly-like substance at the eye's center. These factors contribute to optic nerve damage, impairing the nerve responsible for transmitting inverted images from the eye's blind spot to the brain and leading to vision loss.

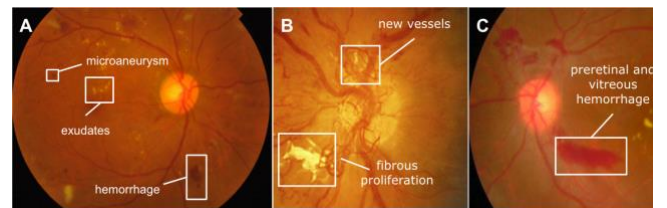


Fig. 1. Retinal images of diabetic retinopathy lesions (a) non-proliferative, (b) and (c) show different types of proliferative lesions.

Timely detection and intervention of DR are crucial to preventing vision deterioration and minimizing the risk of permanent vision loss. However, the increasing prevalence of DR presents a considerable challenge, especially in resource-constrained regions where ophthalmologists are burdened with heavy screening workloads. Therefore, there is an urgent need to develop automated and effective DR diagnosis and prognostic techniques to address the growing number of untreated cases and reduce the strain on eye care providers.

II. RELATED WORK

Over the past decade, academics have developed numerous Machine Learning (ML) and Deep Learning (DL) algorithms for DR detection. Among these, transfer learning has emerged as one of the most popular methods for automatic DR detection [6]. In transfer learning, a model that is already trained for a particular task is repurposed as a starting point for a different task. The advantages of transfer learning include higher initial training accuracy, faster training rates, and better convergence compared to models that do not utilize transfer learning. It is widely used for mitigating the limitations of small datasets with few images. Despite significant advancements in DL-based image analysis for reliable DR detection, there remains considerable room for improving accuracy. Most research efforts have focused on the segmentation and classification of retinal lesions. For instance, Kumar et al. [7] proposed DRISTI, a hybrid model combination VGG16 and capsule networks, which did not employ specialized image preprocessing. Their model achieved impressive training and validation accuracies of 99.96% and 97.05% on the augmented APTOS 2019 dataset; however, its testing accuracy for multi-stage DR classification was only 75.81%. More recent studies have begun to focus on image-preprocessing of retinal fundus images to enhance the lesion features. Alghamdi [8], for example, preprocessed the fundus images using Grad-CAM technique, but the models were unable to capture DR lesions effectively, resulting in poor accuracies. Kassani et al. [9] employed the min-pooling preprocessing technique introduced by Graham [10], the winner of the Kaggle DR grading competition, to remove image variations caused by different lighting conditions or imaging devices. They developed a modified Xception architecture with an aggregation of convolutional neural network (CNN) layers to more effectively fuse deep features and improve the learning process. The achieved accuracy was 83.09% on the APTOS 2019 dataset. Islam et al. [11] applied Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance the image quality and deployed a pre-trained Xception CNN model as the encoder with transfer learning. They proposed a supervised contrastive loss (SCL) function to visualize the embedding space and identify the DR severity stages. The proposed model achieved the test accuracies of 98.36% and 84.36% for DR binary and multi-stage classifications, respectively. Nneji et al. [12] extracted DR-related features from CLAHE fundus images by fine-tuned Inception V3 and extracted the DR features of the contrast-enhanced canny edge detection (CECED) fundus images using a fine-tuned VGG-16. The outputs of both channels were merged in a weighted fusion deep learning network (WFDLN), resulting in a reported accuracy of 98.0% on the Kaggle EYEPACS dataset [13]. This literature review highlights the critical role of image preprocessing, as the quality of preprocessing significantly affects classification results.

Other studies modified the original dataset to improve the performance of DR detection. Shakibania et al. [14] employed pre-trained dual-branch ResNet50 and EfficientNetB0 models, using CLAHE as part of their augmentation methods. When tested on the APTOS 2019 Dataset, the models achieved accuracies of 97.95% for binary classification and 83.17% for multi-class classification. To address the issue of data imbalance, they also added additional 1,000 images from two other public datasets to increase training set for minority classes: Stage 1 (mild DR), 3 (severe DR), and 4 (proliferative DR). This approach

further improved the multi-class classification accuracy to 89.6%. Sikder et al. [15] highlighted that the APTOS 2019 Dataset contains many noisy and duplicate images labeled under different classes, prompting them to handpick and discard 566 images. They reported that this image exclusion process boosted accuracy from 91.8% to 94.2%.

Although significant progress has been made in automatic DR detection using machine learning, several research gaps remain: 1) a lack of systematic studies on applying image processing to DR images to enhance features for better DR detection performance; 2) insufficient research on optimizing image processing techniques in alignment with deep learning architectures; and 3) a lack of practical applications capable of detecting retinal lesions. Addressing these challenges is critical, as enhanced image processing can significantly impact the reliability and accuracy of DR detection systems, particularly in real-world clinical settings. Additionally, improving the alignment between image processing techniques and deep learning models is essential for achieving more robust and generalizable solutions. Lastly, the development of practical applications for detecting retinal lesions will bridge the gap between research and clinical practice, providing valuable tools for early diagnosis and treatment. To tackle these challenges, this study focuses on the exploration of retinal fundus image processing techniques to enhance image quality and improve the accuracy of DR classification using deep learning models. The key contributions of this study are as follows: 1) Recent popular image processing techniques, CLAHE, Gaussian Blur, and ESRGAN, have been systematically studied. Their combinations, with fine-tuned parameters, are proposed for the first time to reduce image noise, enhance contrast, restore blurriness, and highlight DR lesion features, thereby improving multi-DR stage classification performance. 2) This study integrates pre-trained ResNet50 and VGG19 deep learning frameworks with several proposed image preprocessing methods, achieving performance superior to the state-of-the-art methods in prior research. 3) The proposed image enhancement methods are suggested for integration into retinal camera applications, making retinal lesions more visible to the human eye.

III. METHODOLOGY

A. Dataset

This study utilized the publicly accessible APTOS 2019 dataset, shown in Fig. 2. The dataset comprises 3,662 retinal fundus images, each sized 3216×2136 pixels, collected from 193 patients using various cameras in different locations. The distribution of images across the DR stages is as follows: 1805 in the no DR group, 370 in the mild DR group, 999 in the moderate DR group, 193 in the severe DR group, and 295 in the proliferate DR group. The dataset is imbalanced with an uneven distribution among the classes, which can negatively impact classification performance. To mitigate this, data augmentation techniques were applied.

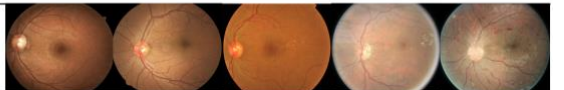
Stage	0 - No DR	1 - Mild	2 - Moderate	3 - Severe	4 - Proliferative
Count	1805	370	999	193	295
Image					

Fig. 2. Sample retinal images of each DR stage in the APTOS 2019 Dataset.

B. Proposed Image Processing

In this study, the combinations of the three image preprocessing techniques were fine-tuned and proposed—Graham method [9][10], CLAHE [11], and ESRGAN [16]—to address issues such as blurriness, low contrast, and inhomogeneous illumination present in the APTOS dataset.

The Graham method was proposed by Ben Graham, the winner of the Kaggle DR grading competition, to help remove image variations due to different lighting conditions or imaging devices. In this work, the Graham method was implemented with OpenCV functions:

```
cv2.addWeighted(image,4,cv2.GaussianBlur(image,(0,0),sigma),-4,128)
```

The role of sigma in the Gaussian filter is to control the variation around its mean value. First, Gaussian Blur was applied with Gaussian kernel size= (0, 0) and sigma= 10, namely Graham10 in this study. Second, the blurred image was blended with the original image with alpha = 4, beta = -4 and gamma = 128. As shown in Fig. 3(b), applying the Graham10 method effectively removes lighting noise from the original image. However, this process also diminishes the

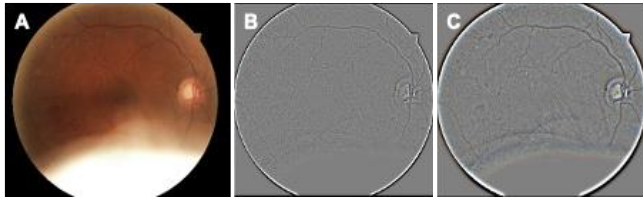


Fig. 3. (a) Original image with lighting noise; (b) Graham method with sigma= 10 applied to the original image; and (c) Graham method with sigma= 20 applied to the original image.

visibility of DR lesions. When the sigma is increased to 20 (Fig. 3(c)), referred to as Graham20 in this study, the vessels, microaneurysms and exudates are more prominently highlighted, making them more visible to the human eye.

Contrast Limited Adaptive Histogram Equalization (CLAHE) is a contrast enhancement method based on Histogram Equalization (HE). While HE increases image contrast by spreading out the most frequently occurring intensity values in the histogram, it also tends to amplify noise. CLAHE is a variant of HE designed to improve image contrast without excessively amplifying noise. In this study, CLAHE was implemented using the OpenCV function with the tile grid size (8, 8) and the clip limit, 2.0.

Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN), proposed by X. Wang et al. [16], were trained with synthetic data to enhance details while removing noisy artifacts to restore blurry images and videos in very high resolution. In this study, ESRGAN was selected for denoising and restoring the quality of blurry retina fundus images. The combinations of the above three image processing methods were experimented to compare the effectiveness of preprocessing on DR classification.

The combination of the above three methods are experimented to compare the effectiveness of preprocessing on DR detection and classification. Fig. 4 shows the effect of preprocessing methods experimented in this study: Graham10, Graham20, CLAHE, ESRGAN, and their combinations. ESRGAN + CLAHE (Fig. 4(f)) means images are first denoised using ESRGAN and the contrast was increased by CLAHE. Graham10 (or Graham20) + CLAHE (Fig. 4(d) and 4(e)) means images were blurred and denoised using the Graham method with Sigma= 10 (or 20) and then

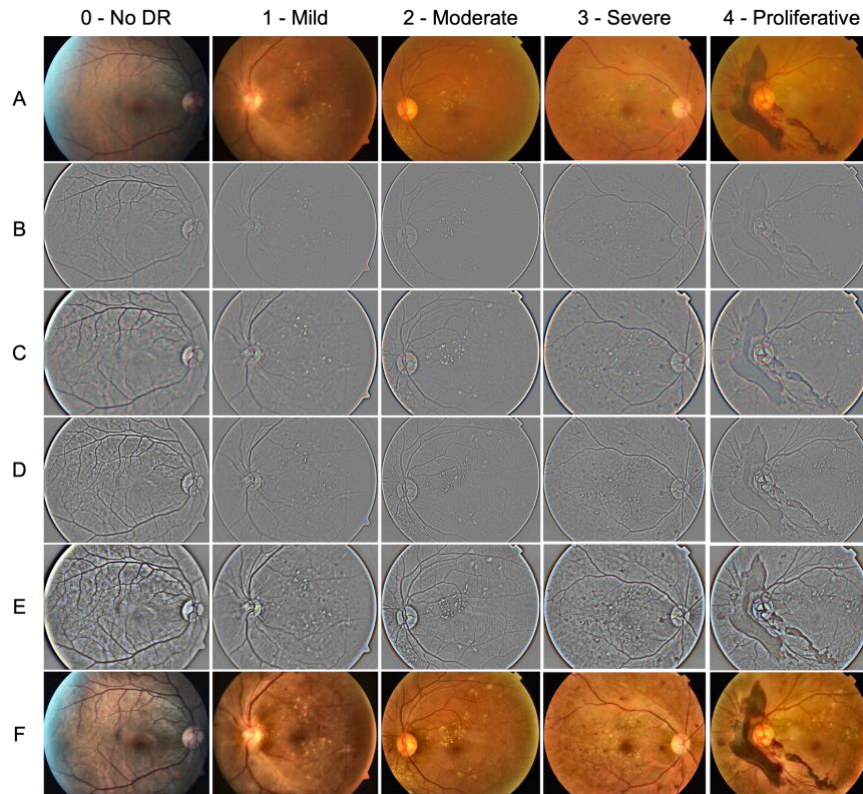


Fig. 4. Images in five DR stages before and after applying image preprocessing methods: (a) Original; (b) Graham10; (c) Graham20; (d) Graham10 + CLAHE; (e) Graham20 + CLAHE; and (f) ESRGAN + CLAHE.

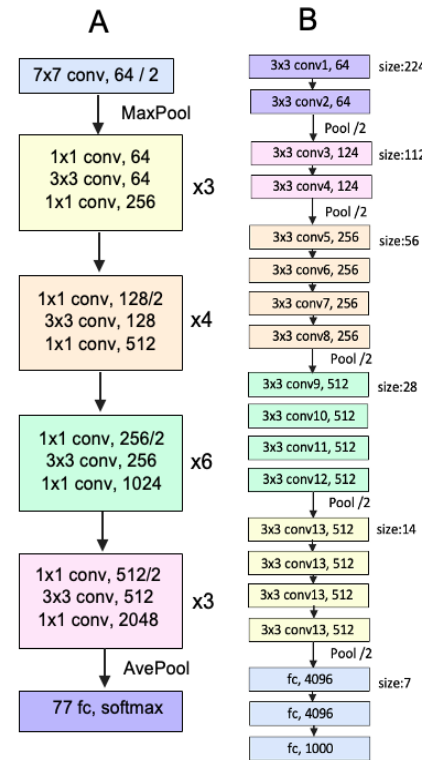


Fig. 5. Diagram of deep learning architectures (a) ResNet50 and (b) VGG19.

CLAHE is applied to enhance the image contrast. Blurring was used before CLAHE so that CLAHE does not enhance the unwanted noises in the images. It is clearly demonstrated that Graham20 + CLAHE highlights the DR lesions more clearly to human eyes and the reduced color contrast is beneficial to increase deep learning accuracy.

C. Convolutional Neural Network and Transfer Learning

In this study, transfer learning method was utilized by using the pre-trained weights of two deep learning models, ResNet50 and VGG19, based on their abilities to handle complex and diverse DR lesions such as microaneurysms, hemorrhages, and exudates. Fig. 5(a) shows the architecture of ResNet50, which is a deep learning model given by Microsoft [17] and has the quality feed output of some layer directly to the input of some other layer by bypassing layers in between. This property, known as identity mapping, helps address the vanishing gradient problem. VGG19 is the model given by the “Visual Graphic Group” at Oxford University [18]. Fig. 5(b) shows the architecture of VGG19 which contains 19 layers, including 16 convolutional layers and 3 fully connected layers. It is characterized by the consistent use of 3×3 convolutional filters with a stride of 1 and max-pooling layers.

D. DR Image Preprocessing and Classification Flow

The APTOS 2019 dataset was utilized to create an automatic DR classification model. The image preprocessing and DR classification workflow are presented in Fig. 6. During the preprocessing step, the smallest rectangle containing the entire field of the retinal image was identified and cropped. The cropped image was padded into a black square to maintain the original image aspect ratio and then preprocessed with one of the image preprocessing techniques described in the previous subsection. Subsequently, the image was resized to 280×280 pixels and center-cropped into 224×224 pixels to remove the black background as much as possible while retaining most of the retinal lesions. The images were then converted into a PyTorch tensor, transforming the Python Image Library (PIL) image with a pixel range of $[0, 255]$ into a PyTorch Tensor with a range $[0.0, 1.0]$. To avoid overfitting due to class imbalance, the minority data was augmented through random horizontal flips, vertical flips, and rotations until it matched the size of the majority data. Next, all images were split into 90% for training and 10% for testing. The RGB channels of the training images were normalized using z-score normalization, with the mean and standard deviations calculated from the entire preprocessed training set. After normalization, the mean and standard deviation of the RGB channels for all training images

were 0.0 and 1.0 respectively. Normalization ensures that the data falls within a specific range and reduces the skewness, which helps to accelerate and improve the learning process.

Ten percent of the training images were randomly selected as the validation set to assess the performance and retain the optimal weight coefficients. Both the Adam and SGD optimizers were simulated and compared, with Adam ultimately chosen for its superior accuracy. A weighted decay of 0.0001 was applied for regularization. The loss was calculated using categorical cross-entropy. All models were trained with a fixed random seed for fair comparison. The model was trained for 50 epochs using 32 batch size and the adaptive learning rate, starting at 0.0003, was gradually reduced to $1e-5$. Weights were stored after each epoch and saved. After the training of the dataset, the validation dataset was used to evaluate the trained model. After the training process was completed, the model was tested on the test images to classify the severity of DR. In this study, PyTorch 1.7.1 was used for the data loader and normalization; skorch-0.15.0 was used for splitting the data; Scikit-learn was used for performance metrics calculation; and Smote was used for data augmentation. The training set size increased from 3,296 to 8,075 images after oversampling, with 1,615 samples in each class. The models were trained on Google Colab, using the V100 GPU hardware accelerator.

IV. RESULTS AND DISCUSSION

A. DR Detection and Five Stages Classification

To evaluate the performance of the proposed framework, the trained models were tested using 366 images as follows: 190 in stage 0, 40 in stage 1, 86 in stage 2, 19 in stage 3, and 31 in stage 4. All models developed in this study were assessed using standard performance metrics for image classification, including accuracy, precision, recall, and F1-score. Table I represents the preprocessing performance comparison for DR detection. The combination of CLAHE + ESRGAN, when trained with VGG19, achieved the highest detection accuracy of 98.91%. The other preprocessing methods also yielded accuracies close to 98%.

Table II shows the performance metrics for DR severity level classification using different image preprocessing techniques. The imbalanced nature of the training dataset posed a challenge in assessing performance using the macro average of these metrics, as discussed in [14][19]. Since the dataset is imbalanced, the weighted average of precision, recall and F1-score was calculated based on the test dataset size. This approach accounts for the contribution of each class to the overall model performance, avoiding bias towards

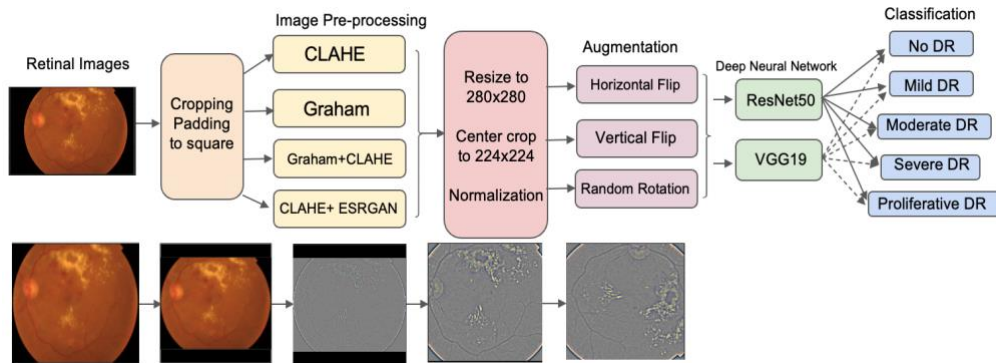


Fig. 6. Image preprocessing and DR Classification Flow.

TABLE I. PERFORMANCE COMPARISON OF MODELS TRAINED ON DIFFERENT PREPROCESSED IMAGES FOR BINARY CLASSIFICATION

Image Enhancement	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
CLAHE	ResNet50	97.54	96.89	98.42	97.65
	VGG19	98.09	97.91	98.42	98.16
Graham10	ResNet50	98.09	99.46	96.84	98.13
	ResNet50	98.36	98.94	97.89	98.41
CLAHE + ESRGAN	ResNet50	97.81	98.92	96.84	97.87
	VGG19	98.91	99.47	98.42	98.94
Graham10 + CLAHE	ResNet50	97.36	98.91	95.79	97.33
Graham20 + CLAHE	ResNet50	98.63	98.94	98.42	98.68
	VGG19	97.37	97.37	97.37	97.37

TABLE II. PERFORMANCE COMPARISON OF MODELS TRAINED ON DIFFERENT PREPROCESSED IMAGES FOR MULTI-CLASS CLASSIFICATION

Image Enhancement	Model	Optimizer	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
CLAHE	ResNet50	Adam	83.88	83.21	83.88	83.11
	VGG19	Adam	83.88	83.64	83.88	83.50
Graham10	ResNet50	Adam	83.61	84.16	83.16	83.57
Graham20	ResNet50	Adam	82.79	82.77	82.79	82.67
CLAHE + ESRGAN	ResNet50	Adam	84.15	84.40	84.15	81.13
	VGG19	Adam	83.88	84.49	83.88	82.91
Graham10 + CLAHE	ResNet50	Adam	80.87	81.07	81.87	80.69
Graham20 + CLAHE	ResNet50	Adam	86.07	85.89	86.07	85.89
	VGG19	SGD	77.23	76.93	77.23	73.12
		Adam	81.69	83.18	81.69	79.64

larger or smaller classes. It provides a more accurate assessment of the model's effectiveness in detecting and grading DR.

This study explores the role of various image processing methods in DR classification, testing different combinations to achieve the best performance of the deep learning model. It was found that two combinations performed the best in DR classification. ResNet50 with Graham20 + CLAHE achieved the highest accuracy of 86.07%, while ResNet50 with CLAHE + ESRGAN showed the second highest accuracy of 84.15%. The pre-trained ResNet50 model performed approximately 2% more accurately than the pre-trained VGG19 model in capturing the distinctive features of the retinal fundus images for five-DR-stage classification. Table II also indicates that Adam was a superior training optimizer compared to SGD, offering faster convergence and approximately 8.84% higher accuracy. Therefore, the Adam optimizer was used throughout the rest of deep learning modeling.

The selected confusion matrices (CM) based on the ResNet50 model using four different preprocessing methods are demonstrated in Fig. 7. The scale of each entry in the matrices is the number of the images. It is evident that Graham20 + CLAHE in Fig. 7(c) produced higher accuracy, with fewer images in Stage 1 and 2 being incorrectly misclassified relative to each other and fewer images in Stage 4 being misclassified as Stage 2. In Fig. 7(a) and 7(b), many images in proliferative DR were misclassified as moderate DR due to Graham's Gaussian Blur setting making fibrous proliferations and vitreous hemorrhages less visible. The model performance results suggest that increasing Sigma from 10 to 20 plus the contrast enhancement by CLAHE helped highlight the lesions in proliferative DR. From the results of the confusion matrix in Fig. 7(d). CLAHE + ESRGAN performed better in detecting proliferative DR, but performs poorly in differentiating between Stage 1 and Stage 2.

The image processing techniques can be customized for the classification of each DR stage. Table III shows the accuracies achieved by the ResNet50 model using the different image processing techniques in this study for each DR stage. The CLAHE + Graham20 method achieves the best

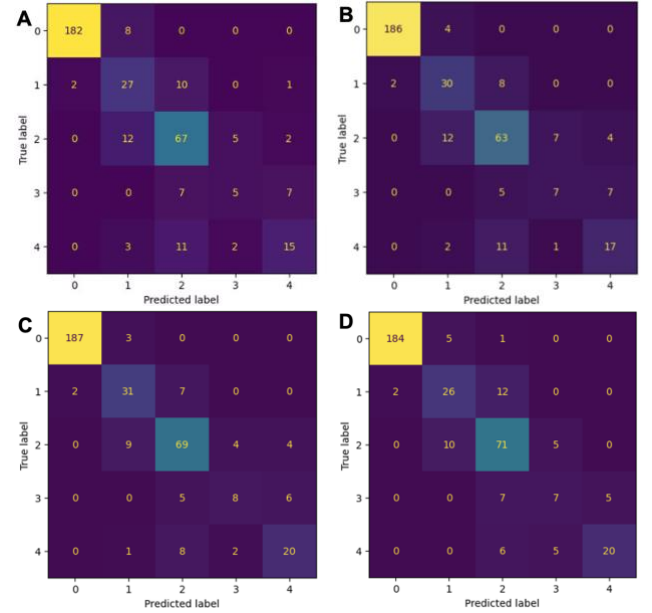


Fig. 7. Confusion Matrices of DR multi-classification using ResNet50 with two different preprocessing methods: (a) Graham10 + CLAHE, (b) Graham20, (c) Graham20 + CLAHE, and (d) CLAHE + ESRGAN.

TABLE III. ACCURACY COMPARISON OF RESNET50 MODEL USING DIFFERENT IMAGE PROCESSING TECHNIQUES FOR DIFFERENT DR STAGES

DR Stage	CLAHE (%)	Graham10 (%)	Graham20 (%)	CLAHE + Graham10 (%)	CLAHE + Graham20 (%)	CLAHE + ESRGAN (%)
0: No DR	98.54	98.09	98.36	98.36	98.63	97.81
1: Mild DR	93.08	91.80	92.35	92.44	93.99	92.08
2: Moderate DR	86.89	87.70	87.16	86.61	89.89	88.80
3: Severe DR	94.17	95.63	94.54	93.99	95.36	93.99
4: Proliferative DR	93.81	93.99	93.17	93.99	94.26	95.63

performance for classifying Stages 1, 2, and 3. However, Graham10 is the most effective for classifying Stage 3, while CLAHE + ESRGAN excels in classifying Stage 4.

B. Comparison with Previous Works

Although there were noisy and duplicated images as mentioned in [15], no images were discarded in this study. The performance of the proposed classification framework was compared with previous studies using the full APTOS dataset from the past six years. Table IV shows that the deep learning model performance is highly dependent on the image preprocessing methods. In comparison with [9] using ResNet50 as the feature extractor and Graham10 as the preprocessing method, our ResNet50 model, incorporating the proposed Graham20 + CLAHE preprocessing method, increased accuracy from 74.64% to 86.07%. Similarly, when compared with [11][14] using CLAHE, our ResNet50 model with the CLAHE + ESRGAN preprocessing method achieved a comparable accuracy. Our proposed Graham20 + CLAHE method further improved accuracy by an additional 2-3%.

However, the proposed model has several limitations: 1) The APTOS 2019 contains only 3662 images, with 566 images identified as noisy or duplicated with different labels [15]. 2) The dataset is imbalanced. Each of the mild, severe, and proliferative DR classes have only 200-300 images, which is insufficient to build a robust classifier for the category. 3) The current model is limited to ResNet50 and VGG19, so other CNN models should be explored. 4) The current model's

performance is based on APTOS dataset. It should be validated through other dataset.

TABLE IV. CLASSIFICATION PERFORMANCE COMPARISON WITH PREVIOUS RESEARCH USING APTOS 2019 DATASET

Ref	Year	Method	Preprocessing	Accuracy (%)	Precision (%)	Recall (%)
Binary Classification						
[7]	2021	Hybrid of VGG16 + capsule networks	None	95.5		
[8]	2022	VGG16	Grad-CAM	73.04	-	-
		DenseNet121	Grad-CAM	72.95	-	-
[11]	2022	Supervised contrastive learning	CLAHE	98.36	98.36	98.37
[14]	2024	Hybrid ResNet50 + EfficientNetB0	CLAHE	97.95	97.84	98.11
This Work	2024	ResNet50	Graham20+CLAHE	98.63	98.94	98.42
		VGG19	CLAHE+ESRGAN	98.91	99.47	98.42
Multi-class classification						
[7]	2021	Hybrid of VGG16 + capsule networks	None	75.81		
[18]	2020	NASNet + v-SVM	None	77.90	76.00	77.00
[9]	2019	ResNet50	Graham10	74.64	-	56.52
		Modified Xception	Graham10	83.09	-	88.24
[8]	2022	VGG16	Grad-CAM	48.43	-	-
[11]	2022	Supervised contrastive learning	CLAHE	84.36	70.51	73.84
[14]	2024	Hybrid ResNet50 + EfficientNetB0	CLAHE	83.17	82.66	83.17
This Work	2024	ResNet50	Graham20+CLAHE	86.07	85.89	86.07
		ResNet50	CLAHE+ESRGAN	84.15	84.40	84.15

V. CONCLUSION

Both deep learning models and proper image preprocessing are crucial for accurately detecting and grading DR. This paper systematically studied several image processing methods: 1) Graham Gaussian Blur to denoise the lighting noise, 2) CLAHE to enhance contrast, and 3) ESRGAN to restore the blurriness. The combinations of these three fine-tuned methods were explored, and the optimal settings were proposed. For early DR detection, the pre-trained VGG19 model with the proposed CLAHE + ESRGAN method achieved an accuracy of 98.91%. To differentiate retinal lesions across different DR stages, ResNet50 was proven to be more effective than VGG19 in capturing distinctive features. The ResNet50 model with Graham20 + CLAHE proposed achieved the highest accuracy of 86.07% for five-stage classification, detecting DR stages 0 through 4 with test accuracies of 98.63%, 93.99%, 89.89%, 95.36%, and 94.26%, respectively. These demonstrated accuracies outperform state-of-the-art methods utilizing APTOS dataset.

This study is current ongoing in two directions. First, in developing a personalized image enhancement system for DR detection, the proposed image enhancement methods can be integrated into an app of ophthalmoscope or fundus cameras in desktop, handheld, or smartphones. Once retinal fundus images are captured, users or ophthalmologists at home or remote locations can select between different image enhancement methods in the app to highlight the DR lesions for self-screening or professional DR screening. Graham20 + CLAHE is generally suitable for all DR stages, effectively removing lighting noise and making retinal lesions more visible to the human eye. The second area of focus is the automated DR classification system. The established ResNet50 deep learning framework can be integrated with electronic health systems in hospitals and clinics to assist medical professionals in diagnosing DR stages so patients can receive appropriate treatment timely.

REFERENCES

- [1] Division of Diabetes, Centers for Disease Control and Prevention: Available online: <https://www.cdc.gov/diabetes/library/socialmedia/infographics/diabetes.html> (accessed on 12 July 2023).
- [2] IDF Diabetes Atlas. Available online: <https://diabetesatlas.org/atlas/ninth-edition> (accessed on 12 July 2023).
- [3] Division of Diabetes, Centers for Disease Control and Prevention: Available online: <https://www.cdc.gov/diabetes/diabetes-complications/diabetes-and-vision-loss.html> (accessed on 13 July 2023).
- [4] APTOS 2019 Blindness Detection, Available online: <https://www.kaggle.com/c/aptos2019-blindness-detection> (accessed on 13 November 2022).
- [5] Z. Yang, T. Tan, Y. Shao, T. Wong, X. Li, "Classification of diabetic retinopathy: Past, present and future," *Front Endocrinol (Lausanne)*, 2022 Dec 16;13:1079217.
- [6] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, F. Li, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision* 115, 2015.
- [7] G. Kumar, S. Chatterjee, C. Chattopadhyay, "DRISTI: a hybrid deep neural network for diabetic retinopathy diagnosis," *Signal, Image and Video Processing*. 2021;15(8):1679-1686.
- [8] H. Alghamdi, "Towards Explainable Deep Neural Networks for the Automatic Detection of Diabetic Retinopathy," *Applied Science* 2022, 12, no 19, 9435.
- [9] S. H. Kassani, P. H. Kassani, R. Khazaeinezhad, M. J. Wesolowski, K. A. Schneider, R. Deters, "Diabetic retinopathy classification using a modified xception architecture," *2019 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)* (2019): 1-6
- [10] B. Graham, "Kaggle diabetic retinopathy detection competition report 2015," Available online: <https://www.kaggle.com/competitions/diabetic-retinopathy-detection/discussion/15801> (accessed on 20 September 2023)
- [11] M. Islam, L. Abdulrazak, M. Nahiduzzaman, M. Goni, M. Anower, M. Ahsan, J. Haider, M. Kowalski, "Applying supervised contrastive learning for the detection of diabetic retinopathy and its severity levels from fundus images," *Comput Biol Med.* 2022 Jul;146:105602.G. Nneji, J. Cai, J. Deng, H. Monday, M. Hossin, S. Nahar, "Identification of Diabetic Retinopathy Using Weighted Fusion Deep Learning Based on Dual-Channel Fundus Scans," *Diagnostics* 2022, 12, no 2, 540.
- [12] G. Nneji, J. Cai, J. Deng, H. Monday, M. Hossin, S. Nahar, "Identification of Diabetic Retinopathy Using Weighted Fusion Deep Learning Based on Dual-Channel Fundus Scans," *Diagnostics* 2022, 12, no 2, 540.
- [13] EYEPACS Kaggle DR Dataset. Available online: <https://www.kaggle.com/c/diabetic-retinopathy-detection/data> (accessed on 22 November 2021).
- [14] H. Shakibania, S. Raoufi, B. Pourafkham, H. Khotanlou and M. Mansoorizadeh, "Dual Branch Deep Learning Network for Detection and Stage Grading of Diabetic Retinopathy", *Biomedical Signal Processing & Control journal*, 2024.
- [15] N. Sikder, M. Masud, A. Bairagi, A. Arif, A. Nahid, H. Alhumyani, "Severity Classification of Diabetic Retinopathy Using an Ensemble Learning Algorithm through Analyzing Retinal Images," *Symmetry* 13: 670, 2021
- [16] X. Wang, L. Xie, C. Dong, Y. Shan, "Real-esrgan: Training real-world blind super-resolution with pure synthetic data," *Proceedings of the IEEE/CVF international conference on computer vision*, 2021.
- [17] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778, 2019.
- [18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, abs/ 1409.1556, 2017.
- [19] J. S. Akosa, "Predictive accuracy: A misleading performance measure for highly imbalanced data," *SAS Global Forum* 942, 2017.
- [20] V. Dondeti, J. D. Bodapati, S. N. Shareef, V. Naralasetti, "Deep convolution features in non-linear embedding space for fundus image classification," *Revue d'Intelligence Artificielle* 34,2020.