IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 3, June 2026, pp. 2554~2566
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i3.pp2554-2566

Image enhancement combined with EfficientNet-B7 for grading classification of diabetic retinopathy

Nina Sevani[1,2], Edy Kristianto[1,3], Albert Salomo[1]
[1]Department of Informatics, Faculty of Engineering and Computer Science, Krida Wacana Christian University, Jakarta, Indonesia
[2]Research Center for Knowledge Management and Collaborative Innovation, Krida Wacana Christian University, Jakarta, Indonesia
[3]Research Center for Connected Intelligence and Security, Krida Wacana Christian University, Jakarta, Indonesia

Article Info

Article history:

Received Oct 17, 2024
Revised Feb 9, 2026
Accepted Mar 5, 2026

Keywords:

Contrast limited adaptive histogram equalization
EfficientNet-B7
Fundus
Image segmentation
Real enhanced super-resolution generative adversarial network

ABSTRACT

Diabetic retinopathy (DR) is a complication caused by poorly managed diabetes that affects the eyes. According to the World Health Organization (WHO), 422 million people worldwide have suffered from DR in the past ten years. Manual detection using retinal fundus images is time-consuming and requires experienced ophthalmologists. This study proposes a deep learning method using the pre-trained model EfficientNet-B7 to identify this disease automatically. Five levels of DR will be classified: no-DR, mild-DR, moderate-DR, severe-DR, and proliferative-DR. The model was trained using "APTOS 2019 blindness detection" dataset, and image augmentation was performed. Image segmentation techniques such as contrast limited adaptive histogram equalization (CLAHE) and real enhanced super resolution generative adversarial network (Real-ESRGAN) were applied during preprocessing to improve the model's accuracy significantly. The implementation of CLAHE resulted in the validation accuracy improvement from 76.6% to 83.4% compared to no segmentation, while the combination of Real-ESRGAN and CLAHE increased the accuracy to 93.7%. Future research can explore the combination of CLAHE with other image processing techniques apart from the Real-ESRGAN model.

Corresponding Author:

Nina Sevani
Department of Informatics, Faculty of Engineering and Computer Science, Krida Wacana Christian University
Tanjung Duren Raya No. 4, West Jakarta, DKI Jakarta, Indonesia
Email: nina.sevani@ukrida.ac.id

1. INTRODUCTION

Diabetic retinopathy (DR) is the leading cause of blindness in persons of mature age. Work worldwide is responsible for more than 24,000 cases of blindness per year [1]. The World Health Organization (WHO) reported that 422 million people suffered from DR worldwide in the last ten years. In 2017, there were 425 million DR patients worldwide, with an estimation that 600 million people will have diabetes in the year 2040, with one-third of cases experiencing DR [2].

DR is a complication of type 1 and type 2 diabetes, primarily in patients with chronic hyperglycemia, poor glycemic control, and enhancement of blood pressure [2]. DR can classified into two stages, namely mild-DR, stage beginning DR, which is marked by existing microaneurysm (MA), proliferative-DR is stage second DR and can result in loss of vision [3]. Moreover, DR can increase the risk of damage, dysfunction, and failure in organs and body tissues, such as kidney disorders, heart disease, and the nervous system [2]. Screening regular treatment for DR cases is needed for early detection and management to prevent blindness [3].

Currently, ophthalmology diagnoses and assesses the severity of DR through visual evaluation with direct inspection and evaluation of the eye with the inspection fundoscopy [4]. This inspection examination and determination of the severity of DR is usually done by analyzing lesion features found on the patient's fundus image [5]. However, this method is time-consuming, labor-intensive, and prone to human error [5]. Considering funduscopic examinations like this still rely on ophthalmology expertise for interpretation, errors in DR detection results can occur [2]. It is crucial to improve the consistency of prediction results and the level of precision and efficiency in diagnosis DR, using the help of artificial intelligence (AI) technology.

Previous research shows that the rapid development of AI can produce a DR disease detection model that can provide results quickly with continuously improving levels of accuracy. Machine learning and deep learning are commonly used for DR detection. Generally, accuracy and precision are the factors most widely used to assess the performance of deep learning models in detecting DR. The use of technology like this can also help the task of ophthalmology in the early detection of the severity of DR disease in patients [6]. Ultimately, technology for detecting the severity of DR disease will help areas with limited numbers of ophthalmologists provide services to the community.

In detecting DR disease, severity level can be classified into several levels. Research by Qian *et al.* [7], which detects DR disease into five levels using a combination of convolutional neural network (CNN) architectures, namely Res2Net and DenseNet, also produces prediction accuracy as high as 83.2%. DenseNet-121 is also used in DR disease detection research, and producing an accuracy of 98.36% [8]. Other studies were carried out by [9] using the EfficientNet-B7 architecture to get a level of prediction accuracy as high as 89.1%. Meanwhile, other studies also try to compare the performance of several deep learning architectures in detecting DR disease, such as the comparison between ResNet-34, VGG-16, Inception-V3, and EfficientNetB4 [10] or comparing between ResNet-50, ResNet-152, and SqueezeNet1 [11].

Previous research also shows that other stages need to be carried out to produce a deep learning model with good accuracy performance. Good deep learning model performance is obtained not only by relying on the proper arrangement of convolution blocks and hyperparameters settings. Image quality data is vital and can be optimized by applying preprocessing techniques, such as image enhancement and segmentation. In processing an image, the preprocessing stage helps reduce complexity parameters and increase the accuracy of a model. In architecture deep learning, the most prominent features are selected to test each class on a dataset. Specifically, in processing image data, image enhancement, and segmentation are crucial for separating and extracting parts of the image.

In principle, DR is a condition that develops from untreated diabetes mellitus. The symptoms experienced by DR patients can vary depending on the severity of the condition. Patients with mild-DR may not experience any symptoms, making it difficult to realize the presence of the disease. As DR becomes more severe, patients may suffer from visual disturbances, ranging from blurred or distorted vision to complete loss of vision. The severity of DR can be identified by examining fundus images obtained through fundoscopic examination. Some parts of the fundus used to detect DR are MA, hemorrhages (HEM), hard exudates, and cotton wool spots (CWS) [12]. MA is an early sign of DR, typically identified by the presence of small, round, and dark red dots on the eye's fundus. HEM refers to bleeding that causes the red dot to grow larger with irregular edges. HEM is yellowish in appearance with an irregular and shiny circular ring shape. CWS are greyish-white images on retina of eye with irregular edges, resembling a roll of fluffy cotton [13].

Currently, research using fundus images to detect the severity of DR disease is mostly employed the deep learning technology combined with various other techniques. Previous research also demonstrated the use of various deep learning architectures for detecting the severity of DR disease with improved results over time. The deep learning architectures used include ResNet, DenseNet, EfficientNet, VGG, and Inception [7], [8], [10], [11], [14]–[18]. The use of a combination of two CNN-based architectures, namely ResNet and DenseNet, has also been carried out and gave the best results at a value of 83.2% [7]. Research using ResNet shows that ResNet-152 provides accuracy results of 94.40%, this value is higher than using ResNet-50 [11]. Other research using DenseNet121 also gave good results, reaching 98.36% [8]. The use of Inception-V3 also provides improved accuracy results when combined with preprocessing techniques such as contrast limited adaptive histogram equalization (CLAHE) showing an improvement in accuracy values of 4.4%. This value is better than when CLAHE was used together with EfficientNet-B4 which only provided an improvement in value of 2.3% [10]. Several previous studies also show that there are other deep learning architectures that are used to detect specific DR diseases based on HEM and MA [18], [19] and provide the best sensitivity reaching 94% [20].

EfficientNet [21] is an architecture deep learning that utilizes concepts of scaling up and scaling down on each neural network (NN) and produces a lightweight model and good performance. EfficientNet good performance is due to this architecture utilizing transfer learning techniques, which the model has already trained on datasets ImageNet. Currently, EfficientNet has developed from EfficientNet-B0 to EfficientNet-B7. EfficientNet-B7 makes use of block convolutions to capture crucial image features, resulting in improved accuracy of 84.4% and faster processing time [21].

Some researches show that deep learning architectures are combined with preprocessing methods or other methods to provide better accuracy. One of them is the use of K-means segmentation in architecture CNN in classifying cataract disease [22]. The research results show that segmentation, which is carried out as a preprocessing technique, apart from providing good accuracy, can also shorten the data time training models. There are also other studies that compare several preprocessing techniques and obtain that CLAHE works well to improve image quality fundus [23]. The application of CLAHE to increase accuracy in DR disease detection was also carried out by [10], [11], [24] with the highest accuracy obtained being 94.4% using ResNet-152. CLAHE can improve accuracy because it can increase contrast histogram equalization (HE) which will make the image more beautiful [25]. Unfortunately, increasing image contrast can also increase the contrast of other parts that are not necessary for DR detection. Therefore, other studies use a combination of CLAHE with enhanced super-resolution generative adversarial network (ESRGAN) for DR disease detection and provides the best accuracy results of 98.36% [8]. ESRGAN works to improve image texture to be more realistic and natural in an image [26].

Previous research shows that preprocessing techniques and deep learning architecture can provide good accuracy for DR disease detection. Moreover, there is still a chance for increased accuracy with faster preprocessing time. Therefore, this research will use the real enhanced super resolution generative adversarial network (Real-ESRGAN) technique, development of ESRGAN, to be combined with the EfficientNet-B7 architecture.

2. METHOD

The method used in this study begins with splitting the dataset into datasets for training, testing, and validation. Followed by preprocessing, namely the application of CLAHE, Real-ESRGAN techniques, and data augmentation, before being processed by EfficientNet-B7 model. There are five classes that describe severity of DR that will be detected by the model, as seen in Figure 1. The research workflow can be seen in Figure 2.

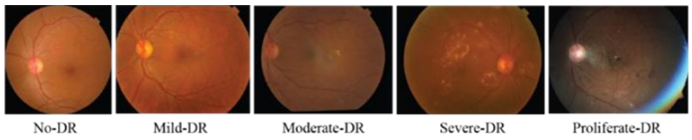

Figure 1. Image based on class label DR

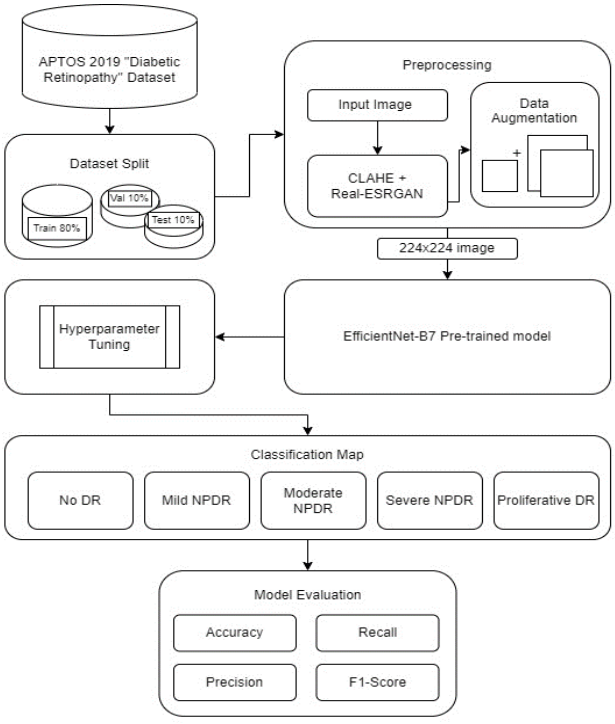

Figure 2. The flow of research work

2.1. Datasets

This study used the public dataset from the Kaggle [27], APTOS 2019 dataset, which is consist of images of DR disease categorized into 5 classes, totaling 3,662 images. The dataset includes PNG files and a CSV file that as the label for each image according to its class. These classes are numbered from 0 to 4 based on the severity of the DR disease. The images in the dataset can be viewed in Figure 1, and the distribution of images for each class label can be seen in Table 1.

The dataset will be divided into three pairs: 80% for training data, 10% for testing, and 10% for validation. Each data segment will be separated using the train_test_split method from the Sklearn library. We will use augmentation techniques to increase the number of images by flipping them vertically and horizontally, as well as scaling them up to 110%. Examples of the augmented images can be seen in Figure 3. The augmentation will ensure that the number of images for each class reaches 1,805, which is the highest number of images in the no-DR class. as the results of the data augmentation can be seen in Figure 4.

Table 1. Number of images based on DR class label

Class	Amount image
No-DR (Normal)	1,805
Mild-DR	370
Moderate-DR	999
Severe-DR	193
Proliferative-DR	295

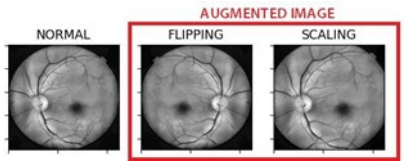

Figure 3. Augmented images in datasets

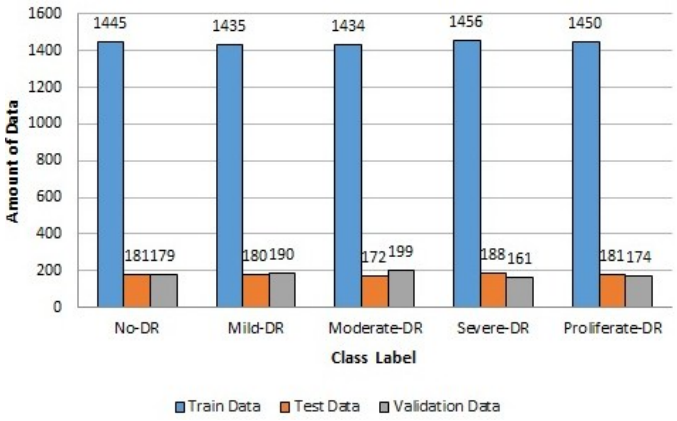

Figure 4. Comparison of the amount of augmented data

2.2. Pre-processing

Preprocessing is done to simplify the study images and achieve optimal results for APTOS 2019, as it has different image sizes. All images will be resized to 224×224 to match the input size of EfficientNet-B7 model and converted to grayscale. Additionally, preprocessing will involve increasing the image contrast using CLAHE. This will enhance the visibility of details in the fundus image for DR detection by increasing contrast and reducing image noise.

2.2.1. Contrast limited adaptive histogram equalization

CLAHE is used to enhance detail to a finer, more detailed textured level, and to increase low contrast from image DR using a distribution to mark brightness on the input image. The process of how CLAHE works to improve the quality of fundus images can be seen in Figure 5. The first stage is the tile

generation stage, where the input image is divided into four equally sized parts. Each part of the image goes through a HE processes, which involves six steps: processing histograms, calculating excess, distributing excess, redistributing excess, cumulative distribution function (CDF), and scaling and mapping processes. In each part of the image, the histogram values are clipped and propagated to the section histogram values of other parts of the image. Then, the histogram is calculated using the CDF per pixel scale of the input image, and mapping is executed based on the CDF values. To increase contrast, bilinear interpolation is used to combine each part of the image into one whole. This technique naturally increases the local contrast and makes the border lines and curves in the image clearer.

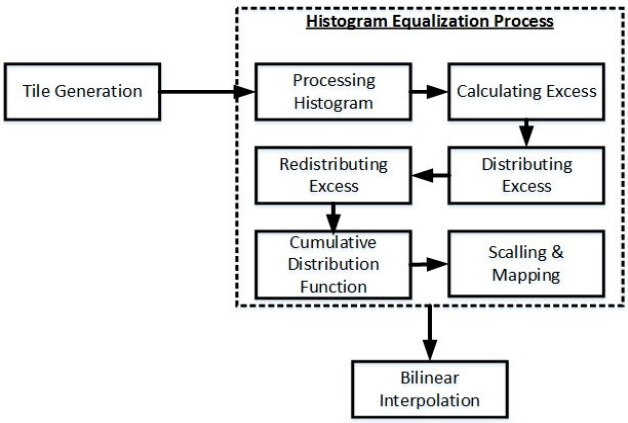

Figure 5. Workflow CLAHE

2.2.2. Real enhanced super resolution generative adversarial network

After applying CLAHE, the next step is to use Real-ESRGAN, which helps remove noise and enhance the quality of the fundus image, particularly at the edges and curves of the lines. Real-ESRGAN is based on the architecture of SRResNet and is an improvement of ESRGAN, integrating multiple residual-in-residual dense block (RRDB) elements. The use of Real-ESRGAN results in a more natural texture for the fundus image and enhances the visibility of details, which is crucial for medical images [28]. For more information on how Real-ESRGAN works, refer to Figure 6.

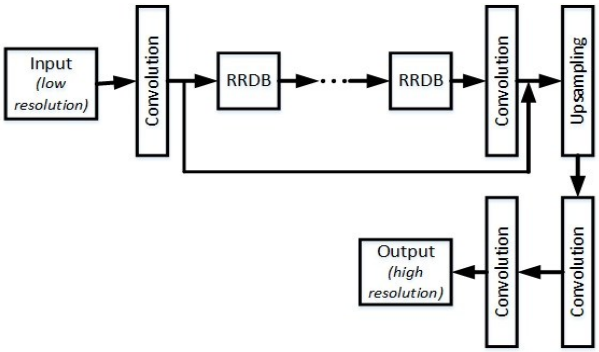

Figure 6. Workflow Real-ESRGAN

2.3. Experiment settings

For assessing the impact of the preprocessing technique, three experimental scenarios will be conducted as follows:

– The first scenario will involve no preprocessing and will use EfficientNet-B7.
– The second scenario will incorporate CLAHE with EfficientNet-B7.
– The third scenario will feature the combined use of Real-ESRGAN and CLAHE with EfficientNet-B7.

All experiments will use a Colab Pro CPU with NVIDIA A100 Tensor GPU, which has a RAM of 83.5 GB for the CPU and 40 GB for the GPU.

2.4. Evaluation performances

To measure the results of model evaluation, this research will use four metrics, namely accuracy, precision, recall or sensitivity, and F1-score. The formulas for these four metrics are written as in (1) to (4) [29].

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1-score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (4)$$

Where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

3. RESULTS AND DISCUSSION

3.1. Contrast limited adaptive histogram equalization segmentation

The CLAHE implementation followed the resizing of the fundus image to 224×224 and conversion to a single-channel grayscale format. Grayscaling was employed to reduce complexity and minimize model training time. Figure 7 displays a sample of the grayscaled APTOS 2019 dataset image results.

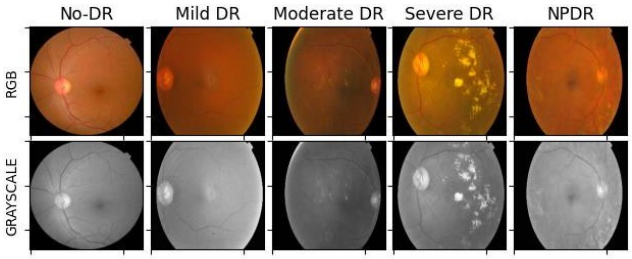

Figure 7. Color change to grayscaling

The implementation of CLAHE on the image tested the clip limit value. This clip limit controls how the image pixel intensity changes as the contrast changes. After testing clip limits ranging from 5 to 10, it was found that the optimal clip limit value is 8, as indicated in bold in Table 2. This value is crucial for providing the most suitable pixel intensity adjustment based on the image characteristic in APTOS2019. To view the results, you can refer to Figure 8, which shows the CLAHE implementation with a clip limit of 8 and a tile grid size with dimensions of 8×8.

Table 2. Clip limit comparison

Tile grid size	Clip limit	Accuracy			
		1st	2nd	3rd	Mean
8×8	5	0.781	0.770	0.753	0.768
	6	0.798	0.783	0.785	0.789
	7	0.804	0.827	0.811	0.814
	8	0.831	0.828	0.834	0.831
	9	0.822	0.818	0.809	0.816
	10	0.819	0.813	0.807	0.813

In Figure 8, it is evident that the use of CLAHE enhances the visibility of HEM details in the fundus image. Similarly, HE details also appear clearer due to increased contrast in the image. Additionally, it is apparent from Figure 8 that CLAHE can augment the contrast in fundus details with low-intensity variations.

CLAHE has also been employed in previous studies to diagnose DR disease using fundus images as part of preprocessing since it has its abilities to enhance image quality [11], [23], [24]. These findings demonstrate that for APTOS 2019, the implementation of CLAHE significantly improves visualization and aids in the detection of crucial features for identifying DR disease. The use of CLAHE, which increases accuracy, is consistent with earlier research [10], [23].

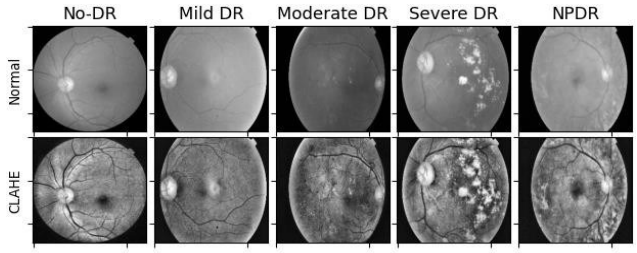

Figure 8. Results of applying CLAHE

3.2. Real enhanced super resolution generative adversarial network segmentation

Real-ESRGAN is employed following the application of CLAHE to enhance the pixel quality of the slightly distorted DR image, resulting in smoother texture and reduced noise. The process involves upscaling the 224×224 DR image to 784×784 (3.5× larger) and then resizing it back to 224×224 for the training model. Implementing CLAHE and Real-ESRGAN in each DR class is achievable, as demonstrated in Figure 9.

In Figure 9, it is apparent that the application of CLAHE and Real-ESRGAN can improve the texture of the fundus image, making low-intensity details such as HEM and HE more visible. Real-ESRGAN improves image texture, allowing low-intensity features to be more visible. This is where real-ESRGAN is employed on a wide range of images, not only medical ones [26], [28]. The overall image quality is also well preserved.

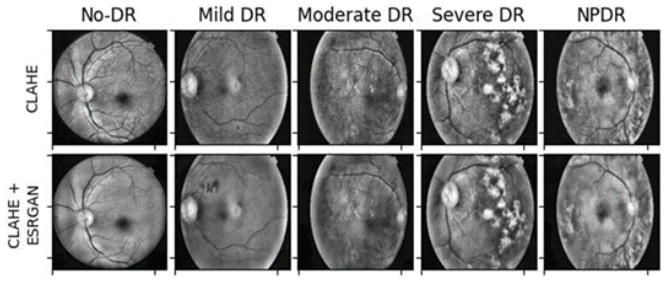

Figure 9. Results of applying CLAHE and real-ESRGAN

3.3. Performance evaluation

The training model was conducted using three different learning rate values: 0.0001, 0.001, and 0.01, with 50 epochs for each value. This value is calculated by taking into account the same values in previous research [7], [9], [10], so that a fairer analysis comparison can be made. The model was trained with the Adam optimizer at a learning rate of 1×10^{-4} and batch size of 32. To reduce overfitting, the optimizer used weight decay regularization with a decay value of 2×10^{-6}, calculated as the ratio of the learning rate to the total number of training epochs (50). To avoid overtraining, an early halting technique was used, which involved monitoring the validation loss over a 10-epoch period. When no improvement in validation loss was seen within this time frame, the training process was stopped, and the model weights corresponding to the lowest validation loss were restored.

The test results for the three scenarios are presented in Tables 3 to 5. The best accuracy, highlighted in bold, was achieved with a learning rate of 0.001. Additionally, it was observed that accuracy increased with the implementation of preprocessing techniques. Specifically, accuracy improved from 76.6% without preprocessing to 83.4% with CLAHE, and further to 93.7% with CLAHE and Real-ESRGAN. Subsequently, the chosen learning rate was used in further experiments to evaluate metric values. The precision, recall

(sensitivity), and F1-score results, achieved with a learning rate of 0.001 for each experimental scenario using EfficientNet-B7, are displayed in Figures 10 to 12.

Table 3. Accuracy results without CLAHE and Real-ESRGAN on EfficientNet-B7

Learning rate	Accuracy	Mean	Time execution (s)
0.00001	0.665	0.706	4,453.56
0.0001	0.687		4,488.16
0.001	0.766		4,473.56

Table 4. Accuracy results with CLAHE on EfficientNet-B7

Learning rate	Accuracy	Mean	Time execution (s)
0.00001	0.813	0.818	4,451.89
0.0001	0.808		4,468.23
0.001	0.834		4,469.01

Table 5. Accuracy results with CLAHE and Real-ESRGAN on EfficientNet-B7

Learning rate	Accuracy	Mean	Time execution (s)
0.00001	0.926	0.928	4,082.65
0.0001	0.921		4,079.74
0.001	0.937		4,086.55

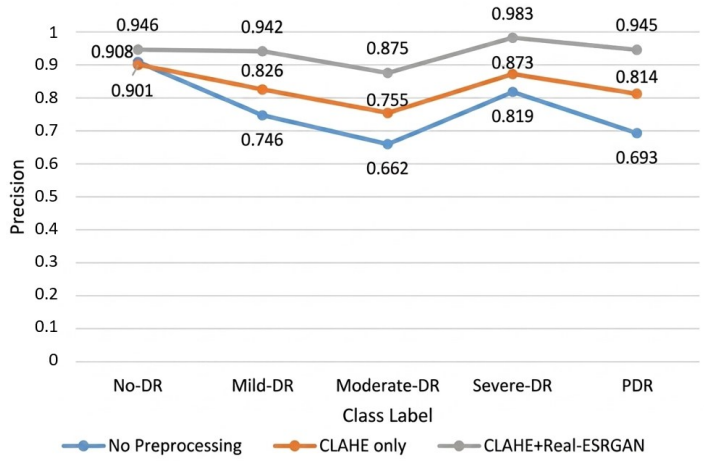

Figure 10. Comparison of precision obtained by EfficientNet-B7 under three experimental settings

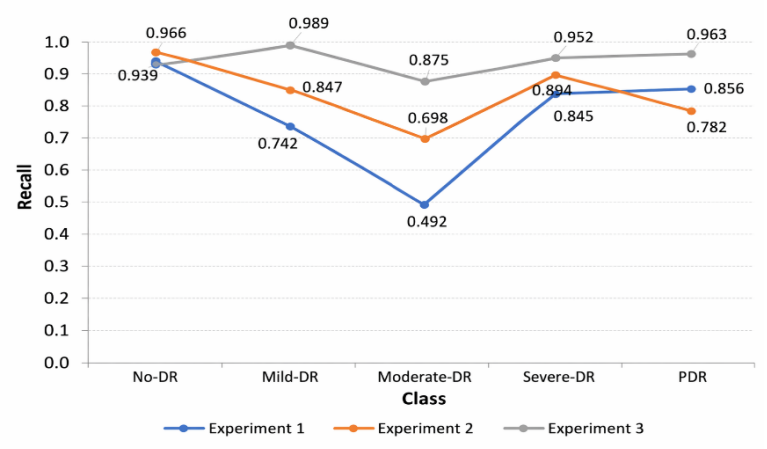

Figure 11. Comparison of recall obtained by EfficientNet-B7 under three experimental settings

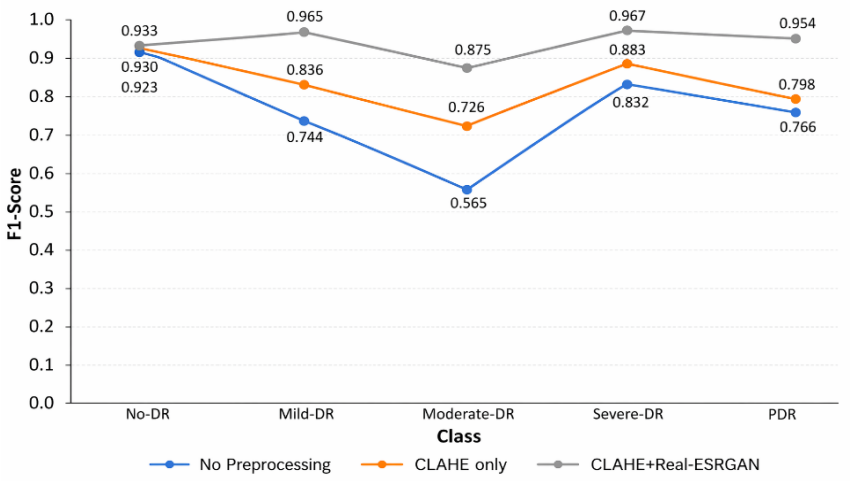

Figure 12. Comparison of F1-scores obtained by EfficientNet-B7 under three experimental settings

Based on the experimental results using the EfficientNet-B7 model, it's evident that applying CLAHE and Real-ESRGAN produces the best results. These results are in line with the results of previous research [8]. Although the previous research still used EfficientNet-B4 and only calculated accuracy. The accuracy value significantly increased from 0.706 to 0.928 after applying CLAHE and Real-ESRGAN. Similarly, in third experiment, which applied CLAHE and Real-ESRGAN, the precision, recall, and F1-score values were 0.983, 0.989, and 0.967, respectively, demonstrating a notable improvement compared to the first experiment that did not utilize preprocessing techniques and achieved values of 0.662, 0.492, and 0.565 for precision, recall, and F1-score. This indicates a 50% increase in the recall value after applying CLAHE and Real-ESRGAN. The significant improvement in values demonstrates the positive impact of implementing CLAHE and Real-ESRGAN on fundus images from the APTOS 2019 dataset, reflecting the potential to create a model with a good level of sensitivity.

Furthermore, the application of CLAHE and Real-ESRGAN yielded the best results in the severe-DR class, with a recall value of 0.989, while the lowest recall value was observed in the moderate-DR class, at 0.875. In the severe-DR class, the employment of CLAHE and Real-ESRGAN successfully smoothed the texture of the fundus image, revealing symptoms of disruption more clearly [26]. However, in the moderate-DR class, CLAHE and Real-ESRGAN alone were insufficient to identify patterns of disturbance clearly. Figures 8 and 9 show that the moderate-DR class had about the same number of indicators of disturbance in terms of place, shape, and color as the mild-DR class had. This caused the system to frequently make wrong predictions, not predicting the moderate-DR class. This demonstrates why the moderate-DR class has a lower recall value than the other classes. The moderate-DR class also showed the lowest recall value of 0.492 when CLAHE and Real-ESRGAN were not applied. This suggests that the moderate-DR class is the least influenced by the implementation of CLAHE and Real-ESRGAN. Visual inspection of fundus images in the moderate-DR class indicated no obvious signs of DR, and enhancing the image through contrast adjustment and noise reduction did not significantly alter the initial image. Details related to DR, such as HEM and HE, did not become clearer. Moreover, without applying CLAHE and Real-ESRGAN, the No-DR class exhibited the highest recall results, with a value of 0.939, slightly better than when CLAHE and Real-ESRGAN were applied, which achieved a recall value of 0.914.

In addition to the visual differences observed in the no-DR class, the application of augmentation techniques also influenced the recall results, although it had minimal impact on accuracy, precision, and F1-score metrics. These results indicate that the model requires a significant number of images to produce a sensitive model. Notably, the no-DR class, which did not undergo augmentation from the beginning, still yielded slightly better recall results than the class that underwent augmentation, both of which achieved values above 0.9. Furthermore, the importance of augmentation is evident through the high precision and F1-score results in the moderate-DR and severe-DR classes, as these classes experienced the most augmentation to incorporate images used in the model training process.

The experiments demonstrate that implementing CLAHE and Real-ESRGAN can enhance precision by more than 4% and increase recall by more than 3%. Balancing recall precision can optimize the performance of medical diagnostic tests. Enhanced precision and recall in a medical context indicate how effectively a model identifies true positive cases among all the positive cases it identifies. High precision is

crucial to minimize unnecessary treatments, reduce patient anxiety, utilize medical resources efficiently, and enable more informed clinical decision-making.

3.4. Baseline comparison

Inception-V3 will be chosen as the comparison model in this research. Experiments with Inception-V3 will be carried out following the hyperparameter settings that provide the most optimal results when using EfficientNet-B7, namely learning rates 0.001 and epochs 50. For CLAHE implementation, use clip limit 8 and tile grid size with size dimensions 8×8. There are three additional experimental scenarios carried out following the three main scenarios. The only difference is in the deep learning architecture used. These additional experiments are:

– The fourth scenario experiment is without preprocessing with Inception-V3,
– The second scenario experiment uses CLAHE with Inception-V3,
– The third scenario experiment uses Real-ESRGAN and CLAHE together with Inception-V3.

A comparison of classification report results of EfficientNet-B7 and Inception-V3 can be seen in Tables 6 to 8. Meanwhile, Figure 13 displays a comparison between the accuracy and time produced by the six scenarios carried out in this research.

Table 6. Comparison performance of EfficientNet-B7 and Inception-V3 without CLAHE and Real-ESRGAN

	EfficientNet-B7	Inception-V3
Precision	0.763	0.714
Recall	0.766	0.716
F1-score	0.760	0.713
Accuracy	0.766	0.714

Table 7. Comparison performance of EfficientNet-B7 and Inception-V3 using CLAHE

	EfficientNet-B7	Inception-V3
Precision	0.831	0.786
Recall	0.834	0.787
F1-score	0.832	0.786
Accuracy	0.834	0.787

Table 8. Comparison performance of EfficientNet-B7 and Inception-V3 using CLAHE and Real-ESRGAN

	EfficientNet-B7	Inception-V3
Precision	0.937	0.861
Recall	0.937	0.852
F1-score	0.937	0.852
Accuracy	0.937	0.852

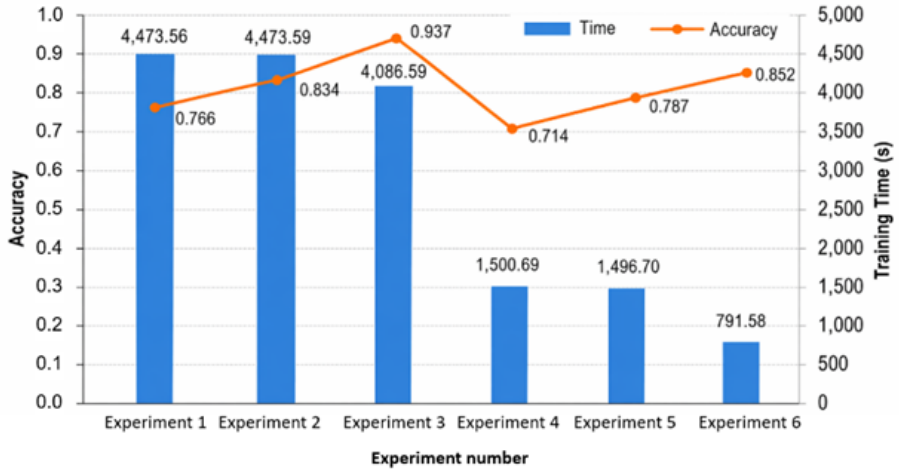

Figure 13. Comparison of time and accuracy of EfficientNet-B7 and Inception-V3

Statistical tests using McNemar were performed as well on both models within discussion [30], [31]. McNemar will assess the performance of EfficientNet-B7 and Inception-V3 at the 0.05 level of significance. This performance comparison statistical analysis has been carried out using Python with early stopping for each class, who had implemented CLAHE and Real-ESRGAN. The statistical test of model performance on the five classes revealed that EfficientNet-B7 outperformed Inception-V3 in almost all classes, with a single exception of moderate-DR. However, significant improvements were found solely in the severe-DR and proliferative-DR classes.

After comparing EfficientNet-B7 and Inception-V3 based on Tables 6 to 8, it has been proven that the EfficientNet-B7 model outperforms Inception-V3. While the results of this study differ slightly from those of similar studies [10], previous research only used CLAHE and EfficientNet-B4. The current study used a more advanced version of EfficientNet, specifically EfficientNet-B7, in conjunction with Real-ESRGAN and other hyperparameters. Although EfficientNet-B7's results are comparable to those of EfficientNet-B4, this study provides consistent outcomes in terms of precision and recall. The results' consistency of 93.7% suggests that the EfficientNet-B7 model, along with CLAHE and Real-ESRGAN, can be an acceptable predictor of DR diseases. EfficientNet-B7 consistently demonstrates better accuracy values compared to Inception-V3, whether using CLAHE and Real-ESRGAN or not, and whether using segmentation or not. However, it should be noted that the training time for EfficientNet-B7 is significantly longer than that of Inception-V3 due to its greater number of layers, necessitating a longer computing time. In terms of CLAHE and Real-ESRGAN processes, EfficientNet-B7 has shown superior results in classification reports compared to those without process segmentation.

The combination of EfficientNet-B7, CLAHE, and Real-ESRGAN yields high accuracy in medical image classification, which is crucial for ensuring effective and safe patient care. This leads to early and precise disease detection, improved treatment planning, optimized medical resources, and supports advancements in medical research. Additionally, it fosters trust in AI-based detection systems and enhances healthcare. Therefore, investing in computing resources and prioritizing high accuracy in medical imaging technologies is vital for improving overall healthcare outcomes.

4. CONCLUSION

This study applied CLAHE and Real-ESRGAN to the EfficientNet-B7 model for DR grading. The experimental results show that a learning rate of 0.001 yields the best performance for EfficientNet-B7. The use of CLAHE and Real-ESRGAN not only improves training efficiency but also significantly enhances model performance, achieving an accuracy of 93.7%, which is 18.25% higher than the model without these techniques. Improvements were consistently observed across recall, precision, and F1-score metrics. This performance also surpasses that of Inception-V3 with CLAHE and Real-ESRGAN, which achieved an accuracy of 85.2%. In addition, the proposed approach reduces training time for both EfficientNet-B7 and Inception-V3. Future research might look into image segmentation, other enhancement approaches, different deep learning architectures, and additional assessment metrics like kappa scores or area under the receiver operating characteristic curve (AUC-ROC) for a more comprehensive analysis.

FUNDING INFORMATION

The research was supported by a grant from Lembaga Penelitian dan Pengabdian Masyarakat from Krida Wacana Christian University No 07/UKKW/LPPM-FTIK/LIT/VI/2024.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Nina Sevani	✓	✓		✓	✓	✓		✓	✓	✓		✓	✓	✓
Edy Kristianto		✓	✓			✓	✓	✓	✓	✓	✓			
Albert Salomo	✓		✓	✓		✓	✓			✓	✓		✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [NS], on request.

REFERENCES

[1] L. F. Nakayama *et al.*, "Diabetic retinopathy classification for supervised machine learning algorithms," *International Journal of Retina and Vitreous*, vol. 8, no. 1, 2022, doi: 10.1186/s40942-021-00352-2.

[2] X. Wang, C. Qiu, X. Ren, Z. Xiong, V. C. M. Leung, and D. Niyato, *Integrating edge intelligence and blockchain*. in Wireless Networks. Cham, Switzerland: Springer International Publishing, 2022, doi: 10.1007/978-3-031-10186-1.

[3] L. Dai *et al.*, "A deep learning system for detecting diabetic retinopathy across the disease spectrum," *Nature Communications*, vol. 12, no. 1, 2021, doi: 10.1038/s41467-021-23458-5.

[4] U. Bhimavarapu and G. Battineni, "Deep learning for the detection and classification of diabetic retinopathy with an improved activation function," *Healthcare*, vol. 11, no. 1, 2023, doi: 10.3390/healthcare11010097.

[5] S. Vujosevic *et al.*, "Screening for diabetic retinopathy: new perspectives and challenges," *The Lancet Diabetes and Endocrinology*, vol. 8, no. 4, pp. 337–347, 2020, doi: 10.1016/S2213-8587(19)30411-5.

[6] F. Arcadu, F. Benmansour, A. Maunz, J. Willis, Z. Haskova, and M. Prunotto, "Deep learning algorithm predicts diabetic retinopathy progression in individual patients," *npj Digital Medicine*, vol. 2, no. 1, 2019, doi: 10.1038/s41746-019-0172-3.

[7] Z. Qian, C. Wu, H. Chen, and M. Chen, "Diabetic retinopathy grading using attention based convolution neural network," *IEEE Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, 2021, pp. 2652–2655, doi: 10.1109/IAEAC50856.2021.9390963.

[8] G. Alwakid, W. Gouda, and M. Humayun, "Enhancement of diabetic retinopathy prognostication using deep learning, CLAHE, and ESRGAN," *Diagnostics*, vol. 13, no. 14, 2023, doi: 10.3390/diagnostics13142375.

[9] A. E. Minarno, M. H. C. Mandiri, Y. Azhar, F. Bimantoro, H. A. Nugroho, and Z. Ibrahim, "Classification of diabetic retinopathy disease using convolutional neural network," *International Journal on Informatics Visualization*, vol. 6, no. 1, pp. 12–18, 2022, doi: 10.30630/joiv.6.1.857.

[10] M. Hayati *et al.*, "Impact of CLAHE-based image enhancement for diabetic retinopathy classification through deep learning," *Procedia Computer Science*, vol. 216, pp. 57–66, 2022, doi: 10.1016/j.procs.2022.12.111.

[11] T. M. Usman, Y. K. Saheed, D. Ignace, and A. Nsang, "Diabetic retinopathy detection using principal component analysis multi-label feature extraction and classification," *International Journal of Cognitive Computing in Engineering*, vol. 4, pp. 78–88, 2023, doi: 10.1016/j.ijcce.2023.02.002.

[12] Y. M. S. Reddy and R. S. E. Ravindran, "Retinal image lesions assisted diabetic retinopathy screening system through machine learning," *International Journal of Intelligent Engineering and Systems*, vol. 15, no. 2, pp. 175–189, 2022, doi: 10.22266/ijies2022.0430.17.

[13] M. N. Alsaleem and M. A. Berbar, "Literature review on the diabetic retinopathy in retinal images," *International Journal of Advanced Research in Computer Science and Electronics Engineering*, vol. 8, no. 11, pp. 50–60, 2019.

[14] Y. Li and L. Liu, "Image quality classification algorithm based on InceptionV3 and SVM," *MATEC Web of Conferences*, vol. 277, 2019, doi: 10.1051/matecconf/201927702036.

[15] M. T. Hagos and S. Kant, "Transfer learning based detection of diabetic retinopathy from small dataset," 2019, *arXiv: 1905.07203*.

[16] X. Zhang *et al.*, "Automated detection of severe diabetic retinopathy using deep learning method," *Graefe's Archive for Clinical and Experimental Ophthalmology*, vol. 260, no. 3, pp. 849–856, 2022, doi: 10.1007/s00417-021-05402-x.

[17] A. Bora *et al.*, "Predicting the risk of developing diabetic retinopathy using deep learning," *The Lancet Digital Health*, vol. 3, no. 1, pp. e10–e19, 2021, doi: 10.1016/S2589-7500(20)30250-8.

[18] D. U. N. Qomariah, H. Tjandrasa, and C. Fatichah, "Segmentation of microaneurysms for early detection of diabetic retinopathy using MResUNet," *International Journal of Intelligent Engineering and Systems*, vol. 14, no. 3, pp. 359–373, 2021, doi: 10.22266/ijies2021.0630.30.

[19] D. U. N. Qomariah, H. Tjandrasa, and C. Fatichah, "Exudate segmentation for diabetic retinopathy using modified FCN-8 and dice loss," *International Journal of Intelligent Engineering and Systems*, vol. 15, no. 2, pp. 508–520, 2022, doi: 10.22266/ijies2022.0430.45.

[20] R. Parmar and R. Lakshmanan, "Detecting diabetic retinopathy from retinal images using CUDA deep neural network," *International Journal of Intelligent Engineering and Systems*, vol. 10, no. 4, pp. 284–292, Aug. 2017, doi: 10.22266/ijies2017.0831.30.

[21] M. Tan and Q. V. Le, "EfficientNet: rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning, Long Beach, California, PMLR 97*, Long Beach, California, 2019, pp. 1-10.

[22] N. Sevani, H. Tampubolon, J. Wijaya, L. Cuvianto, and A. Salomo, "A study of convolution neural network based cataract detection with image segmentation," *IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*, pp. 216–221, 2022, doi: 10.1109/COMNETSAT56033.2022.9994549.

[23] S. K. Yadav, S. Kumar, B. Kumar, and R. Gupta, "Comparative analysis of fundus image enhancement in detection of diabetic retinopathy," *IEEE Region 10 Humanitarian Technology Conference 2016, R10-HTC 2016*, Agra, India, 2016, pp. 1-5, doi: 10.1109/R10-HTC.2016.7906814.

[24] S. Zhu, C. Xiong, Q. Zhong, and Y. Yao, "Diabetic retinopathy classification with deep learning via fundus images: a short survey," *IEEE Access*, vol. 12, pp. 20540–20558, 2024, doi: 10.1109/ACCESS.2024.3361944.

[25] A. Arulraj, J. S. Mariadhason, and R. R. Ronjalis, "An improved neighbourhood-based contrast limited adaptive histogram equalization method for contrast enhancement on retinal image," *International Journal of Ophthalmology*, vol. 18, no. 12, pp. 2225-2236, 2025, doi: 10.18240/ijo.2025.12.02.

[26] X. Wang *et al.*, "ESRGAN: enhanced super-resolution generative adversarial networks," in *Computer Vision – ECCV 2018 Workshops (ECCV 2018)*, 2019, pp. 63–79, doi: 10.1007/978-3-030-11021-5_5.

[27] Asia Pacific Tele-Ophthalmology Society, "APTOS 2019 blindness detection," *kaggle.com*. Accessed: Jan. 15, 2024. [Online]. Available: https://www.kaggle.com/competitions/aptos2019-blindness-detection
[28] F. Akhyar, E. N. Furqon, and C. Y. Lin, "Enhancing precision with an ensemble generative adversarial network for steel surface defect detectors (EnsGAN-SDD)," *Sensors*, vol. 22, no. 11, 2022, doi: 10.3390/s22114257.
[29] P. Ahammed, M. F. Faruk, N. Raihan, and M. Mondal, "Inception V3 based transfer learning model for the prognosis of acute lymphoblastic leukemia from microscopic images," *2022 4th International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE)*, Rajshahi, Bangladesh, 2022, pp. 1-4, doi: 10.1109/ICECTE57896.2022.10114522.
[30] Y. Xu *et al.*, "The diagnostic accuracy of an intelligent and automated fundus disease image assessment system with lesion quantitative function (SmartEye) in diabetic patients," *BMC Ophthalmology*, vol. 19, no. 1, 2019, doi: 10.1186/s12886-019-1196-9.
[31] Y. Papanikolaou, G. Tsoumakas, M. Laliotis, N. Markantonatos, and I. Vlahavas, "Large-scale online semantic indexing of biomedical articles via an ensemble of multi-label classification models," *Journal of Biomedical Semantics*, vol. 8, no. 1, 2017, doi: 10.1186/s13326-017-0150-0.

BIOGRAPHIES OF AUTHORS

Nina Sevani holds a Doctor of Computer Science degree from Universitas Indonesia, in 2023. She also received her bachelor degree in Computer Science from Bina Nusantara University, Indonesia, in 2002 and master degree in Computer Science from Institute Pertanian Bogor, Indonesia in 2009. She is currently an associate professor at Department of Informatics in Krida Wacana Christian University, Jakarta, Indonesia. Her research includes machine learning, data mining, transfer learning, deep learning, and health informatics. She can be contacted at email: nina.sevani@ukrida.ac.id.

Edy Kristianto received a Ph.D. degree in Computer Science from the National Chung Cheng University, Taiwan, in 2023. He also received a bachelor's degree in Informatics Engineering from Duta Wacana Christian University, Indonesia, in 2000 and a master's degree from Universitas Gadjah Mada, Indonesia, in 2007. He is an associate professor from Krida Wacana Christian University, Indonesia. His research interests include deep learning, cyber security, wireless communications, the internet of things, vehicular security technology, and cloud and edge computing. He can be contacted at email: edy.kristianto@ukrida.ac.id.

Albert Salomo graduated with a Bachelor of Computer Science (B.Sc.) from Krida Wacana Christian University in 2023 and has gained various professional certifications and skills. He currently serves as a software engineer at Krida Wacana Christian University, Jakarta, Indonesia. His research interest in artificial intelligence (AI), machine learning, and large language models (LLM). He can be contacted at email: albertsalomo.work@gmail.com.