An Enhanced Deep Learning Framework for Diabetic Retinopathy Classification Using Multiple Convolutional Neural Network Architectures

Zaid Romegar Mair[1], Agus Harjoko[2*], Rendra Gustriansyah[3],
Septa Cahyani[4], Rudi Heriansyah[5], Indah Permatasari[6], Muhammad HavizIrfani[7]
Informatics Engineering-Faculty of Computer and Natural Science,
Indo Global Mandiri University, Palembang, Indonesia[1,3,4,5,6,7]
Department of Computer Science and Electronics-Faculty of Mathematics and Natural Sciences,
Gadjah Mada University, Indonesia[2]

Abstract*—Diabetic retinopathy (DR) is a leading cause of blindness, requiring early and accurate diagnosis. Although deep learning, particularly Convolutional Neural Networks (CNNs), has shown promising results in automating DR classification, selecting the optimal architecture and extracting effective features for specific clinical datasets remains a challenge. This study aims to conduct a comprehensive performance evaluation of six CNN architectures—DenseNet121, MobileNet, NasNet_Mobile, ResNet50, VGG16, and VGG19—for DR classification on a dataset from the Community Eye Hospital of South Sumatra Province. The main novelty of our approach lies in a specific preprocessing workflow that integrates grayscale conversion and Canny edge detection to enhance the visibility of critical retinal features, such as blood vessels and lesions, before classification. Using a dataset of 3000 fundus images across five classes (No_DR, Mild, Moderate, Severe, and Proliferative DR), the model was trained with data augmentation and the Adam optimizer. Experimental results indicate that the VGG16 architecture achieves a peak accuracy of 73%, outperforming baseline implementations from previous studies. This study highlights the potential of combining classical CNN models with tailored preprocessing for improved DR detection, thus providing a benchmark for model selection on similar clinical datasets. These findings highlight the robustness and stability of VGG16, demonstrating its suitability as an early DR screening tool.

***Keywords*—Diabetic retinopathy; diabetic retinopathy classification; deep learning; Convolutional Neural Network (CNN); VGG16**

I. Introduction

According to the International Diabetes Federation (IDF), over 537 million adults are currently living with diabetes worldwide [1], and this number is projected to reach 700 million by 2045 [2]. Diabetic retinopathy can be a serious problem, potentially leading to permanent blindness if not identified or treated [3]. Because DR is often asymptomatic in its early stages, patients may not be aware of retinal degeneration until severe visual impairment develops [4]. PDR and NPDR are the two main categories recognized within the disease [5]. NPDR is characterized by microaneurysms, vascular leakage, and macular edema, while PDR involves the growth of abnormal blood vessels that may rupture and cause vitreous hemorrhage, potentially leading to total blindness.

Currently, ophthalmologists detect DR manually through fundus image examination. This process is time-consuming and prone to subjective errors, especially in early-stage DR detection [6]. Furthermore, due to the complex nature of retinal lesions and differences in image quality, existing automated systems still face challenges in achieving consistent accuracy [7].Since nearly all long-term diabetes patients are at risk of developing retinal microvascular complications [8], early screening and blood glucose control are essential preventive measures.

Convolutional neural networks (CNNs) have demonstrated significant effectiveness in identifying various medical conditions from images [10][11][12]. Deep learning (DL)-based diagnostic techniques have received considerable attention due to their ability to improve the accuracy and efficiency of medical image analysis by automatically extracting distinguishing features from retinal images [9]. Numerous studies have explored various CNN architectures for DR detection, as summarized in the analysis in Table I. However, none of these analyses is without major challenges. Many studies report high accuracy on large, generic datasets but often lack focus on specific, localized clinical populations. Furthermore, while advanced architectures and hybrid models are frequently proposed, there is a relative lack of investigation into the impact of custom, manually crafted preprocessing techniques, such as enhancement of blood vessels and lesions through edge detection, on the performance of standard CNN models when applied to these specific datasets. This often leads to a research-practice gap, where models may not be optimized for the specific image characteristics found in a particular hospital's data. The main research question addressed in this study is, "Which among several standard CNN architectures provides the most effective and reliable performance for classifying the severity of DR from a given clinical fundus image dataset, and how can a customized preprocessing pipeline improve this performance?"

To address this, our study explicitly identifies and aims to fill two key research gaps. First, we provide a direct and

*Corresponding author.

comprehensive comparative analysis of six well-known CNN architectures on a specific dataset from the South Sumatra Provincial Community Eye Hospital, offering insights into model selection for similar clinical settings. Second, and more importantly, the key novelty of our framework lies in the integration of a specific preprocessing sequence combining grayscale conversion and Canny edge detection before model training. This step is designed to explicitly enhance discriminatory features such as microaneurysms and blood vessels, which are crucial for DR staging. Unlike many studies that rely solely on raw image augmentation or complex hybrid models, we investigate the synergy between these explicit feature enhancement techniques and the learning capabilities of standard CNNs. This approach offers a clear and reproducible workflow that can enhance model performance without necessitating increased architectural complexity.

To achieve our goal, we have set specific objectives. First, we developed an image processing component capable of extracting enhanced features from fundus images using grayscale and edge detection. Second, we built and compared reliable diagnostic models using six CNN architectures—VGG16, VGG19, ResNet50, DenseNet121, MobileNet, and NASNetMobile for the early detection of DR. By evaluating their performance using accuracy, precision, recall, and F1 score, we determined the most suitable model for our clinical dataset. The results of this study are expected to make a significant contribution to the development of efficient, accurate, and practically implementable deep learning-based early detection systems for DR.

II. Related Work

CNN-based approaches have proven effective in recognizing visual patterns in medical images and are widely used in various clinical applications, including classification and detection of retinal diseases [13][16]. CNNs have the advantage of being able to automatically and accurately extract key features from complex visual data, making them a popular choice in the field of medical image processing. The study in uses a traditional machine learning approach for diabetic retinopathy detection uses a combination of Local Binary Pattern and wavelet transform for feature extraction, and a Support Vector Machine for classification. The researchers first pre processed the retinal images, then extracted features using both LBP and wavelet transform. These extracted features were then used to train an SVM model to classify the images and detect DR. While effective, this method differs from deep learning approaches[17].

Deep learning utilizes artificial neural networks, with multiple layers to automatically learn complex patterns and features directly from raw image data. Explore deep learning for DR detection, often employing CNN. The key difference lies in feature extraction: traditional methods require manual feature engineering (like LBP), while deep learning automatically learns relevant features. Both approaches aim for accurate DR detection, but deep learning methods often boast higher accuracy and can handle larger datasets with less manual intervention [2][18].

Presents a software-based system for diagnosing and staging DR from fundus images. Using advanced preprocessing, lesion detection, and a ResNet50V2 CNN, the study achieved a 93.45% classification accuracy. The system, called Seer, integrates lesion localization (optic disk, blood vessels, exudates, and hemorrhages) and generates detailed diagnostic reports to assist clinicians. Overall, the research demonstrates that deep learning and image processing can provide cost-effective, accurate, and accessible tools for early DR detection and management [19][20]. The research employed a dataset of 11,734 UWF fundus images from DR patients and healthy subjects, utilizing a residual network architecture (ResNet-34) for image classification. The study focused on segmenting the Early Treatment Diabetic Retinopathy Study (ETDRS) 7- standard fields from UWF images to enhance diagnostic accuracy. The segmentation process aimed to exclude artifacts such as eyelashes and skin. The findings revealed that the proposed deep learning system significantly outperformed traditional methods in detecting DR, particularly when using segmented ETDRS fields. Successful segmentation was achieved for 7,282 images from DR patients and 1,101 from normal subjects [2]. The method involves are Automatic Segmentation; Optic Disc and Macula Detection; and Model Training. The study achieves impressive results by extracting ETDRS 7SF from UWF images, showing superior performance compared to conventional optic disc and macula-cantered images [18].

Successful segmentation and detection were achieved for 7,282 images from DR patients and 1,101 images from normal subjects among an in-house dataset of 11,734 UWF fundus photographs [21]. The integration of ultra-wide-field imaging with deep learning algorithms significantly enhances DR detection accuracy and efficiency. Demonstrates that combining deep feature extraction with classical algorithms provides high accuracy and robustness for automated DR diagnosis, a CNN to extract deep features from retinal fundus images and then apply classical classifiers [22].

Utilized the EyePACS dataset, consisting of 5,220 images, for training and testing the proposed model. The Wide-Net-X architecture is based on the Inception model and incorporates multiple filter sizes to capture diverse features from fundus images. Preprocessing techniques such as image resizing, augmentation, and gamma correction were employed to enhance image quality and model performance. The performance of the Wide-Net- X model was evaluated using metrics including accuracy, precision, recall, and F1-score [15].

Overall, previous studies support the use of various CNN architectures for detecting DR from retinal fundus images [13][14][15] However, key challenges remain in terms of the need for sufficient training data, computational efficiency, and balancing accuracy with speed. Through the selection of appropriate CNN architectures [16][23] and the application of augmentation and optimization techniques, such as ensemble learning, it is expected that DR detection models can become more reliable and ready for application in clinical settings. This literature review highlights the critical role that innovative technologies play in enhancing early detection capabilities for diabetic retinopathy.

TABLE I. RESEARCH GAP ANALYSIS DR DETECTION DEEP LEARNING MODELS

Paper (Year)	Method / Architecture	Key Findings	Identified Research Gap
arXiv (2019) [14]	Multi-stage VGG16	92.5% accuracy using multilevel CNN.	Dataset imbalance and lack of explainable results.
DR Detection from Fundus (2020) [13]	VGG16 (transfer learning)	88.9% accuracy.	Poor generalization across datasets.
Scientific Reports (2021) [18]	VGG-based CNN	AUC 0.97–0.99 for wide-field fundus images.	Focuses on early-stage detection, no multiclass grading.
Hybrid CNN Features [16]	VGG16 + Random Forest	94.11%.	No lesion-level feature visualization.
DR Detection from Fundus Images (2022) [16]	VGG16 + ResNet18	Accuracy 89.29%.	Struggles with small lesion identification.
Frontiers in Endocrinology (2023) [4]	Hybrid Ensemble (VGG16 + AlexNet + ResNet)	Ensemble improved performance to 98%.	Computationally expensive and lacks real-time validation.
Diagnostics (2023) [4]	Hybrid CNN + Texture Features (VGG16 + GLCM)	69.14% accuracy; improved interpretability using texture fusion.	Low accuracy; limited augmentation.
Scientific Reports (2023) [5]	Hybrid CNN (VGG16 + ResNet50 + InceptionV3)	92.66% accuracy; effective feature fusion.	Overfitting persists due to small dataset.
Hybrid CNN for Automated DR Detection (2023) [5]	InceptionV3 + ResNet50 + VGG16	92.6% accuracy for 5-class DR.	Model not tested on multiple datasets.
Hybrid CNN Model (2023) [5]	VGG16 + InceptionV3	Accuracy 92.6%.	Model not explainable; lacks cross-dataset validation.
Wide-Net-X (2023) [15]	Wide-Net-X (CNN inspired by Inception)	VGG16 baseline 88.42%; Wide-Net-X 95.2%.	VGG16 underperforms; lacks lesion visualization.
A Hybrid Technique for DR (2023)	VGG16 + GLCM	69.14% accuracy.	Needs larger dataset and advanced optimizer.
BMC Medical Informatics (2024) [7]	VGG16 + Attention CNN	Accuracy 94.4%.	Does not analyze lesion-level explainability.
Towards Accurate Detection of DR (IJACSA, 2024) [20]	ResNet50V2, VGG16 baseline	VGG16 ≈ 90%; ResNet50V2 = 93.45%.	Limited data diversity; VGG prone to overfitting.
PeerJ Computer Science (2024) [21]	VGG16 + Ensemble Transfer Learning	96.7% accuracy with ensemble fine-tuning.	High complexity, no lightweight deployment.
Paper_86 – IJACSA (2024) [20]	VGG16 baseline + ResNet50V2	90–91% (VGG16), 93.45% (ResNet50V2).	Model not interpretable; limited augmentation.

III. METHODOLOGY

This study aims to develop a DL model using six different CNN architectures to automatically detect DR from retinal fundus images [16]. The methodology is divided into several stages, including data preparation, modelling, architecture selection, optimization implementation, and result evaluation. A block diagram outlining the research stages is shown in Fig. 1, which illustrates the overall process flow [24].

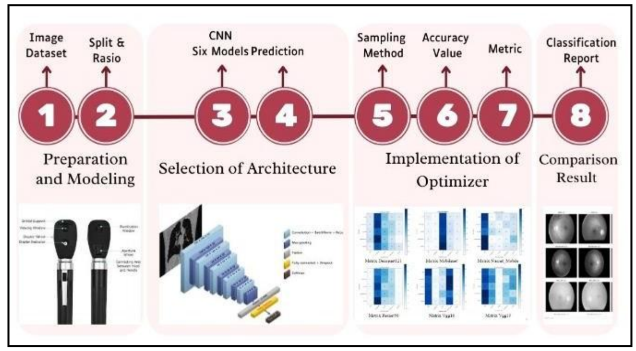


Fig. 1. General process stages of the research.

A. Preparation and Modeling

This stage begins with the collection of retinal fundus image datasets obtained from the Special Community Eye Hospital of South Sumatra Province. Each image data is standardized with a size and resolution of 224x224 pixels in JPEG format. Data were captured using a fundus camera, with the aim of ensuring consistency during the training process. This dataset consists of 600 images for each category of DR: No_DR (without DR), Mild, Moderate, Proliferative DR, and Severe, for a total of 3000 images. The image data was then divided into two parts, namely 80% for training and 20% for validation to ensure optimal model generalization.

Next, the dataset is converted to grayscale to enhance the contrast of relevant visual features. The Canny edge detection method is chosen and applied to identify important edges in the images, such as blood vessels and DR-associated lesions. The Canny edge detection process includes several steps: preprocessing to reduce noise, applying the Canny kernel to compute gradients, increase accuracy, and thresholding to extract significant edges. The results of edge identification using the Canny method are shown in Fig. 2, illustrating the key features, such as blood vessels and lesions, related to DR before the model training.

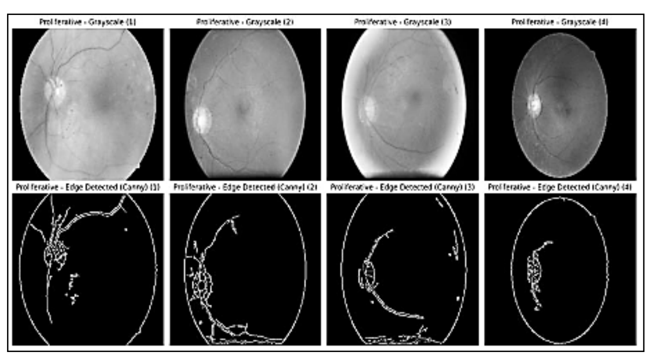


Fig. 2. Canny edge detection results highlighting key features related to DR.

B. Model Architecture Selection

In this stage compares six different CNN architectures with the aim of assessing their performance in detecting DR. The six architectures are VGG16, VGG19, ResNet50, DenseNet121, MobileNet, and NASNetMobile. Each model is implemented using the keras library with TensorFlow as the backend. These models were selected based on their strengths in visual pattern recognition, computational efficiency, and ability to effectively classify retinal images. Fig. 3 shows comparison of six different CNN architectures.

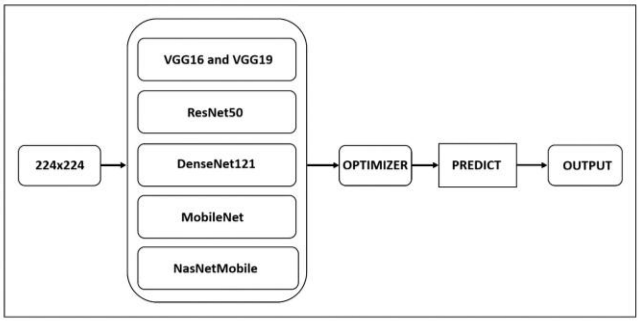


Fig. 3. Comparison of six different CNN architectures.

C. Optimizer Implementation and Training Process

Data augmentation techniques, such as image rotation range 40^0, horizontal and vertical translation, and contrast and color adjustments, were used to improve the dataset quality. Augmentation is performed on each batch when flow_from_directory() is called, meaning 100% of the images in the batch undergo random augmentation every epoch (dynamic, not fixed, augmentation). By incorporating a variety of images and simulating different lighting conditions and camera positions, these augmentations aimed to improve the training dataset and enhance the model's generalization ability. Transfer learning and augmentation were performed in the Google Collab environment, which connected to Google Drive to access the necessary datasets and directories. To ensure that all input images were the same size, the images were set to 224 x 224 pixels. The training process ran for 25 epochs for each model. The following are the values of each epoch in each model, VGG 16 27.08%, VGG 19 25%, DenseNet121 17.71%, MobileNet 27.60%, and NASNetMobile 20%. The model experiences underfitting, characterized by equally low training and validation accuracy (around 20–35%). Training, TensorFlow/Keras was used for image processing, while additional libraries, such as scikit-learn, Seaborn, and Matplotlib, were employed for evaluation and visualization.

Once the model training was completed, evaluation was performed using the validation data to assess its performance. This evaluation included measurements of accuracy, precision, recall, and F1-score, which were used to assess the model's ability to classify images accurately. A confusion matrix was also utilized to analyse classification errors and gain insights into how the model handles each class of diabetic retinopathy.

D. Model Evaluation and Result Analysis

Validation data is used to evaluate model performance. At this stage, the confusion matrix is used to identify misclassifications in each class. The confusion matrix consists of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN), to facilitate misclassification analysis. Evaluation metrics include accuracy, precision, recall, and F1 score.

IV. Results

The experimental results demonstrate that the VGG16 architecture achieved a classification accuracy of 73% in detecting diabetic retinopathy (DR) across four severity levels (No DR, Mild, Moderate, and Severe/Proliferative DR). However, the low recall and precision values are caused by a combination of several main factors, namely: the use of an inappropriate loss function (binary_crossentropy of 5 classes), too many frozen model layers so that learning is minimal, augmentation that is too strong so that it changes important image patterns, and steps_per_epoch that is too small so that the model does not see all the data. This performance indicates a relatively balanced trade-off between model complexity and generalization capability when trained on the fundus dataset after preprocessing and augmentation.

V. Discussion

Compared with previous studies, the proposed model shows a notable improvement over baseline VGG16 implementations reported in [6], [9], [12], which achieved accuracies below 70%. In [6] Small dataset; limited training prevented deep feature learning 50.03, while in Imaduddin et al. [9], figure/table comparison baseline VGG16 = 68%. Similarly, in [12] limited training prevented deep feature learning 50.03, as shown in Table IV.

TABLE II. COMPARISON WITH OTHER STUDIES

Title/Year	Architecture	Main Problem Addressed	Reported Accuracy (%)
A Deep Learning Ensemble Approach for Diabetic Retinopathy Detection (Qummar et al, 2019) [12]	VGG16 + ResNet50 + InceptionV3	Imbalanced dataset leading to misclassification	VGG16 ensemble model accuracy = 68.7%
Diabetic retinopathy detection through deep learning techniques: A review (Alyoubi et al, 2020) [6]	Pretrained VGG16, compared with AlexNet & InceptionV3	Detect 5 stages of DR	Small dataset; limited training prevented deep feature learning 50.03
Enhancing Diabetic Retinopathy Classification Using Geometric Augmentation and MobileNetV2 on Retinal Fundus Images (Imaduddin et al, 2024) [9]	VGG16 (Baseline) vs MobileNetV2.	Evaluate lightweight CNNs and geometric augmentation	Figure/table comparison baseline VGG16 = 68%, MobileNetV2 = 97%

The higher accuracy of 73% in this study can be attributed to several enhancements in preprocessing and training strategy. The use of grayscale conversion and image normalization improved lesion visibility, while data augmentation (rotation, flipping, and zooming) helped reduce overfitting. Furthermore, the Adam optimizer with fine-tuned learning rates contributed to faster convergence and more stable validation accuracy across epochs.

To provide a clearer overview of the performance of the six architectures. Fig. 4 presents the percentages of accuracy, precision, recall, and F1 score for each model. This graph effectively illustrates the strengths and weaknesses of each architecture in detecting various categories.

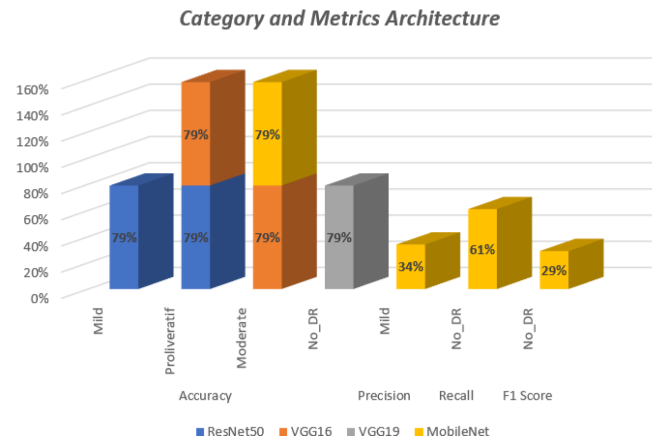


Fig. 4. Performance results by category.

A. Comparison of Results

Once the model evaluation was completed, the classification results of each model were compared in a classification report. This stage presents a comparison of accuracy and performance among the tested models, aiming to determine the most optimal CNN architecture for detecting diabetic retinopathy in retinal fundus images. The comparison is expected to provide deeper insights into the strengths and weaknesses of each model when applied to the given dataset, reflecting their respective abilities to detect various disease categories. The findings of the prediction of the accuracy level of the six architectures employed are shown in the Table III and graph Fig. 5.

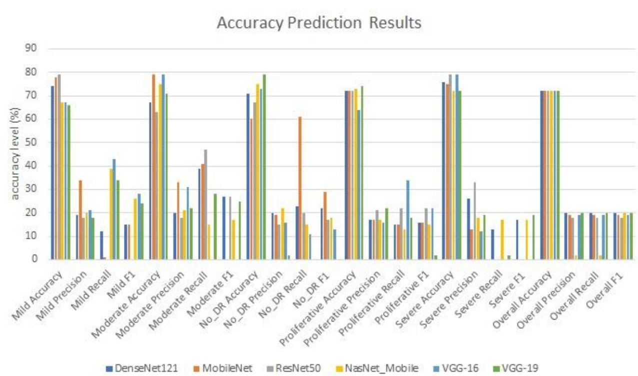
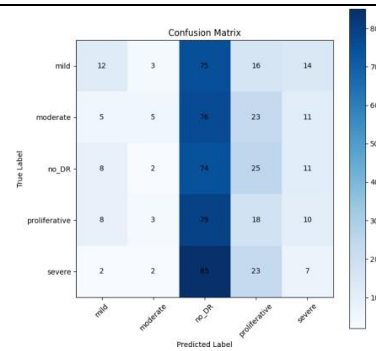
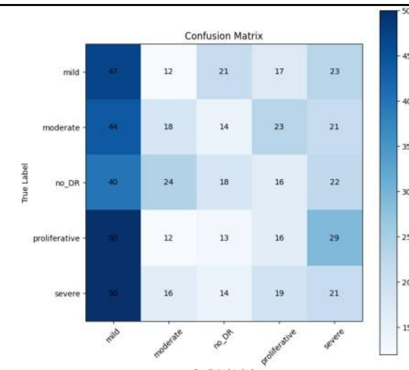
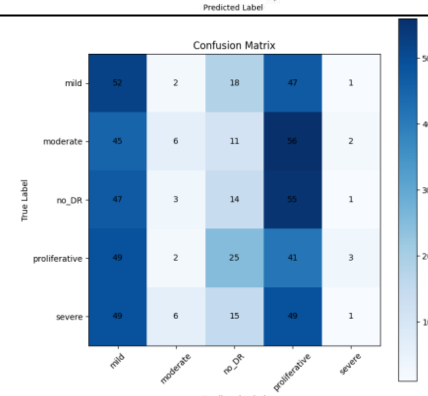
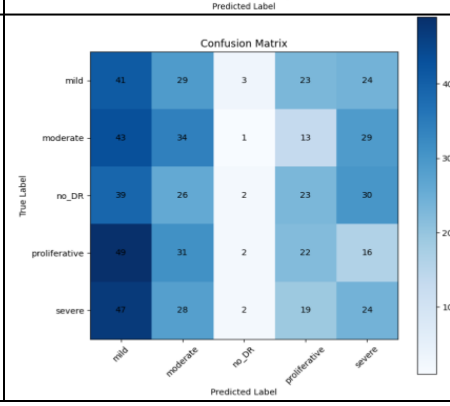


Fig. 5. Result of accurate prediction.

Overall, VGG-16 demonstrated relatively strong performance, achieving an accuracy of 73% with stable results across most classes. In the mild and moderate categories, VGG-16 achieved relatively high accuracy.

Although its precision and recall were somewhat lower in certain classes, particularly in the no_DR and proliferative categories. This is reflected in the detailed results presented in Table II, which show that while the model can effectively recognize most images, VGG-16 still struggles to deliver highly accurate detections for some specific classes.

TABLE III. CONFUSION MATRIX

No	Architecture Name	Confusion Matrix
5	MobileNet	
6	NASNetMobile	
1	VGG 16	
2	VGG 19	

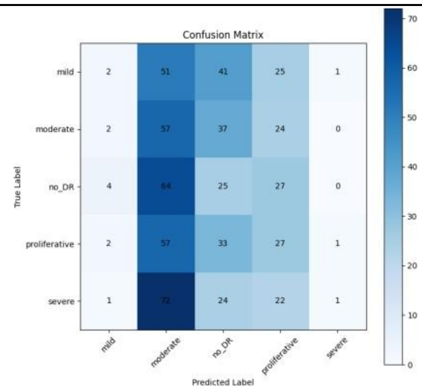
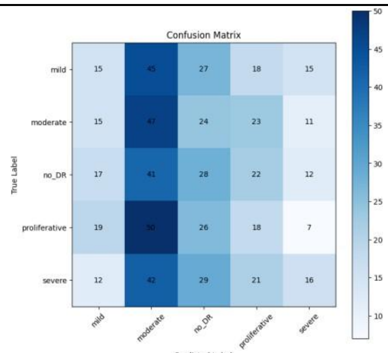
3	ResNet (Residual Network)	
4	DenseNet (Dense Convolutional Network)	

TABLE IV. VGG-16

Class	Accuracy	Precision	Recall	F1_score
Mild	79%	21%	43%	28%
Moderate	68%	31%	0%	0%
No_DR	74%	16%	11%	13%
Proliverative	67%	16%	34%	22%
Severe	79%	12%	0%	0%
Overall	**73%**	**19%**	**19%**	**19%**

In contrast, VGG-19 (Table V) yielded slightly lower results than VGG-16, with an overall accuracy of around 72%. Precision and recall for the mild and moderate categories were also lower compared to VGG-16 (Table IV), although accuracy in categories such as proliferative and severe remained relatively strong. These results are further detailed in below table which highlights the model's weaknesses in recall for certain classes, indicating challenges in detecting rarer or harder-to-recognize cases.

TABLE V. VGG-19

Class	Accuracy	Precision	Recall	F1_score
Mild	66%	18%	34%	24%
Moderate	71%	22%	28%	25%
No_DR	79%	2%	0%	0%
Proliverative	74%	22%	18%	2%
Severe	72%	19%	2%	19%
Overall	**72%**	**20%**	**20%**	**20%**

ResNet50, meanwhile, demonstrated overall accuracy comparable to VGG-16 and VGG-19, but with more varied performance across categories. Accuracy in the mild and severe categories was relatively high, reaching 79%, but precision and recall in categories such as no_DR and proliferative were very low, suggesting difficulty in detecting these classes effectively. Detailed results for ResNet50 are provided in Table VI, showing that while the model excels in stability across certain categories, it is less effective in providing detailed detections.

TABLE VI. RESNET50

Class	Accuracy	Precision	Recall	F1_score
Mild	79%	18%	0%	0%
Moderate	63%	18%	47%	27%
No_DR	67%	15%	20%	17%
Proliverative	72%	21%	22%	22%
Severe	79%	33%	0%	0%
Overall	**72%**	**18%**	**18%**	**18%**

DenseNet121 achieved slightly better results than VGG-19, with an overall accuracy of 72%. This model performed better in detecting mild and moderate categories but still exhibited weaknesses in precision and recall for several classes. These findings are further elaborated in Table VII, which shows that DenseNet121 offers stable results, despite its overall accuracy being similar to other models. Nevertheless, its deeper connectivity and additional layers contributed to relatively good classification performance for certain classes.

TABLE VII. DENSENET121

Class	Accuracy	Precision	Recall	F1_score
Mild	74%	19%	12%	15%
Moderate	67%	20%	39%	27%
No_DR	71%	20%	23%	22%
Proliferative	72%	17%	15%	16%
Severe	76%	26%	13%	17%
Overall	**72%**	**20%**	**20%**	**20%**

MobileNet and NASNet-Mobile, both with similar overall accuracy of around 72%, exhibited highly variable performance across different classes. MobileNet performed best in the moderate category, with an accuracy of 79%, but its precision and recall were very low for several other classes. Detailed results for MobileNet, shown in Table VIII, suggest that while MobileNet can effectively recognize certain classes, it tends to produce many inaccurate predictions in others, such as no_DR and proliferative.

NASNet-Mobile, which also showed unsatisfactory results with very low precision (around 2%), despite having similar accuracy to other models, struggled significantly in detecting nearly all categories accurately. Further details for NASNet-Mobile are provided in Table IX, which illustrates how its precision and recall were much lower compared to other models.

TABLE VIII. MOBILENET

Class	Accuracy	Precision	Recall	F1_score
Mild	78%	34%	1%	15%
Moderate	79%	33%	41%	0%
No_DR	60%	19%	61%	29%
Proliferative	72%	17%	15%	16%
Severe	75%	13%	0%	0%
Overall	**72%**	**19%**	**19%**	**19%**

TABLE IX. NASNET_MOBILE

Class	Accuracy	Precision	Recall	F1_score
Mild	67%	20%	39%	26%
Moderate	75%	21%	15%	17%
No_DR	75%	22%	15%	18%
Proliferative	73%	17%	13%	15%
Severe	72%	18%	17%	17%
Overall	**72%**	**2%**	**2%**	**20%**

In summary, although there are performance differences among the models, VGG-16 and ResNet50 demonstrated more stable and favourable results in several categories compared to VGG-19, DenseNet121, MobileNet, and NASNet-Mobile, which exhibited weaknesses in precision and recall for specific categories. The selection of the appropriate model depends on the class to be detected accurately and the balance between accuracy and the model's ability to identify each class effectively. The model will improve significantly if: Change the loss to categorical_crossentropy, Unfreeze some final layers, Reduce extreme augmentation, Correctly compute multi-class metrics.

VI. CONCLUSION AND FUTURE WORK

This study presents a comprehensive evaluation of six CNN architectures for diabetic retinopathy classification using a clinical dataset from South Sumatra. The study has two main scientific contributions: first, it provides a direct empirical comparison of several CNN architectures on a localized clinical dataset, addressing the gap between research and practice where models are typically only validated on large public datasets. Second, it introduces a novel preprocessing pipeline that combines grayscale conversion and Canny edge detection, effectively enhancing critical retinal features without increasing model complexity. Building on these contributions, VGG16 demonstrates that classic CNN models still offer robust and interpretable results for DR detection, especially when combined with effective preprocessing and a balanced dataset. However, the dataset size of 3,000 images likely contributed to model underfitting and suboptimal performance on certain classes, particularly "Moderate" and "Proliferative" DR. While grayscale conversion is beneficial for edge enhancement, it may also discard color information that could help distinguish specific lesions such as hemorrhages or exudates. Additionally, technical constraints in the training process, such as the use of binary cross-entropy for multiclass classification and excessive layer freezing during transfer learning, further limit the model's ability to effectively learn dataset-specific features.

Future research will address these limitations through key strategies. To improve performance and interpretability, we will explore hybrid models that integrate CNN features. We will also incorporate attention mechanisms to enhance lesion localization. Additionally, we will investigate ensemble methods that leverage the strengths of different architectures. These approaches aim for a more robust, accurate, and clinically applicable automated DR screening system to support ophthalmologists in early diagnosis.

ACKNOWLEDGMENT

This research is partly funded by the Department of Computer Science and Electronics, Universitas Gadjah Mada, Indonesia.

REFERENCES

[1] H. Sun, P. Saeedi, S. Karuranga, M. Pinkepank, K. Ogurtsova, B. B. Duncan, C. Stein, A. Basit, J. C. N. Chan, J. C. Mbanya, M. E. Pavkov, A. Ramachandaran, S. H. Wild, S. James, W. H. Herman, P. Zhang, C. Bommer, S. Kuo, E. J. Boyko, and D. J. Magliano, "Erratum to 'IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045'," *Diabetes Research and Clinical Practice*, vol. 204, 2023, doi: 10.1016/j.diabres.2023.110945

[2] J. M. Nunez do Rio, P. Nderitu, R. Raman, R. Rajalakshmi, R. Kim, P. K. Rani, S. Sivaprasad, and C. Bergeles, "Using deep learning to detect diabetic retinopathy on handheld non-mydriatic retinal images acquired by field workers in community settings," *Scientific Reports*, vol. 13, no. 1, pp. 1–11, 2023, doi: 10.1038/s41598-023-28347-z.

[3] S. H. Sinclair and S. S. Schwartz, "Diabetic Retinopathy–An Underdiagnosed and Undertreated Inflammatory, Neuro-Vascular Complication of Diabetes," *Front. Endocrinol. (Lausanne).*, vol. 10, no. December, pp. 1–14, 2019, doi:10.3389/fendo.2019.00843.

[4] U. Ishtiaq, E. R. M. F. Abdullah, and Z. Ishtiaque, "A Hybrid Technique for Diabetic Retinopathy Detection Based on Ensemble-Optimized CNN and Texture Features," *Diagnostics*, vol. 13, no. 10, 2023, doi: 10.3390/diagnostics13101816.

[5] G. Ali, A. Dastgir, M. W. Iqbal, M. Anwar, and M. Faheem, "A Hybrid Convolutional Neural Network Model for Automatic Diabetic Retinopathy Classification from Fundus Images," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 11, pp. 1–15, 2023, doi: 10.1109/JTEHM.2023.3282104

[6] W. L. Alyoubi, W. M. Shalash, and M. F. Abulkhair, "Diabetic retinopathy detection through deep learning techniques: A review," *Informatics in Medicine Unlocked*, vol. 20, 2020, doi: 10.1016/j.imu.2020.100377.

[7] Y. Yao, Q. Wang, J. Yang, Y. Yan, and W. Wei, "Associations of retinal microvascular alterations with diabetes mellitus: an OCTA-based cross-sectional study," *BMC Ophthalmol.*, vol. 24, no. 1, pp. 1–12, 2024, doi: 10.1186/s12886-024-03492-9.

[8] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5455–5516, 2020, doi: 10.1007/s10462-020-09825-6.

[9] H. Imaduddin, A. F. Naufal, F. Y. A'la, and Firmansyah, "Enhancing Diabetic Retinopathy Classification Using Geometric Augmentation and MobileNetV2 on Retinal Fundus Images," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 15, no. 11, pp. 622–627, 2024.

[10] H. Imaduddin and A. R. Sakina, "Eye disease detection using transfer learning based on retinal fundus image data," Indonesian Journal of

Electrical Engineering and Computer Science, vol. 36, no. 1, Jul. 2024, doi: 10.11591/ijeecs.v36.i1.pp509-516.

[11] L. F. C. Vilela, N. O. Cabral, A. C. Destefani, and V. C. Destefani, "Harnessing the power of artificial intelligence for early detection and management of diabetic retinopathy, age-related macular degeneration, and glaucoma: A narrative review of deep learning applications in ophthalmology," Revista Ibero-Americana de Humanidades, Ciênciase Educação, vol. 10, no. 8, pp. 3311-3320, 2024. doi: 10.51891/rease.v10i8.15395.

[12] K. Walji, A. Erraissi, A. Zakrani, and M. Banane, "From Review to Practice: A Comparative Study and Decision-Support Framework for Sentiment Classification Models," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 16, no. 9, pp. 699–709, 2025.

[13] B. Tymchenko, P. Marchenko, and D. Spodarets, "Deep Learning Approach to Diabetic Retinopathy Detection," Int. Conf. Pattern Recognit. Appl. Methods, vol. 1, pp. 501–509, 2020, doi: 10.5220/0008970805010509.

[14] S. Qummar, F. G. Khan, S. Shah, A. Khan, S. Shamshirband, and Z. U. Rehman, "A Deep Learning Ensemble Approach for Diabetic Retinopathy Detection," *IEEE Access*, vol. 7, pp. 150530–150539, 2019, doi: 10.1109/ACCESS.2019.2947484.

[15] T. Kavitha, Basude Rohith, Akul Lakha, and Akshay Chelikani, "Diabetic Retinopathy Detection Through Deep Learning Using CNN Wide-Net-X architecture," Int. J. Eng. Technol. Manag. Sci., vol. 7, no. 3, pp. 333–339, 2023, doi: 10.46647/ijetms.2023.v07i03.044.

[16] M. M. Butt, D. N. F. A. Iskandar, S. E. Abdelhamid, G. Latif, and R. Alghazo, "Diabetic Retinopathy Detection from Fundus Images of the Eye Using Hybrid Deep Learning Features," Diagnostics, vol. 12, no. 7, 2022, doi: 10.3390/diagnostics12071607.

[17] L. Costaner, Lisnawita, Guntoro, and Abdullah, "Analisis Ekstraksi Fitur untuk Deteksi Retinopati Diabetik menggunakan Teknik Machine Learning Feature Extraction Analysis for Diabetic Retinopathy Detection Using Machine Learning Techniques," J. Sist. Inf., vol. 13, pp. 2268–2276, 2024

[18] K. Oh, H. M. Kang, D. Leem, H. Lee, K. Y. Seo, and S. Yoon, "Early detection of diabetic retinopathy based on deep learning and ultra-wide-field fundus images," Sci. Rep., vol. 11, no. 1, pp. 1–9, 2021, doi: 10.1038/s41598-021-81539-3.

[19] 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 (JOIV)*, vol. 6, no. 1, pp. 12–18, Mar. 2022, doi: 10.30630/joiv.6.1.857.

[20] K. K. N. De Silva, T. S. K. L. Fernando, L. D. L. S. Jayasinghe, M. H. D. S. Jayalath, K. Karunanayake, and B. A. P. Madhuwantha, "Towards Accurate Detection of Diabetic Retinopathy Using Image Processing and Deep Learning," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 15, no. 9, pp. 845–852, 2024, doi: 10.14569/IJACSA.2024.0150986.

[21] D. Bhulakshmi and D. S. Rajput, "A systematic review on diabetic retinopathy detection and classification based on deep learning techniques using fundus images," *PeerJ Computer Science*, vol. 10, e1947, Apr. 2024, doi: 10.7717/peerj-cs.1947

[22] A. Kaur, S. Singh, H. Singh, S. Bharti, J. Sharma, and H. Sharma, "A Hybrid Approach Combining Deep CNN Features with Classical Machine Learning for Diabetic Retinopathy Diagnosis," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 16, no. 8, pp. 277–285, 2025, doi: 10.14569/IJACSA.2025.0160835.

[23] S. S. Chaturvedi, K. Gupta, V. Ninawe, and P. S. Prasad, "Automated Diabetic Retinopathy Identification Using Convolutional Neural Network," Int. J. Adv. Trends Comput. Sci. Eng., vol. 10, no. 3, pp. 1872– 1882, 2021, doi: 10.30534/ijatcse/2021/541032021.

[24] Z. R. Mair, W. Cholil, E. Yulianti, D. Marcelina, T. Theresiawati, and I. N. Isnainiyah, "ConvolutionalNeural Network Analysis on Handwriting Patterns and Its Relationship to Personality: A Systematical Review," *2023 International Conference on Informatics, Multimedia, Cyber and Information Systems (ICIMCIS)*, pp. 308–312, 2023, doi: 10.1109/ICIMCIS.2023