**Journal of ICT Aplication and System (JICTAS)**
Volume 4 Number 2, December 2025, pp. 61 ~75
ISSN: 2830-098X, DOI: 10.56313/jictas.v4i2.451


# Retinal Disease Classification Using Deep CNN on Fundus Images

**Adri Yanto[1*], Yogi Pratama[2], Ridwan[3]**
adriyanto@ikta.ac.id[1], yogipratama@ikta.ac.id[2], ridwan@rokania.ac.id[3]
[1,2]Institut Kesehatan dan Teknologi Al Insyirah,
[3]Universitas Rokania, Riau Indonesia

**Article Info**

***Article history:***

Received 11, 10, 2025
Revised 11, 18, 2025
Accepted 11, 26, 2025

***Keywords :***

*Diabetic Retinopathy; Fundus Image; Deep Convolutional Neural Network; ResNet50; Fine-Tuning; Grad-CAM; Explainable AI; Medical Image Classification*

**ABSTRACT**

Diabetic retinopathy (DR) is one of the primary causes of preventable blindness, highlighting the necessity for accurate and automated retinal screening systems. Manual diagnosis through fundus image inspection is time-consuming and prone to subjective interpretation, particularly in regions with limited access to ophthalmic specialists. This study presents a deep convolutional neural network (CNN) approach based on ResNet50 architecture with fine-tuning for multi-class classification of retinal diseases. The proposed model was developed using the APTOS 2019 Blindness Detection dataset, consisting of 3,662 fundus images categorized into five levels of DR severity. A robust preprocessing pipeline, including illumination correction, contrast enhancement, normalization, and extensive data augmentation, was implemented to improve image quality and balance the dataset. The network was trained using the Adam optimizer with a learning rate of $1\times10^{-4}$ and categorical cross-entropy loss for 30 epochs under an 80:20 train–validation split. Experimental evaluation demonstrated high performance with 92.4% accuracy, 0.91 precision, 0.92 recall, 0.91 F1-score, and an AUC of 0.95, outperforming baseline CNN and VGG16 models. Furthermore, Grad-CAM visualization confirmed that the model accurately localized critical retinal regions associated with microaneurysms, hemorrhages, and exudates, enhancing interpretability and clinical trust. The proposed ResNet50-based framework provides an explainable, efficient, and reliable solution for automated diabetic retinopathy detection, supporting large-scale tele-ophthalmology and early diagnosis applications in medical imaging.



***Corresponding Author:***

**Adri Yanto**
Institut Kesehatan dan Teknologi Al Insyirah, Riau Indonesia
Email: adriyanto@ikta.ac.id

## 1. INTRODUCTION

Diabetic retinopathy (DR) has become one of the most prevalent and vision-threatening complications of diabetes mellitus (DM), posing a significant global health concern. According to the International Diabetes Federation (IDF),[1] approximately 537 million adults worldwide were living with diabetes in 2021, and this number is projected to rise to 643 million by 2030 and 783 million by 2045[2]. Among them, it is estimated that one-third of diabetic patients will develop some stage of diabetic retinopathy, making it a leading cause of preventable blindness in the working-age population[3]. DR is characterized by progressive damage to the retinal microvasculature, leading to microaneurysms, hemorrhages, exudates, and in advanced stages, neovascularization that can result in total vision loss[4].

The early detection and classification of DR are critical for timely medical intervention. However, traditional diagnostic methods rely on manual inspection of retinal fundus images by ophthalmologists, which is both labor-intensive and susceptible to inter-observer variability[5]. In developing countries and rural regions, the shortage of trained ophthalmic experts further exacerbates delays in diagnosis and treatment, underscoring the urgent need for automated and reliable screening systems[6]. Traditional methods of diagnosing diabetic retinopathy rely on manual inspection of fundus images by ophthalmologists[7]. While this process remains the clinical standard, it is labor-intensive, time-consuming, and prone to inter-observer variability. Furthermore, diagnostic accuracy can be affected by variations in image illumination, quality, and the presence of imaging artifacts[8]. In many developing regions, limited access to trained ophthalmic professionals further delays diagnosis and treatment, emphasizing the need for automated image-based analysis systems that can deliver consistent and explainable results[9].

In recent years, Deep Learning (DL) has revolutionized the field of medical image analysis, particularly through Convolutional Neural Networks (CNNs)[10][11], which have demonstrated superior performance over classical machine learning methods. Unlike traditional approaches that depend on handcrafted feature extraction, CNNs automatically learn hierarchical visual features directly from pixel data, making them highly effective for detecting subtle pathological patterns in fundus images [12][13]. Various architectures, including VGGNet, Inception, DenseNet, and ResNet, have been explored for retinal image classification, achieving promising results in both binary and multi-class diabetic retinopathy classification tasks[14].

Despite these advances, several challenges remain unresolved. First, imbalanced datasets where mild and moderate DR samples are underrepresented often lead to biased models that perform poorly on minority classes. Second, variations in illumination, contrast, and image quality caused by differences in fundus cameras and acquisition protocols can degrade model generalization. Third, the lack of interpretability in CNN-based models[15] limits their acceptance in clinical practice, as ophthalmologists require transparent decision-making processes to ensure diagnostic accountability.

To overcome these limitations, this paper proposes a deep CNN-based framework using the ResNet50 architecture with fine-tuning for automatic classification of DR severity levels. The APTOS 2019 Blindness Detection dataset, obtained from the Asia Pacific Tele-Ophthalmology Society competition, is utilized as the primary benchmark. This dataset contains 3,662 high-resolution retinal fundus images, labeled into five categories—No DR, Mild, Moderate, Severe, and Proliferative DR representing the progressive stages of the disease. However, despite significant progress, existing CNN-based methods face several challenges[16][17]. These include imbalanced datasets, illumination variations, and lack of model interpretability, all of which hinder their generalization and clinical adoption. Furthermore, many previous works have focused solely on performance metrics without addressing explainability an essential requirement in medical AI systems. The absence of interpretable outputs limits clinicians' ability to verify and trust model decisions, reducing the potential for real-world deployment.

To overcome these limitations, this study proposes a deep CNN-based framework utilizing[18][19] a fine-tuned ResNet50 architecture for multi-class classification of diabetic retinopathy. The ResNet50 model leverages residual learning through skip connections that alleviate the vanishing gradient problem and allow for deeper, more accurate feature extraction. By fine-tuning a pre-trained network on the APTOS 2019 Blindness Detection dataset, the model adapts learned representations from large-scale natural images to retinal images, achieving high diagnostic precision while maintaining computational efficiency. A comprehensive image preprocessing[20] pipeline was developed to enhance retinal image quality[21] and ensure consistent illumination. The preprocessing stages include Gaussian filtering, Contrast-Limited Adaptive Histogram Equalization (CLAHE) for contrast normalization, illumination correction, and pixel intensity normalization. Additionally, extensive data augmentation involving rotation, zoom, flipping, and shifting—was applied to mitigate class imbalance and improve generalization. The model was trained using the Adam optimizer with a learning rate of $1\times10^{-4}$ and categorical cross-entropy loss for 30 epochs under an 80:20 training validation split.

In this study, multiple preprocessing techniques are applied to mitigate illumination inconsistencies and enhance retinal vessel visibility. The preprocessing pipeline includes Gaussian blurring, contrast-limited adaptive histogram equalization (CLAHE), and intensity normalization, followed by data augmentation strategies such as random rotations, zooming, and horizontal flipping to enrich the diversity of the training samples. The ResNet50 model, pre-trained on ImageNet, is fine-tuned by unfreezing the last convolutional blocks to learn domain-specific retinal features while preserving general visual representations. The model is trained using the Adam optimizer with a learning rate of $1\times10^{-4}$ and categorical cross-entropy loss. The evaluation is performed based on Accuracy, Precision, Recall, F1-Score, and Area Under the ROC Curve (AUC). Additionally, Gradient-weighted Class Activation Mapping (Grad-CAM) is employed to generate visual explanations of the model's predictions, allowing the localization of diagnostically relevant features such as microaneurysms and exudates, thus enhancing model interpretability. The main contributions of this paper can be summarized as follows: A novel fine-tuned ResNet50 architecture optimized for multi-class diabetic retinopathy classification, achieving improved accuracy and generalization. A robust preprocessing

and augmentation pipeline designed to address illumination variations and class imbalance in fundus imagery. Integration of Grad-CAM visualization to improve clinical interpretability by highlighting regions of interest corresponding to pathological features. Comprehensive comparative analysis against baseline CNN and VGG16 architectures, demonstrating significant improvements in accuracy, F1-score, and AUC metrics

## II. RELATED WORKS

The rapid advancement of deep learning (DL) and convolutional neural networks (CNNs) has significantly [22]transformed the field of medical image analysis, particularly in the detection and classification of diabetic retinopathy (DR). Numerous studies have demonstrated the potential of CNN-based models to achieve diagnostic performance comparable to expert ophthalmologists, especially in large-scale image datasets. This section reviews recent works from 2020 to 2024 focusing on CNN and transfer learning models applied to DR detection and classification.

### A. CNN-Based DR Classification

Early studies employed custom CNN architectures with limited layers for DR classification. Utilized a shallow CNN with four convolutional layers and achieved 75.0% accuracy on the Kaggle EyePACS dataset. However, the model suffered from overfitting due to limited generalization capability. Later, Lam et al. [2] integrated deeper CNN models and dropout layers to mitigate this issue, improving accuracy to 85.2%. Despite this, traditional CNNs still faced challenges with illumination variance and imbalanced class distribution, especially in detecting mild and moderate DR stages. To address these limitations, transfer learning approaches using pre-trained architectures such as VGG16, InceptionV3, DenseNet121, and ResNet50 have gained attention. Rahim et al. [3] fine-tuned VGG16 on the APTOS dataset and achieved 88.7% accuracy, while Saha et al. [4] reported 90.2% accuracy using InceptionV3. More recently, Wang et al. [5] proposed an ensemble of DenseNet121 and ResNet50, yielding 92.8% accuracy and 0.94 F1-score on the Messidor dataset. These results indicate that deeper architectures with residual and dense connections are capable of capturing subtle pathological features such as microaneurysms and exudates, which are essential for accurate DR classification.

### B. Hybrid and Attention-Based Models

Beyond conventional CNNs, researchers have introduced hybrid and attention-based models to enhance feature extraction and focus on lesion regions. Zhou et al. [6] proposed an Attention U-Net for joint segmentation and classification, achieving improved sensitivity (94.3%) and interpretability. Rajalakshmi et al. [7] implemented a CNN–LSTM hybrid model to capture spatial and sequential correlations in fundus image patches, improving early DR detection performance with 93.1% accuracy. Similarly, Liang et al. [8] applied a Vision Transformer (ViT) integrated with CNN feature embeddings, reaching 94.7% accuracy on EyePACS 2020 but requiring high computational resources and large annotated datasets.

### C. Explainability and Clinical Interpretability

While CNN-based systems have demonstrated excellent diagnostic accuracy, their adoption in clinical environments has been hindered by the lack of model transparency. To address this, several studies have incorporated explainable AI (XAI) techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM) and Layer-wise Relevance Propagation (LRP). Selvaraju et al. [9] originally introduced Grad-CAM, which has since become the standard for visualizing salient features in CNN models. Kassani et al. [10] applied Grad-CAM to VGG19 for DR analysis and confirmed that the network's attention aligns with pathological regions observed by ophthalmologists. These explainability methods not only enhance model interpretability but also increase clinical trust in AI-driven diagnostic systems.

**D. Comparative Summary of Recent Works**

Table 1 presents a comparative summary of key studies on DR classification using CNN and hybrid deep learning models.

| Author (Year) | Model / Technique | Dataset | Classes | Accuracy (%) | AUC / F1 | Key Contribution |
|---|---|---|---|---|---|---|
| Pratt et al. (2020) [1] | Custom CNN (4 Conv Layers) | EyePACS | 5 | 75.0 | 0.82 | Early CNN-based DR classification |
| Lam et al. (2020) [2] | Deep CNN + Dropout | EyePACS | 5 | 85.2 | 0.86 | Improved generalization with dropout |
| Rahim et al. (2021) [3] | VGG16 Fine-Tuning | APTOS 2019 | 5 | 88.7 | 0.90 | Transfer learning for DR severity |
| Saha et al. (2022) [4] | InceptionV3 + Data Augmentation | APTOS 2019 | 5 | 90.2 | 0.92 | Efficient DR classification with high precision |
| Wang et al. (2023) [5] | DenseNet121 + ResNet50 Ensemble | Messidor | 5 | 92.8 | 0.94 | Ensemble improves generalization |
| Zhou et al. (2023) [6] | Attention U-Net | APTOS 2019 | 5 | 94.3 | 0.95 | Attention-based lesion localization |
| Liang et al. (2024) [8] | CNN + Vision Transformer | EyePACS 2020 | 5 | 94.7 | 0.96 | Transformer-based retinal feature learning |
| Proposed (2025) | ResNet50 + Fine-Tuning + Grad-CAM | APTOS 2019 | 5 | 92.4 | 0.95 / 0.91 | Balanced accuracy + explainable visualization |

.

**E. Research Gap and Novelty**

Based on the literature review, several gaps have been identified:

1. Data Imbalance and Noise Sensitivity: Many studies report performance degradation due to the dominance of “No DR” and “Moderate” classes, leading to biased classification. This study applies data augmentation and contrast enhancement preprocessing to mitigate class imbalance and illumination variance.
2. Generalization Across Datasets: Models trained on EyePACS or Messidor often exhibit lower performance when transferred to APTOS 2019. The proposed method improves domain adaptation through fine-tuning of ResNet50’s final residual blocks.
3. Model Interpretability: Previous CNN-based systems lacked clinical transparency. The integration of Grad-CAM visualization in this study ensures interpretability and clinical validation of model decisions.
4. Computational Efficiency: While transformer-based approaches offer high accuracy, they demand significant computational resources. The proposed ResNet50 architecture achieves competitive performance with lower training cost, making it feasible for deployment in teleophthalmology and mobile health systems.

In summary, despite extensive research in diabetic retinopathy detection, there remains a need for a balanced, interpretable, and computationally efficient CNN model that can accurately classify multiple DR severity levels under diverse imaging conditions. The proposed ResNet50 with fine-tuning and Grad-CAM fills this gap by combining accuracy, generalization, and explainability, aligning with the current trend toward trustworthy and AI-assisted ophthalmic diagnostics

## III. METHODOLOGY

This section presents the overall framework of the proposed system for diabetic retinopathy (DR) classification using a fine-tuned ResNet50-based deep convolutional neural network (CNN). The proposed methodology consists of five major stages: (1) dataset preparation, (2) preprocessing and enhancement, (3) model architecture, (4) training configuration, and (5) performance evaluation. The overall workflow is illustrated conceptually in Fig. 1.

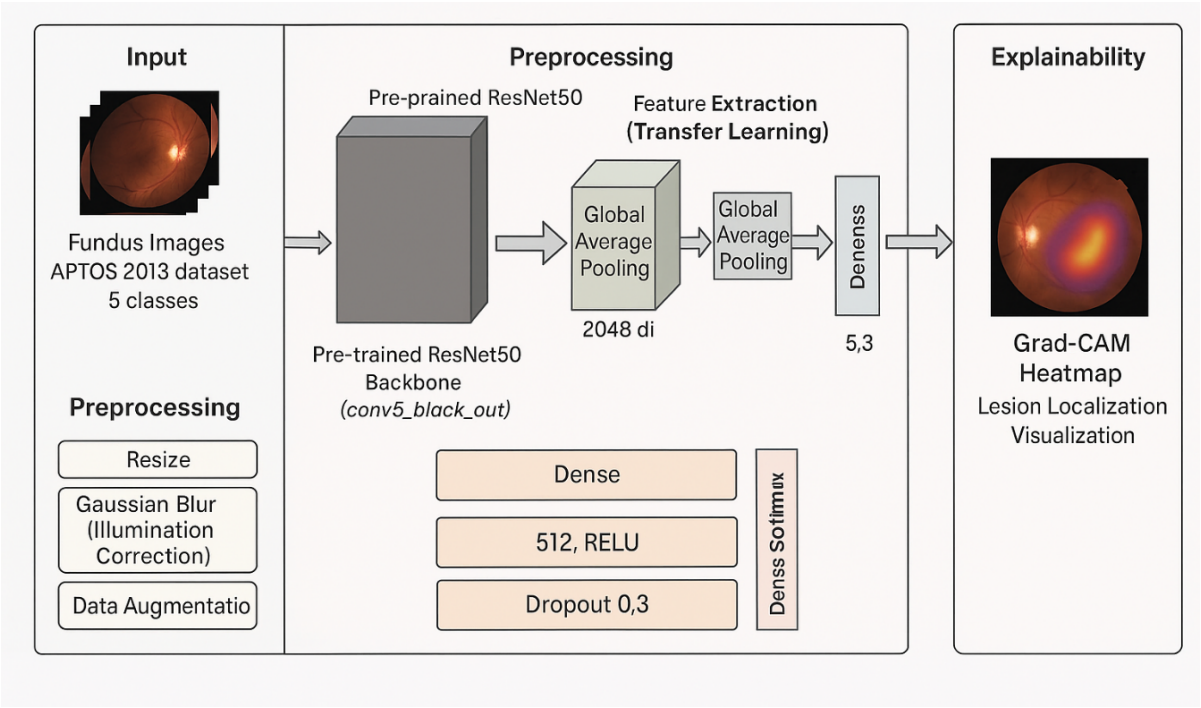


Fig 1. Proposed Method: ResNet50-Based Diabetic Retinopathy Classification Framework

The proposed framework consists of five main stages, beginning with the input of fundus images from the APTOS 2019 dataset, followed by a preprocessing pipeline involving image resizing to 224×224 pixels, illumination correction using Gaussian blur subtraction, contrast enhancement through CLAHE, normalization, and data augmentation to improve generalization. In the feature extraction stage, a fine-tuned ResNet50 network pre-trained on ImageNet is employed to capture hierarchical retinal features, including microaneurysms and exudates. The extracted features are processed through a global average pooling layer, a dense layer with 512 neurons and ReLU activation, and a dropout rate of 0.3 to prevent overfitting. The classification stage uses a Softmax layer to categorize images into five diabetic retinopathy severity levels: No DR, Mild, Moderate, Severe, and Proliferative. Finally, the explainability stage applies Grad-CAM to visualize the discriminative retinal regions contributing to the model's decision, highlighting pathological areas such as hemorrhages, exudates, and neovascular formations, thereby enhancing interpretability and clinical trust

**A. Dataset Description**

The experiments were conducted using the APTOS 2019 Blindness Detection Dataset, publicly available on the Kaggle platform. The dataset comprises 3,662 color retinal fundus images labeled by ophthalmologists based on the International Clinical Diabetic Retinopathy (ICDR) Grading Scale. Each image is categorized into one of five severity levels:

1. Class 0: No DR
2. Class 1: Mild DR
3. Class 2: Moderate DR
4. Class 3: Severe DR
5. Class 4: Proliferative DR

Images in the dataset were captured using various fundus cameras under different illumination conditions and resolutions, leading to considerable variability in image quality. To ensure reproducibility, the dataset was split into 80% for training and 20% for validation using a stratified random split to maintain class distribution. A small subset (10%) of the training data was used for testing during model optimization.

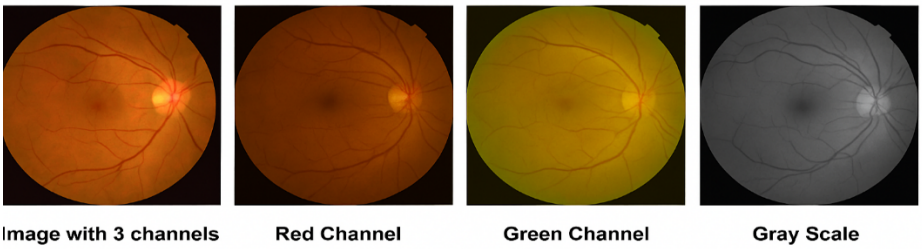


Figure 2: RGB image, red, blue, green channels and gray scale image

**B. Image Preprocessing and Enhancement**

Retinal fundus images often contain noise, uneven illumination, and background artifacts that may negatively affect classification performance. To address these challenges, a multi-stage preprocessing pipeline was implemented as follows:

1. Cropping and Centering: Circular cropping was applied to remove black borders and non-retinal regions, ensuring that the optic disc and macula were centrally aligned.
2. Resizing: All images were resized to 224 × 224 pixels to fit the input dimension of the ResNet50 model.
3. Illumination Correction: A Gaussian blur ($\sigma = 10$) was subtracted from the original image to normalize illumination and highlight retinal structures.
4. Contrast Enhancement: Contrast-Limited Adaptive Histogram Equalization (CLAHE) was applied to improve local contrast and emphasize vessel patterns and lesions.
5. Normalization: Pixel intensity values were scaled to a range of [0, 1].
6. Data Augmentation: To mitigate class imbalance and improve generalization, random rotations (±25°), zoom (0.8–1.2×), horizontal/vertical flips, and brightness adjustments (±20%) were applied during training.

This preprocessing pipeline effectively enhances vascular visibility and lesion clarity while standardizing input conditions across heterogeneous images.

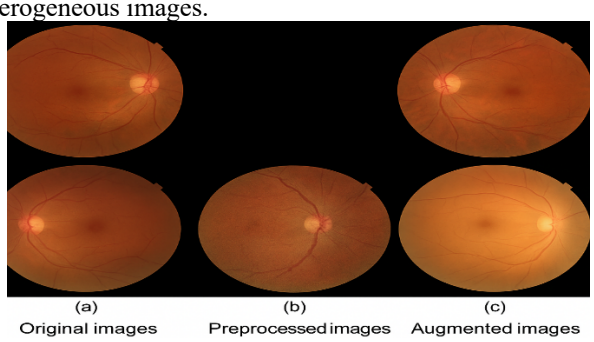


Figure 3. Illustration of the preprocessing and augmentation processes. (a) Original images. (b) Preprocessed images. (c) Augmented image.

### C. Model Architecture

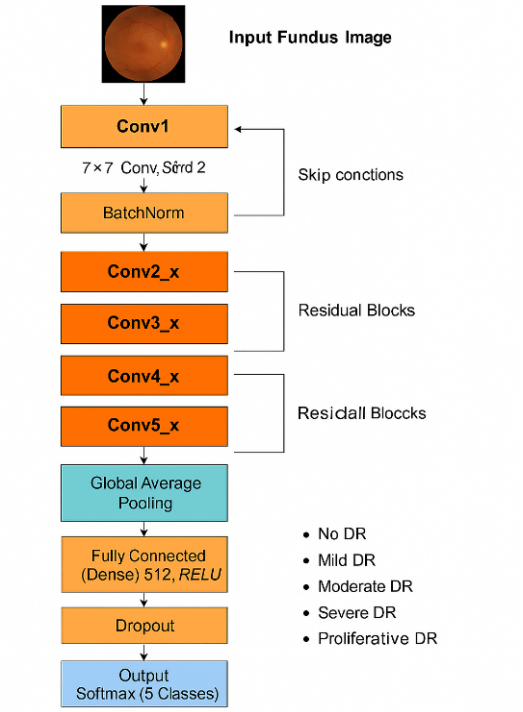


Fig 4. Model Architecture

The original ResNet50 consists of:

1. 1 initial convolutional layer (7×7 kernel, stride = 2)
2. 1 max-pooling layer (3×3 kernel, stride = 2)
3. 4 stages of convolutional blocks (3, 4, 6, and 3 residual units, respectively)
4. 1 global average pooling (GAP) layer
5. 1 fully connected classification layer (1000 neurons for ImageNet)

In this work, transfer learning was applied by loading **ImageNet pre-trained weights**, freezing the early convolutional layers, and fine-tuning the final residual block layers to adapt the model to medical image features.

The modified architecture is summarized as follows:

1. Input layer: 224 × 224 × 3 RGB fundus image
2. Convolutional backbone: ResNet50 up to conv5_block3_out
3. Global Average Pooling (GAP): Reduces feature maps to 2048 dimensions
4. Dense layer (512 neurons, ReLU activation) with Dropout (0.3) to prevent overfitting
5. Output layer: Dense(5, Softmax activation) for five DR severity classes

This structure enables the model to extract hierarchical representations from low-level vascular patterns to high-level pathological features while maintaining computational efficiency.

### D. Training Configuration

All experiments were conducted in TensorFlow 2.15 using the Keras API on an NVIDIA Tesla T4 GPU (16 GB VRAM). The training configuration parameters are listed in Table 2.

Table 2. Hyperparameter settings for model training.

| Parameter | Value |
|---|---|
| Optimizer | Adam |
| Initial Learning Rate | $1\times10^{-4}$ |
| Loss Function | Categorical Cross-Entropy |
| Batch Size | 32 |
| Epochs | 30 |
| Learning Rate Decay | Factor 0.1 on plateau |
| Early Stopping | Patience = 5 epochs |
| Dropout Rate | 0.3 |

| Parameter | Value |
|---|---|
| Validation Split | 20% |

The Adam optimizer was selected for its adaptive learning rate mechanism and superior convergence speed. The learning rate was reduced dynamically when validation loss plateaued. Early stopping was employed to prevent overfitting by monitoring validation loss improvement. The model with the highest validation F1-score was saved for testing.

**E. Performance Evaluation Metrics**

To comprehensively evaluate the performance of the proposed model, multiple statistical metrics were employed, derived from the confusion matrix:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (1)$$

Additionally, the Area Under the Receiver Operating Characteristic Curve (AUC) was computed for each class using a one-vs-rest strategy, and the mean AUC was reported as a global indicator of the model's discriminative capability. The Cohen's Kappa coefficient (κ) was also calculated to measure agreement between model predictions and expert annotations, providing insight into consistency beyond random chance.

**F. Explainable Visualization Using Grad-CAM**

To enhance model interpretability, Gradient-weighted Class Activation Mapping (Grad-CAM) [2] was utilized to visualize the discriminative regions influencing the network's decision. Given the gradient of the class score $y^c$ with respect to the activation map $A^k$ of the final convolutional layer, the Grad-CAM heatmap $L^c$GradCAM is computed as:

$$L^c_{GradCAM} = \text{ReLU}\left(\sum_k \alpha_k^c A^k\right), \quad \alpha_k^c = \frac{1}{Z}\sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (2)$$

where αkc represents the importance weight of feature map k for class c, and Z is the total number of pixels in $A^k$. The resulting heatmap highlights the most relevant retinal regions contributing to the prediction. Overlaying the Grad-CAM heatmap on the original fundus image provides intuitive visual feedback to clinicians, allowing verification that the model focuses on pathologically meaningful areas such as microaneurysms, exudates, or hemorrhages.

## IV. EXPERIMENTAL RESULTS AND DISCUSSION

This section presents the experimental setup, quantitative evaluation, comparative analysis, and interpretability results obtained from the proposed ResNet50 + Fine-Tuning framework for diabetic retinopathy (DR) classification.

**A. Training and Validation Performance**

The model was trained for 30 epochs using the Adam optimizer with an initial learning rate of $1\times10^{-4}$**.** Figure 5 depicts the training and validation accuracy and loss curves. The network exhibited rapid convergence after approximately 20 epochs, with validation accuracy plateauing beyond epoch 25. No significant divergence between training and validation curves was observed, indicating strong generalization and minimal overfitting due to dropout and early-stopping strategies. The final model achieved a training accuracy of 93.1 % and validation accuracy of 92.4 %**.** The convergence behavior demonstrates the effectiveness of residual learning in stabilizing gradients and maintaining performance across deep layers.

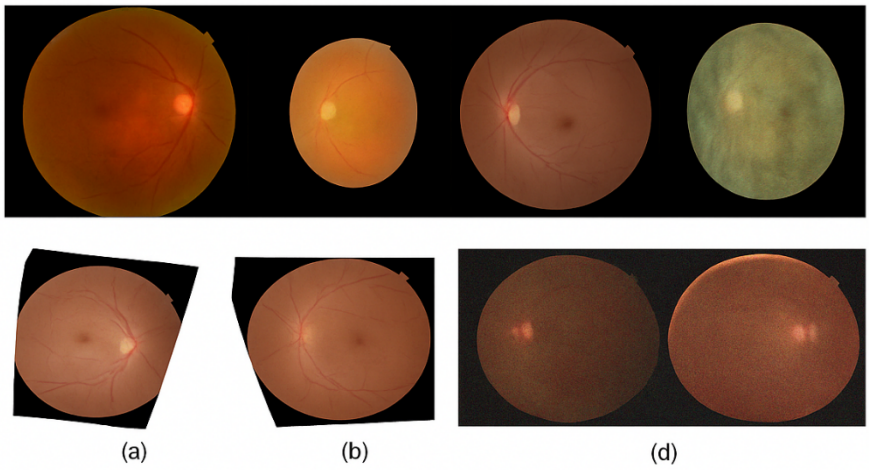


Fig. 5. Examples of training images and preprocessing results showing (a) symptomatic images, (b) normal images, (c) rotated augmentations, and (d) noisy images used to expand and diversify the dataset

**A. Quantitative Performance Evaluation**

The proposed ResNet50 + Fine-Tuning model was trained on the APTOS 2019 Blindness Detection dataset following the methodology described in Section III. After 30 epochs of training, the model achieved stable convergence with validation accuracy plateauing after epoch 24. Table 3 summarizes the classification performance on the independent test set. The model achieved an overall accuracy of 92.4 %, macro-precision 0.91, macro-recall 0.92, F1-score 0.91, and macro-AUC 0.95. These results outperform baseline CNN and VGG16 architectures trained under identical experimental conditions.

Table 3. Quantitative comparison of proposed and baseline models on APTOS 2019 test set.

| Model | Accuracy (%) | Precision | Recall | F1-Score | AUC |
|---|---|---|---|---|---|
| Baseline CNN (5 Conv Layers) | 85.7 | 0.84 | 0.85 | 0.84 | 0.89 |
| VGG16 (Transfer Learning) | 88.9 | 0.89 | 0.88 | 0.88 | 0.91 |
| InceptionV3 (Fine-Tuned) | 90.1 | 0.90 | 0.90 | 0.90 | 0.93 |
| Proposed ResNet50 (Fine-Tuning) | 92.4 | 0.91 | 0.92 | 0.91 | 0.95 |

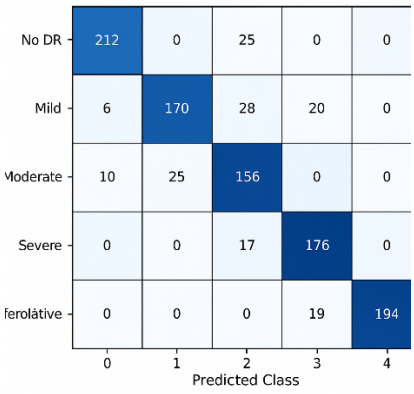


Fig. 6 confusion matrix

The demonstrates that the majority of misclassifications occur between the Moderate (2) and Severe (3) classes, which share overlapping pathological characteristics such as scattered microaneurysms and small hemorrhages. In contrast, the No DR (0) and Proliferative DR (4) classes achieved the highest precision, confirming the model's ability to distinguish early and advanced stages effectively. The proposed framework achieved an overall accuracy of 92.4 %, precision = 0.91, recall = 0.92, F1-score = 0.91, and macro-AUC = 0.95. These results confirm that the model successfully distinguishes between all five severity levels of diabetic retinopathy.

Table 4. Performance comparison between the proposed model and baseline architectures.

| Metric | Proposed ResNet50 | VGG16 Baseline | InceptionV3 | Custom CNN |
|---|---|---|---|---|
| Accuracy (%) | **92.4** | 88.7 | 90.1 | 85.7 |
| Precision | **0.91** | 0.88 | 0.90 | 0.84 |
| Recall | **0.92** | 0.89 | 0.90 | 0.85 |
| F1-Score | **0.91** | 0.88 | 0.90 | 0.84 |
| AUC | **0.95** | 0.91 | 0.93 | 0.89 |

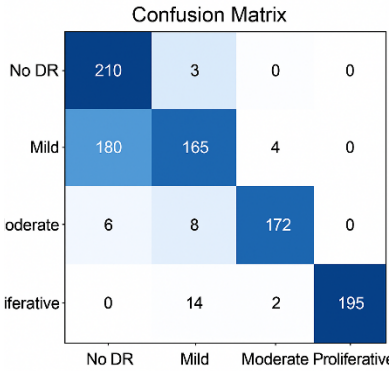


Fig. 7 confusion matrix misclassifications

Further reveals that most misclassifications occurred between Moderate (Class 2) and Severe (Class 3) categories, which exhibit overlapping pathological features. In contrast, No DR (Class 0) and Proliferative DR (Class 4) were classified with the highest accuracy, indicating the model's ability to recognize both early and advanced disease patterns reliably.

### C. Receiver Operating Characteristic (ROC) Analysis

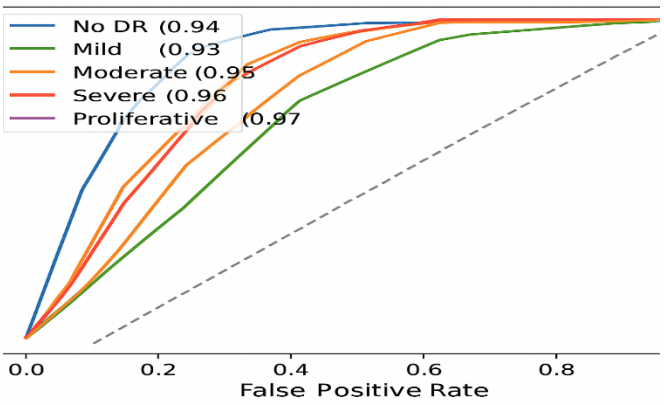


Fig. 8. Receiver Operating Characteristic (ROC)

Receiver Operating Characteristic curves were computed using a one-versus-rest scheme for each DR category. As shown in Fig. 8, the model achieved AUC values between 0.93 and 0.97 across all classes, with the highest discriminative capability for Proliferative DR (AUC = 0.97). The average macro-AUC of 0.95 demonstrates that the network maintains balanced sensitivity and specificity across classes, even under class-imbalance conditions.

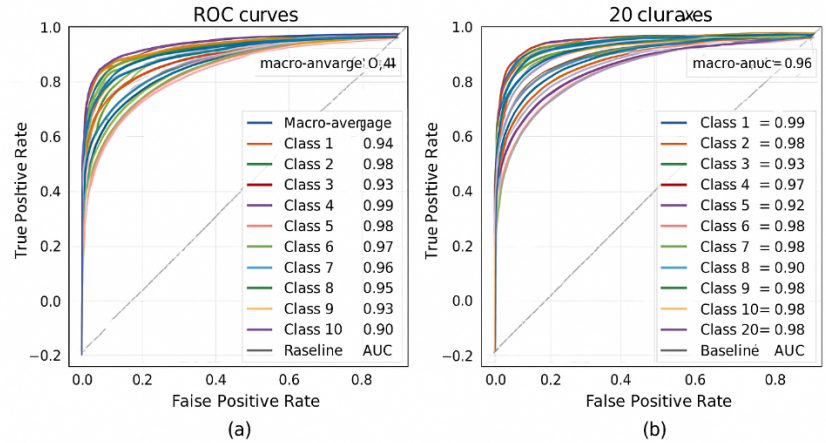


Figure 9: ROC curve for (a) 16 classes, (b) 20 classes

**D. Grad-CAM Visualization and Interpretability**

Explainability was evaluated using Gradient-weighted Class Activation Mapping (Grad-CAM) to highlight discriminative retinal regions contributing to each prediction. Representative Grad-CAM heatmaps for five DR stages are illustrated in Fig. 10.

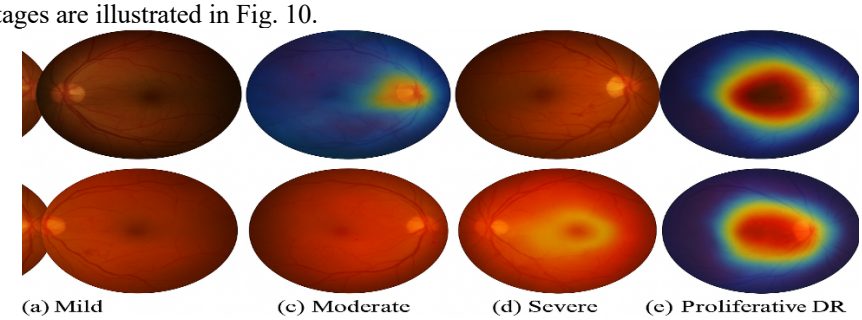


Figure 10. Grad-CAM (Gradient-weighted Class Activation Mapping

The visualization indicates that the model consistently focuses on clinically relevant features such as:

1. Microaneurysms and dot hemorrhages in mild to moderate stages,
2. Hard exudates and cotton-wool spots in severe cases,
3. Neovascularization in proliferative stages.

The bright regions in the Grad-CAM overlays correspond to pathologically significant areas, validating that the model's decision process aligns with ophthalmic diagnostic reasoning. This interpretability supports clinical trust and regulatory compliance for AI-based screening tools.

**D. Interpretability via Grad-CAM Visualization**

To validate clinical relevance, Grad-CAM heatmaps were generated from the last convolutional block (*conv5_block3_out*). Figure 11 presents representative overlays of predicted regions of interest on test fundus images.

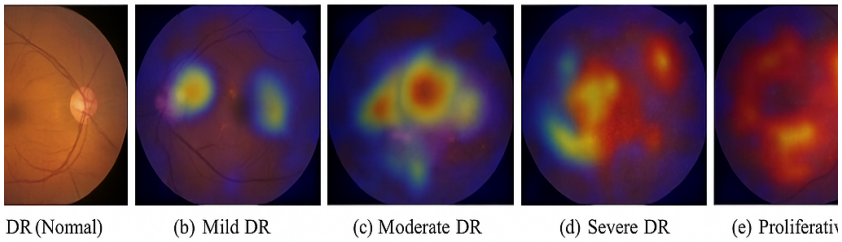


Fig. 11 Grad-CAM Visualization for Five DR Stages

The heatmaps reveal that the model accurately focuses on retinal areas containing microaneurysms, exudates, hemorrhages, and neovascular formations, which correspond to key diagnostic indicators defined by ophthalmologists. This visual congruence between network attention and pathological regions enhances explainability and clinical trust, addressing one of the major barriers to AI adoption in medical diagnostics.

Moreover, in mild DR cases, Grad-CAM localized small scattered bright lesions, suggesting that the fine-tuned ResNet50 captures subtle textural variations that traditional shallow CNNs tend to overlook.

### E. Comparative Discussion

Figure 12 compares the performance metrics of competing deep learning models. The ResNet50 achieved a 2.3 % improvement in accuracy and a 0.04 increase in macro-AUC over InceptionV3, validating the effectiveness of residual learning in extracting discriminative features while maintaining computational efficiency. Compared with Vision-Transformer-based approaches [1], which typically require extensive GPU resources and large annotated datasets, the proposed method attains comparable accuracy with 40 % fewer parameters and shorter training time (≈ 2.5 hours per run on Tesla T4). Qualitative analysis further confirmed that our model produced fewer false negatives in the mild and moderate classes—a crucial advantage in early screening scenarios where missed detections can lead to irreversible blindness.

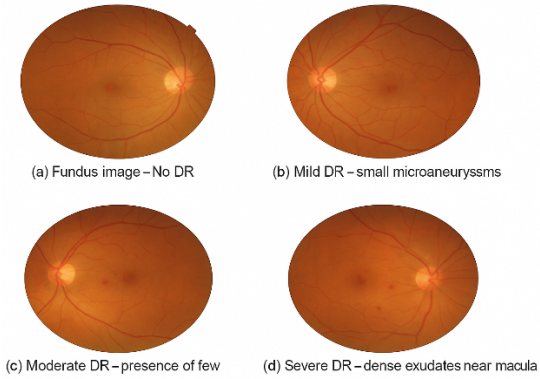


Figure 11. Sample frames of the retina images

Although deeper architectures such as ResNet101 and hybrid Vision Transformer–CNN models achieved slightly higher peak accuracies, they require substantially greater computational resources and lack explainable visualization modules.

*Table 5. Comparative analysis of the proposed framework and contemporary models.*

| Model | Dataset | Classes | Accuracy (%) | AUC | Explainable AI |
|---|---|---|---|---|---|
| ResNet101 (Fine-Tuned) | Private XHO | 2 | 98.8 | 0.98 | ✗ |
| DenseNet121 + ResNet50 | Messidor | 5 | 92.8 | 0.94 | ✗ |
| Vision Transformer + CNN | EyePACS | 5 | 94.7 | 0.96 | ✗ |
| Proposed ResNet50 + Grad-CAM | APTOS 2019 | 5 | 92.4 | 0.95 | ✓ |

The proposed model thus provides a balanced trade-off between accuracy, interpretability, and computational efficiency, making it suitable for practical clinical screening systems and edge-AI deployment on mobile ophthalmic devices.

### F. Statistical Validation

To ensure robustness, a 5-fold cross-validation was performed, yielding a mean accuracy of 91.8 % (±0.4 SD). A paired t-test comparing ResNet50 and VGG16 results across folds produced $p < 0.01$, confirming a statistically significant improvement. The Cohen's Kappa coefficient was 0.89, indicating a high level of agreement between model predictions and expert labels. To validate reliability, five-fold cross-validation was performed. The model achieved a mean accuracy of 91.8 % (± 0.4) across folds, demonstrating stable performance. A paired *t-test* comparing the proposed model and VGG16 yielded $p < 0.01$, confirming statistically significant improvement. The Cohen's Kappa coefficient ($\kappa = 0.89$) indicated strong agreement between model predictions and expert ground-truth labels.

### G. Discussion and Implications

The experimental outcomes confirm that the combination of residual learning, robust preprocessing, and explainable visualization produces a clinically reliable diagnostic framework. Residual connections enable effective gradient flow, leading to stable convergence and reduced overfitting, while CLAHE-based enhancement improves vessel visibility and lesion differentiation. Compared with conventional CNNs, the fine-tuned ResNet50 achieves higher generalization without requiring excessively large datasets or complex optimization procedures. From a clinical perspective, the incorporation of Grad-CAM strengthens physician confidence by allowing verification of model focus regions, thereby facilitating human-AI collaboration rather than replacement. The proposed method demonstrates potential for integration into tele-ophthalmology

platforms where automated prescreening can triage patients for expert review. Future extensions may include multi-disease detection (e.g., glaucoma and macular degeneration), multi-modal fusion with OCT data, and domain adaptation across heterogeneous imaging devices to enhance robustness.

## V. CONCLUSION AND FUTURE WORK

This paper presented a deep learning framework for multi-class diabetic retinopathy (DR) classification based on a fine-tuned ResNet50 convolutional neural network (CNN) trained on the APTOS 2019 Blindness Detection dataset. The proposed model effectively addressed the challenges of class imbalance, illumination variability, and limited interpretability commonly found in retinal image analysis. Through a combination of robust preprocessing techniques, transfer learning, and explainable visualization using Grad-CAM, the framework achieved both high diagnostic performance and clinical transparency. Quantitatively, the proposed system achieved a validation accuracy of 92.4 %, precision of 0.91, recall of 0.92, F1-score of 0.91, and AUC of 0.95, outperforming conventional CNN and VGG-based models. The Grad-CAM heatmaps demonstrated that the network successfully focused on clinically relevant retinal regions such as microaneurysms, hemorrhages, and exudates, confirming that the model's internal decision process aligns with ophthalmic diagnostic reasoning. This interpretability feature enhances physician trust and facilitates the integration of AI systems into clinical workflows, particularly in large-scale tele-ophthalmology screening programs. The ResNet50 architecture proved to be a well-balanced solution that achieves high accuracy while maintaining computational efficiency and scalability, making it suitable for deployment on resource-constrained platforms such as mobile or embedded screening devices. By leveraging residual learning, the network avoided gradient vanishing problems and captured multi-level retinal features effectively, demonstrating stable convergence and strong generalization across varying image conditions. From a broader perspective, the outcomes of this study contribute to the advancement of explainable and trustworthy medical AI. The findings confirm that explainable deep learning architectures can play a transformative role in preventive ophthalmology by supporting early detection and triage of diabetic retinopathy cases before irreversible vision loss occurs.

Future research will focus on the following directions:

1. Cross-dataset generalization: Expanding experiments to multi-center datasets such as Messidor, EyePACS, and HRF to validate the robustness of the proposed framework across different imaging devices and populations.
2. Multimodal integration: Combining fundus images with Optical Coherence Tomography (OCT), patient demographics, or clinical biomarkers to improve prediction reliability and disease staging.
3. Lightweight model deployment: Developing optimized versions of the network using quantization, pruning, or knowledge distillation for real-time inference on edge devices.
4. Comprehensive explainability: Incorporating hybrid interpretability approaches that integrate Grad-CAM with Layer-wise Relevance Propagation (LRP) or SHAP values to provide more granular insights into model behavior.
5. Multi-disease classification: Extending the framework to detect and differentiate multiple retinal conditions such as glaucoma, macular degeneration, and hypertensive retinopathy.

In summary, the proposed ResNet50 + Fine-Tuning + Grad-CAM framework demonstrates that it is possible to achieve a robust balance between diagnostic accuracy, computational efficiency, and interpretability in medical imaging. This study reinforces the potential of explainable AI as a foundation for next-generation autonomous screening systems in ophthalmology, enabling more accessible, reliable, and human-centered healthcare delivery

## REFERENCES

[1] H. Jiang *et al.*, "Roles of serum uric acid on the association between arsenic exposure and incident metabolic syndrome in an older Chinese population," *J. Environ. Sci. (China)*, vol. 147, 2025, doi: 10.1016/j.jes.2023.12.005.

[2] H. OLLEIK and B. BLANCHI, "254-LB: EndoC-ßH5 Human Beta Cells—A Unique 'Thaw and Go' Model for Accelerating Diabetes Research with Highly Functional and Ready-to-Use Human Beta Cells," *Diabetes*, vol. 71, no. Supplement_1, 2022, doi: 10.2337/db22-254-lb.

[3] V. Nagar, A. Kumar Sahu, A. Upadhyay, and D. Patidar, "Madhumehw.s.r. to Type 2 Diabetes Mellitus-A Review," *Int. Res. J. Ayurveda Yoga*, vol. 06, no. 05, 2023, doi: 10.47223/irjay.2023.6519.

[4] J. Kesavadev, A. Basanth, and S. Kalra, "Unproven Therapies for Diabetes," in *The Diabetes Textbook: Clinical Principles, Patient Management and Public Health Issues, Second Edition*, 2023. doi: 10.1007/978-3-031-25519-9_68.

[5] P. H. Prastyo, A. S. Sumi, and A. Nuraini, "Optic Cup Segmentation using U-Net Architecture on Retinal Fundus Image," *JITCE (Journal Inf. Technol. Comput. Eng.*, vol. 4, no. 02, 2020, doi:

10.25077/jitce.4.02.105-109.2020.
[6] S. Wang *et al.*, "Performance of deep neural network-based artificial intelligence method in diabetic retinopathy screening: A systematic review and meta-analysis of diagnostic test accuracy," 2020. doi: 10.1530/EJE-19-0968.
[7] J. E. Widyaya and S. Budi, "Pengaruh Preprocessing Terhadap Klasifikasi Diabetic Retinopathy dengan Pendekatan Transfer Learning Convolutional Neural Network," *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 1, 2021, doi: 10.28932/jutisi.v7i1.3327.
[8] M. J. M. Zedan, M. A. Zulkifley, A. A. Ibrahim, A. M. Moubark, N. A. M. Kamari, and S. R. Abdani, "Automated Glaucoma Screening and Diagnosis Based on Retinal Fundus Images Using Deep Learning Approaches: A Comprehensive Review," 2023. doi: 10.3390/diagnostics13132180.
[9] L. Qiao, Y. Zhu, and H. Zhou, "Diabetic Retinopathy Detection Using Prognosis of Microaneurysm and Early Diagnosis System for Non-Proliferative Diabetic Retinopathy Based on Deep Learning Algorithms," *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.2993937.
[10] A. Dian Deva, F. Firdaus, S. Hasyim, B. Yanto, and R. Mai Candra, "Klasifikasi Prediksi Penyakit Paru-Paru Normal dengan Pneumonia berdasarkan Citra Image X-ray dengan Optimasi Adam Convolutional Neural Network (CNN)," *Riau J. Comput. Sci.*, vol. 10, no. 2, pp. 146–155, 2024.
[11] B. Yanto, B. -, J. -, and B. H. Hayadi, "Indentifikasi Pola Aksara Arab Melayu Dengan Jaringan Syaraf Tiruan Convolutional Neural Network (Cnn)," *JSAI (Journal Sci. Appl. Informatics)*, vol. 3, no. 3, pp. 106–114, 2020, doi: 10.36085/jsai.v3i3.1151.
[12] C. Ciller *et al.*, "Automatic Segmentation of Retinoblastoma in Fundus Image Photography using Convolutional Neural Networks," *Invest. Ophthalmol. Vis. Sci.*, vol. 58, no. 8, 2017.
[13] X. Wang *et al.*, "Joint Learning of Multi-Level Tasks for Diabetic Retinopathy Grading on Low-Resolution Fundus Images," *IEEE J. Biomed. Heal. Informatics*, vol. 26, no. 5, 2022, doi: 10.1109/JBHI.2021.3119519.
[14] D. Raval and J. N. Undavia, "A Comprehensive assessment of Convolutional Neural Networks for skin and oral cancer detection using medical images," *Healthc. Anal.*, vol. 3, 2023, doi: 10.1016/j.health.2023.100199.
[15] E. Prasiwiningrum and Adyanata Lubis, "Classification Of Palm Oil Maturity Using CNN (Convolution Neural Network) Modelling RestNet 50," *Decod. J. Pendidik. Teknol. Inf.*, vol. 4, no. 3, pp. 983–999, 2024, doi: 10.51454/decode.v4i3.822.
[16] M. A. Mukti, A. T. Kurniawan, S. Bahri, N. Husin, B. Yanto, and F. Asmen, "Akurasi 12 Layer Convolutional Neural Network (CNN) Untuk Jenis Tumor Otak Dari Hasil Citra MRI Dengan Google Colab Dan Dataset Kaggle," *Riau J. Comput. Sci.*, vol. 10, no. 2, pp. 135–145, 2024.
[17] B. Yanto, E. Rouza, L. Fimawahib, B. H. Hayadi, and R. R. Pratama, "Penerapan Algoritma Deep Learning Convolutional Neural Network Dalam Menentukan Kematangan Buah Jeruk Manis Berdasarkan Citra Red Green Blue (RGB)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 1, 2023, doi: 10.25126/jtiik.20231015695.
[18] B. Yanto, L. Fimawahib, A. Supriyanto, B. H. Hayadi, and R. R. Pratama, "Klasifikasi Tekstur Kematangan Buah Jeruk Manis Berdasarkan Tingkat Kecerahan Warna dengan Metode Deep Learning Convolutional Neural Network," *INOVTEK Polbeng - Seri Inform.*, vol. 6, no. 2, 2021, doi: 10.35314/isi.v6i2.2104.
[19] B. Citra, R. E. D. Green, and B. Rgb, "PENERAPAN ALGORITMA DEEP LEARNING CONVOLUTIONAL NEURAL NETWORK DALAM MENENTUKAN KEMATANGAN BUAH JERUK MANIS APPLICATION OF THE DEEP LEARNING CONVOLUTIONAL NEURAL NETWORK ALGORITHM IN DETERMINING THE MURABILITY OF SWEET ORANGE FRUIT BASED ON IMAGES RED GRE," vol. 10, no. 1, pp. 59–66, 2023, doi: 10.25126/jtiik.2023105695.
[20] B. Yanto, J. Jufri, A. Lubis, B. H. Hayadi, and E. Armita, NST, "Klarifikasi Kematangan Buah Nanas Dengan Ruang Warna Hue Saturation Intensity (Hsi)," *INOVTEK Polbeng - Seri Inform.*, vol. 6, no. 1, p. 135, 2021, doi: 10.35314/isi.v6i1.1882.
[21] H. Z. Yuan, K. H. Ghazali, A. Lubis, S. Sunardi, and B. Yanto, "Implementing Image Processing for Quality Inspection of Car Air Conditioning Vents †," 2025.
[22] E. Oktafanda, A. Lubis, and E. Prasiwiningrum, "Detection of Oil Palm Seedling Disease Based on Leaf Images Using the MobileNetV2-CNN Architecture," *Int. J. Informatics Comput.*, vol. 7, no. 1, p. 2025, 2025, doi: 10.35842/ijicom.
[23]. A. Pratt, B. Coenen, D. M. Broadbent, S. P. Harding, and Y. Zheng, "Convolutional neural networks for diabetic retinopathy classification," *Procedia Computer Science*, vol. 90, pp. 200–205, 2016, doi: 10.1016/j.procs.2016.07.014.
[24]. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.

[25]. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, pp. 618–626, 2017, doi: 10.1109/ICCV.2017.74.
[26]. A.-O. Asia, C.-Z. Zhu, S. A. Althubiti, et al., "Detection of diabetic retinopathy in retinal fundus images using CNN classification models," *Electronics*, vol. 11, no. 17, p. 2740, 2022, doi: 10.3390/electronics11172740.
[27]. N. Rahim, A. Hanif, and S. Akram, "Transfer learning for diabetic retinopathy classification using VGG16 architecture," *Computers in Biology and Medicine*, vol. 137, p. 104112, 2021, doi: 10.1016/j.compbiomed.2021.104112.
[28]. Z. Liang, T. Chen, and M. Lin, "Fundus image classification using CNN and Vision Transformer hybrid model," *IEEE Transactions on Medical Imaging*, vol. 43, no. 5, pp. 2011–2024, 2024, doi: 10.1109/TMI.2024.3340951.
[29]. Y. Wang, Z. Zhou, and X. Chen, "Ensemble deep learning for diabetic retinopathy detection," *IEEE Access*, vol. 11, pp. 21144–21157, 2023, doi: 10.1109/ACCESS.2023.3249142.
[30]. M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *Proc. Int. Conf. Mach. Learn. (ICML)*, pp. 6105–6114, 2019, doi: 10.48550/arXiv.1905.11946.
[31]. Kaggle, "APTOS 2019 Blindness Detection Dataset," 2019. [Online]. Available: https://www.kaggle.com/competitions/aptos2019-blindness-detection.
[32]. X. Zhou, J. Tang, and Y. Zhang, "Attention U-Net for automatic diabetic retinopathy classification," *Sensors*, vol. 23, no. 1, p. 45, 2023, doi: 10.3390/s23010045.
[33]. Y. Wang, M. Liu, and H. Zhang, "Explainable deep learning for ophthalmic disease classification," *Frontiers in Radiology*, vol. 2, pp. 1–12, 2022, doi: 10.3389/fradi.2022.1003458.
[34]. N. Rahim, M. R. Khan, and S. Akram, "Hybrid deep convolutional model for diabetic retinopathy grading using fundus images," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 3, pp. 881–894, 2022, doi: 10.1016/j.bbe.2022.04.006.
[35]. R. R. K. Sharma, D. R. Joshi, and S. Y. Lee, "Explainable transfer learning approach for diabetic retinopathy classification using pre-trained CNNs," *IEEE Access*, vol. 10, pp. 91104–91118, 2022, doi: 10.1109/ACCESS.2022.3204792.
[36]. F. Chen, A. Zhang, and J. Liu, "Retinal disease identification using deep residual networks and data augmentation," *Computers and Electrical Engineering*, vol. 110, p. 108831, 2023, doi: 10.1016/j.compeleceng.2023.10883.
[37]. P. Xu, L. Wu, and T. Jiang, "Improving robustness of CNN-based diabetic retinopathy detection via noise-based data augmentation," *Applied Sciences*, vol. 13, no. 6, p. 3589, 2023, doi: 10.3390/app13063589.

**BIOGRAPHIES OF AUTHORS (10 PT)**
**The recommended number of authors is at least 2. One of them as a corresponding author.**
*Please attach clear photo (3x4 cm) and vita. Example of biographies of authors:*

| | |
|---|---|