Indonesian Journal of Data and Science

Volume 6 Issue 1
ISSN 2715-9936
https://doi.org/10.56705/ijodas.v6i1.232

Research Article

Comparison of ResNet-50 and DenseNet-121 Architectures in Classifying Diabetic Retinopathy

I Putu Gede Yoga Pramana Putra [1,*], Ni Wayan Jeri Kusuma Dewi [2], Putu Surya Wedra Lesmana [3], I Gede Totok Suryawan [4], Putu Satria Udyana Putra [5]

[1] Institut Bisnis dan Teknologi Indonesia, Kota Denpasar, Bali 80225, Indonesia, yogapramanaputra26@gmail.com
[2] Institut Bisnis dan Teknologi Indonesia, Kota Denpasar, Bali 80225, Indonesia, wayan.kusumadewi@instiki.ac.id
[3] Institut Bisnis dan Teknologi Indonesia, Kota Denpasar, Bali 80225, Indonesia, suryawedra@stiki-indonesia.ac.id
[4] Institut Bisnis dan Teknologi Indonesia, Kota Denpasar, Bali 80225, Indonesia, totok.suryawan@instiki.ac.id
[5] Institut Bisnis dan Teknologi Indonesia, Kota Denpasar, Bali 80225, Indonesia, satria@instiki.ac.id

Correspondence should be addressed to I Putu Gede Yoga Pramana Putra; yogapramanaputra26@gmail.com

Received 22 December 2024; Accepted 20 March 2025; Published 31 March 2025

Abstract:

Introduction: Diabetic Retinopathy (DR) is a vision-threatening complication of diabetes that requires early and accurate diagnosis. Deep learning offers promising solutions for automating DR classification from retinal images. This study compares the performance of two convolutional neural network (CNN) architectures—ResNet-50 and DenseNet-121—for classifying DR severity levels. **Methods**: A dataset of 2,000 pre-processed and augmented retinal images was used, categorized into four classes: normal, mild, moderate, and severe. Both models were trained using two approaches: standard train-test split and Stratified K-Fold Cross Validation (k=5). Data augmentation techniques such as flipping, rotation, zooming, and translation were applied to enhance model generalization. The models were trained using the Adam optimizer with a learning rate of 0.001, dropout of 0.2, and learning rate adjustment via ReduceLROnPlateau. Performance was evaluated using accuracy, precision, recall, and F1-score. **Results**: ResNet-50 outperformed DenseNet-121 across all evaluation metrics. Without K-Fold, ResNet-50 achieved 84% accuracy compared to DenseNet-121's 80%; with K-Fold, ResNet-50 scored 83% and DenseNet-121 81%. ResNet-50 also demonstrated better balance in class-wise classification, with higher recall and F1-score, especially for moderate and severe DR classes. Confusion matrices confirmed fewer misclassifications with ResNet-50. **Conclusions**: ResNet-50 provides superior accuracy and robustness in classifying DR severity levels compared to DenseNet-121. While K-Fold Cross Validation enhances model stability, it slightly reduces overall accuracy. These findings support the use of ResNet-50 in developing reliable deep learning-based screening tools for early DR detection in clinical practice.

Keywords: Deep Learning, DenseNet-121, Diabetic Retinopathy, Image Classification, ResNet-50.
Dataset link: https://www.kaggle.com/datasets/tanlikesmath/diabetic-retinopathy-resized

1. Introduction

Diabetic Retinopathy (DR) is a major complication of diabetes that can lead to blindness if not detected and treated early [1]. The increasing prevalence of diabetes has made DR a growing concern, with approximately 1.5% of the Indonesian population affected [2]. [3] Early detection is critical but remains challenging due to the limitations of manual diagnosis, which relies on expert evaluation, is time-consuming, and is prone to human error [4].

Deep learning, particularly Convolutional Neural Networks (CNNs), has demonstrated remarkable success in medical image analysis, offering a potential solution for automating DR classification [5]. Among CNN architectures, ResNet-50 and DenseNet-121 are widely used due to their ability to extract deep features from images [6]. [7] ResNet-50 introduces residual learning, allowing deeper networks without vanishing gradients, while DenseNet-121 employs dense connectivity to improve feature propagation and efficiency [8]. However, few studies have directly compared these architectures in DR classification, highlighting a gap in research [9].

This study aims to compare ResNet-50 and DenseNet-121 in classifying Diabetic Retinopathy using retinal images [10], [11]. [12] A dataset consisting of 2000 retinal images categorized into four DR severity levels is used, and the models are evaluated using Stratified K-Fold Cross Validation (k=5) and a non-K-Fold approach [13]. Performance is assessed through accuracy, precision, recall, and F1-score, providing insights into the most effective model for automated DR detection [14], [15].

The key contribution of this research lies in its comparative analysis of these architectures, offering a benchmark for future deep learning-based DR classification systems [16]. The findings can help optimize AI-driven early screening tools, supporting ophthalmologists in accurate and efficient diagnosis [17].

2. Method:

This study employs a quantitative experimental approach by implementing deep learning for the classification of Diabetic Retinopathy (DR) based on retinal images. Two Convolutional Neural Network (CNN) architectures, ResNet-50 and DenseNet-121, are compared to evaluate their effectiveness in classifying DR severity levels. The models are tested using two validation strategies.

a. Stratified K-Fold Cross Validation (k=5) – to address class imbalance in the dataset and enhance model generalization.
b. Without K-Fold – as a comparison to evaluate the model's performance using a standard dataset split.

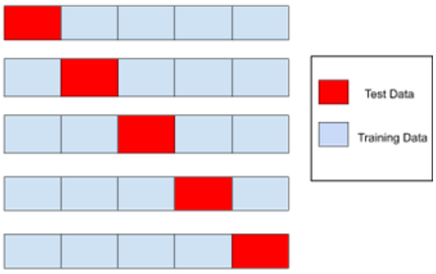

Figure 1. *K-Fold Cross Validation*

Model evaluation is conducted using accuracy, precision, recall, and F1-score, with further analysis using a confusion matrix to understand misclassification patterns for each DR category.

Data Selection

The dataset used in this study is Processed_Augmented_Dataset, sourced from Kaggle, which has undergone pre-processing and data augmentation. The dataset consists of 2000 retinal images, categorized into four classes:

a. Normal Eye
b. Mild DR
c. Moderate DR
d. Severe DR

Data augmentation was applied to address class imbalance in the raw dataset and to improve the model's generalization capability. The process aimed to generate new data variations from existing images without altering their labels. The augmentation techniques included random rotation, vertical/horizontal flipping, zoom in/out, and translation, resulting in a total of 6,800 augmented images.

The dataset is divided into three main parts:

a. Training data: Used to train the model
b. Validation data: Used to prevent overfitting during training
c. Testing data: Used to evaluate the final model performance

For the non-K-Fold method, the dataset is split into 80:20 for training and validation. In K-Fold Cross Validation, the dataset is divided into five folds, ensuring that each class is evenly distributed across all iterations of training. To evaluate model performance, 5-Fold Cross Validation was used. In this method, the dataset was split into five equal parts. Each iteration used four folds (5,120 images) for training and one-fold (1,280 images) for validation. This approach, applied on 80% of the total dataset (6,400 images), is considered one of the most reliable techniques for estimating a model's ability to generalize on unseen data.

Tools and Technology Used

This study utilizes various tools and technologies to build, train, and evaluate the deep learning models, including:

a. Programming Language: Python

b. Optimizer Adam [18]

c. Deep Learning Frameworks: TensorFlow and Keras [19]

d. Pretrained Models: ResNet-50 and DenseNet-121

e. Augmentation Techniques: ImageDataGenerator to enhance image variability

Call-backs Used:

a. EarlyStopping: Stops training if no improvement in performance is detected.
b. ModelCheckpoint: Saves the best model based on validation performance.
c. ReduceLROnPlateau: Reduces the learning rate if the performance stagnates.

Model training is performed using GPU acceleration to speed up computations, as deep learning requires high processing power for handling image data

Data Collection Process

The data collection process involves several steps:

a. Downloading the dataset from Kaggle, which has been curated and labelled based on DR severity levels.

b. Pre-processing data, including:

1. Pixel value normalization to a range of [0,1] to match CNN input requirements.
2. Resizing images to 224×224 pixels to fit the ResNet-50 and DenseNet-121 architectures.
3. Data augmentation using ImageDataGenerator to increase image variations, such as rotation, flipping, and brightness adjustments

c. Splitting the dataset into training, validation, and testing according to the validation method used.

d. Training the models using the preprocessed dataset, with hyperparameter tuning to optimize model performance.

Data Analysis Methods

Model evaluation is conducted using key performance metrics to assess prediction quality:

a. Accuracy: Measures the percentage of correct predictions compared to the total test data [20].

$$Accuracy \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

b. Precision: Calculates the model's accuracy in identifying each DR class [21].

$$Precision \frac{TP}{TP + FP} \tag{2}$$

c. Recall: Measures how well the model recognizes each DR category [22].

$$\boldsymbol{R}ecall \frac{TP}{TP + FN} \tag{3}$$

d. F1-Score: A combination of precision and recall to ensure balanced evaluation [23].

$$\boldsymbol{F}1 - Score = 2.\frac{Precision\ .\ Recall}{Precision +\ Recall} \tag{4}$$

TP : is a prediction result that is positive and correct.
TN: is a prediction result that is negative and correct.
FP : is a prediction result that is positive but incorrect.
FN : is a prediction result that is negative but incorrect.

e. Confusion Matrix: Used to analyze misclassification patterns and determine if the model frequently misidentifies certain categories [24], [25].

The results from both models are compared to determine the most effective CNN architecture for Diabetic Retinopathy classification, as well as to assess the impact of K-Fold Cross Validation on model performance

3. Results and Discussion

Results

The dataset used in this study underwent a series of preprocessing steps to ensure data quality and consistency before training the model. The preprocessing steps included pixel value normalization to the [0,1] range, resizing images to 224×224 pixels, and data augmentation techniques such as rotation, flipping, and brightness adjustments to increase variation in retinal images.

After preprocessing, the dataset was divided into training (train), validation (validation), and testing (test) data, with an 80:20 ratio in the non-K-Fold method, while Stratified K-Fold Cross Validation (k=5) was used to improve class balance and model generalization. The model evaluation results are presented in the **Table 1**:

Table 1. Model Performance Comparison on Test Data

Model	LR	OP	Accuracy	Precision	Recall	F1-Score
ResNet-50 (Without K-Fold)	0,0001	Adam	0,84%	0,84%	0,84%	0,84%
DenseNet-121 (Without K-Fold)	0,0001	Adam	0,80%	0,80%	0,79%	0,79%
ResNet-50 (With K-Fold)	0,0001	Adam	0,83%	0,83%	0,83%	0,83%
DenseNet-121 (With K-Fold)	0,0001	Adam	0,81%	0,81%	0,81%	0,80%

From the **Table 1**, it can be seen that ResNet-50 without K-Fold achieved the highest accuracy of 84%, while DenseNet-121 without K-Fold reached 80%. The K-Fold Cross Validation method slightly reduced accuracy, but provided a more stable model in detecting different DR classes.

A confusion matrix graph was also used to observe misclassification patterns across the four DR categories.

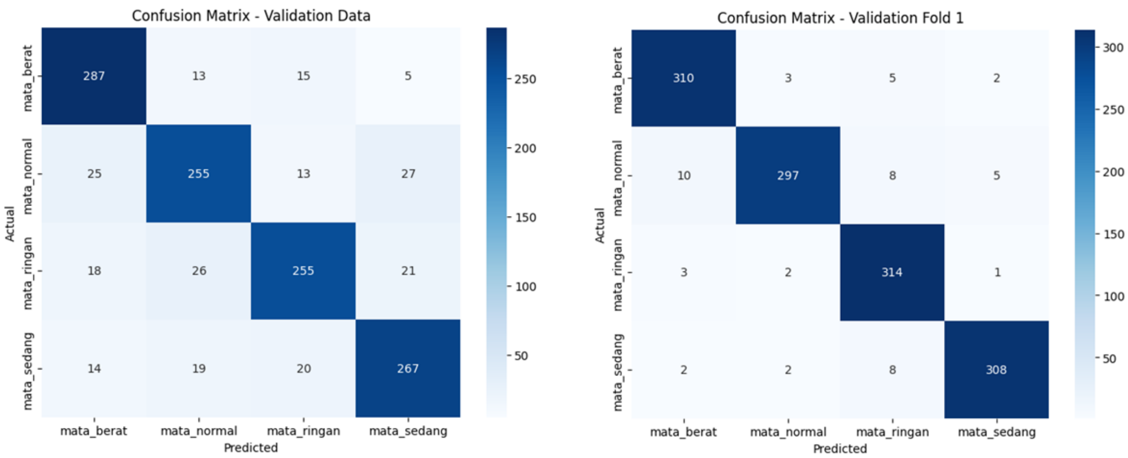

Figure 2. Confusion Matrix ResNet-50 Without (Left) and With (Right) K-Fold

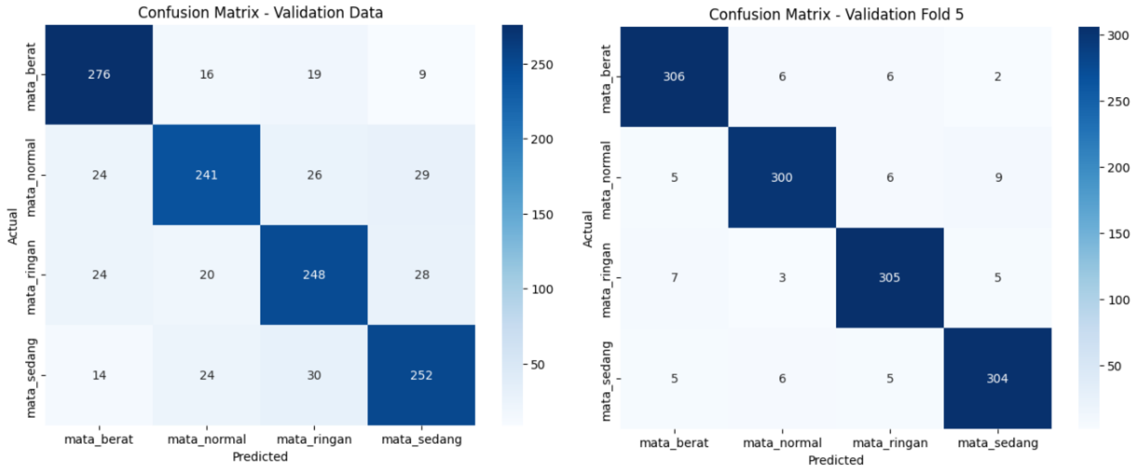

Figure 3. Confusion Matrix DenseNet-121 Without (Left) and With (Right) K-Fold

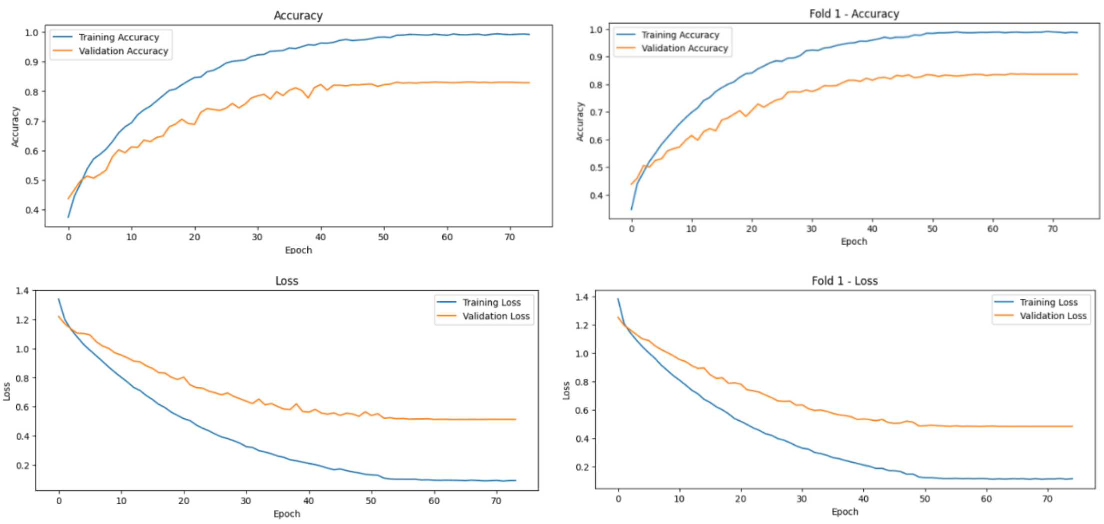

Figure 4. Accuracy and Loss ResNet-50 Without(Left) and With(Right) K-Fold

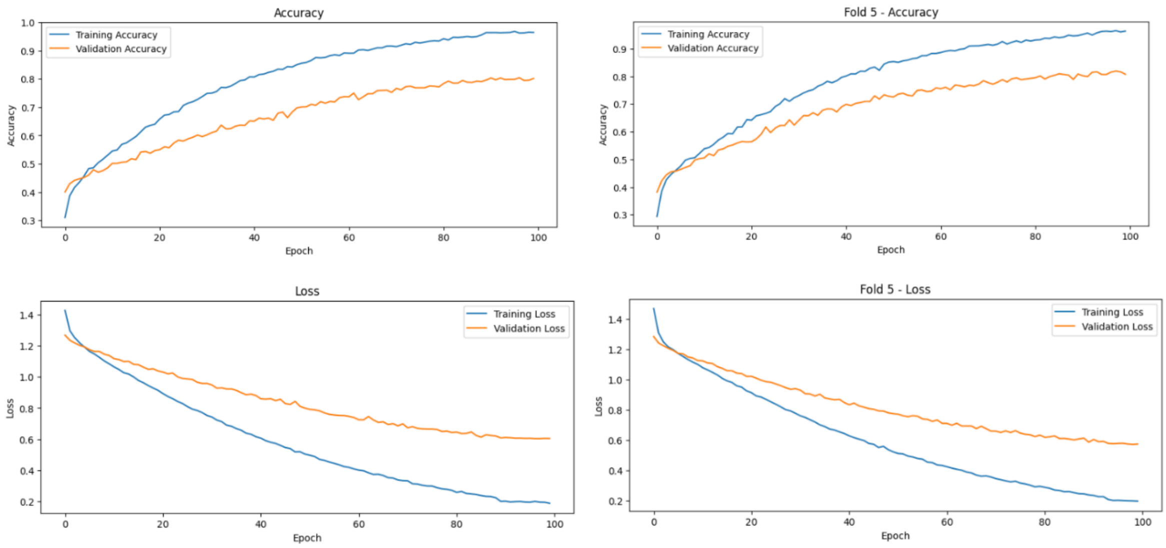

Figure 5. Accuracy and Loss DenseNet-121 Without(Left) and With(Right) K-Fold

The confusion matrix on the training and validation data for the ResNet-50 model shows strong performance in classifying images correctly. In one result, the model correctly predicted 287 images for the Severe class, 255 for Normal, 255 for Mild, and 267 for Moderate. In another result, it correctly classified 310 Severe, 297 Normal, 314 Mild, and 308 Moderate images. The dark blue diagonal in the matrix illustrates a high number of correct classifications across all classes, indicating good model accuracy.

The training and validation accuracy graph for ResNet-50 shows a consistent increase in training accuracy throughout the training process. The validation accuracy also improves initially but later fluctuates, suggesting signs of slight overfitting. Despite this, the model maintains stable performance overall. Similarly, the training loss continues to decrease, while validation loss also decreases in the early stages before fluctuating slightly. The relatively small gap between training and validation loss at the end indicates that the model generalizes well, even though some overfitting occurs.

In the case of the DenseNet-121 model, the confusion matrix also shows good classification performance. The model correctly predicted 276 Severe, 241 Normal, 248 Mild, and 252 Moderate images in one result, and 306 Severe, 300 Normal, 305 Mild, and 304 Moderate in another. The high concentration of correct predictions on the diagonal of the matrix highlights the model's reliability in recognizing each class.

The training and validation accuracy graph for DenseNet-121 indicates a steady rise in training accuracy. Validation accuracy increases early on but later fluctuates, indicating a potential for slight overfitting, similar to ResNet-50. The gap between training and validation accuracy toward the end is relatively large, suggesting that although the model learns the training data well, its performance on unseen data may vary slightly. The training loss decreases steadily, and while validation loss shows some fluctuations, the gap between them remains small, indicating the model still performs effectively with a reasonable level of generalization. ResNet-50 demonstrated better performance than DenseNet-121 in terms of accuracy and classification balance.

a. K-Fold Cross Validation produced a more stable model, but slightly reduced overall accuracy.
b. Higher recall in ResNet-50 indicates that the model is better at recognizing all DR severity levels, reducing the likelihood of undetected cases.

Significant Findings

The performance comparison between ResNet-50 and DenseNet-121 shows that ResNet-50 performs slightly better than DenseNet-121. ResNet-50 achieved a higher accuracy (84%) compared to DenseNet-121 (80%), indicating that this model makes correct predictions more frequently. Precision, recall, and F1-score are also higher in ResNet-50, meaning the model is better at recognizing each class in a more balanced manner. While DenseNet-121 still demonstrates good performance, it is slightly lower compared to ResNet-50.

The performance comparison between ResNet-50 and DenseNet-121 using K-Fold Validation also indicates that ResNet-50 performs slightly better than DenseNet-121. ResNet-50 achieved a higher accuracy (83%) compared to DenseNet-121 (81%), showing that this model makes correct predictions more frequently. Precision, recall, and F1-score are also higher in ResNet-50, indicating that the model is better at identifying each class more evenly. DenseNet-121 still performs well, but slightly lower compared to ResNet-50.

a. ResNet-50 outperforms DenseNet-121 in DR classification, particularly in accuracy and recall.
b. K-Fold Cross Validation helps mitigate class imbalance, although it slightly lowers accuracy.
c. Data augmentation contributes to improved model generalization, as evidenced by evaluation results on test data.

Discussion

Interpretation and Evaluation of the Results

Based on the experimental results, ResNet-50 proved to be more effective than DenseNet-121 in DR classification tasks. This can be attributed to the residual learning architecture in ResNet-50, which allows deeper network training without losing essential information. Additionally, the use of Stratified K-Fold Cross Validation helped address class imbalance, ensuring a more stable model in detecting all DR categories, although it slightly reduced overall accuracy.

Relationship with Previous Studies

The findings of this study align with previous research, which has shown that ResNet-50 outperforms other architectures in medical image classification, particularly in detecting fine details that differentiate disease categories. Other studies have also found that DenseNet-121 is advantageous in parameter efficiency, but in the case of DR classification, ResNet-50's residual learning approach proves to be more effective in boosting model accuracy.

Practical Implications of the Research

The results of this study have several important implications:

a. Using ResNet-50 in automated diagnostic systems can improve the speed and accuracy of early DR detection.
b. Implementing deep learning in ophthalmology can help reduce the workload of ophthalmologists by automating the analysis of thousands of retinal images.
c. K-Fold Cross Validation can be applied to other medical classification models to minimize bias caused by class imbalance.

Research Limitations

Several limitations of this study include:

a. The dataset was limited to 2000 retinal images, which may not fully represent real-world data complexity.
b. The study only compared two CNN architectures (ResNet-50 and DenseNet-121) without considering other models such as EfficientNet or Inception.
c. Computational cost analysis was not conducted, which is an important factor in implementing AI-based systems in clinics or hospitals.

Recommendations for Future Research

a. Using a larger and more diverse dataset to improve model generalization.
b. Comparing more CNN architectures to identify the best-performing model for DR classification.
c. Implementing the model in cloud-based or mobile applications, allowing for real-world medical application use.

4. Conclusion

This study compared the performance of ResNet-50 and DenseNet-121 in classifying Diabetic Retinopathy (DR) using retinal images. The findings show that ResNet-50 consistently outperformed DenseNet-121, achieving higher accuracy and better class recognition, particularly in critical DR stages like Moderate and Severe. The application of K-Fold Cross Validation improved model robustness and class balance, despite a slight decrease in overall accuracy. These results demonstrate that ResNet-50 is more effective and reliable for DR classification tasks, while also highlighting the potential of deep learning models in supporting automated early diagnosis. This research emphasizes the value of model comparison and proper validation techniques in medical imaging. For future work, using larger, more diverse datasets, evaluating other CNN architectures, and analyzing computational efficiency for clinical deployment are strongly recommended. The study contributes as a practical reference for developing AI-based diagnostic tools in ophthalmology.

Acknowledgments

The author expresses gratitude to the research supervisor. Special thanks are also extended to the dataset provider *Kaggle*, *TensorFlow* and *Keras* development community, who provided essential resources supporting this research.

References:

[1] Y. Miao and S. Tang, "Classification of Diabetic Retinopathy Based on Multiscale Hybrid Attention Mechanism and Residual Algorithm," *Wirel. Commun. Mob. Comput.*, vol. 2022, 2022, doi: 10.1155/2022/5441366.

[2] Y. Huang, L. Lin, P. Cheng, J. Lyu, R. Tam, and X. Tang, "Identifying the Key Components in ResNet-50 for Diabetic Retinopathy Grading from Fundus Images: A Systematic Investigation," *Diagnostics*, vol. 13, no. 10, 2023, doi: 10.3390/diagnostics13101664.

[3] A. Mustapha, L. Mohamed, H. Hamid, and K. Ali, "Diabetic Retinopathy Classification Using ResNet50 and VGG-16 Pretrained Networks," *Int. J. Comput. Eng. Data Sci.*, vol. 1, no. 1, pp. 2737–8543, 2021.

[4] M. Saini and S. Susan, "Diabetic retinopathy screening using deep learning for multi-class imbalanced datasets," *Comput. Biol. Med.*, vol. 149, no. August, 2022, doi: 10.1016/j.compbiomed.2022.105989.

[5] S. C. Pravin, S. P. K. Sabapathy, S. Selvakumar, S. Jayaraman, and S. V. Subramani, "An Efficient DenseNet for Diabetic Retinopathy Screening," *Int. J. Eng. Technol. Innov.*, vol. 13, no. 2, pp. 125–136, 2023, doi: 10.46604/IJETI.2023.10045.

[6] Dewi Widyawati and Amaliah Faradibah, "Comparison Analysis of Classification Model Performance in Lung Cancer Prediction Using Decision Tree, Naive Bayes, and Support Vector Machine," *Indones. J. Data Sci.*, vol. 4, no. 2, pp. 80–89, 2023, doi: 10.56705/ijodas.v4i2.76.

[7] A. Sai Bharadwaj Reddy and D. Sujitha Juliet, "Transfer learning with RESNET-50 for malaria cell-image classification," *Proc. 2019 IEEE Int. Conf. Commun. Signal Process. ICCSP 2019*, no. August, pp. 945–949, 2019, doi: 10.1109/ICCSP.2019.8697909.

[8] S. Qummar *et al.*, "A Deep Learning Ensemble Approach for Diabetic Retinopathy Detection," *IEEE Access*, vol. 7, pp. 150530–150539, 2019, doi: 10.1109/ACCESS.2019.2947484.

[9] C. L. Lin and K. C. Wu, "Development of revised ResNet-50 for diabetic retinopathy detection," *BMC Bioinformatics*, vol. 24, no. 1, pp. 1–18, 2023, doi: 10.1186/s12859-023-05293-1.

[10] S. R. Syed and M. A. Saleem Durai, "A diagnosis model for detection and classification of diabetic retinopathy using deep learning," *Netw. Model. Anal. Heal. Informatics Bioinforma.*, vol. 12, no. 1, pp. 1–18, 2023, doi: 10.1007/s13721-023-00432-3.

[11] M. M. Farag, M. Fouad, and A. T. Abdel-Hamid, "Automatic Severity Classification of Diabetic Retinopathy Based on DenseNet and Convolutional Block Attention Module," *IEEE Access*, vol. 10, no. Ml, pp. 38299–38308, 2022, doi: 10.1109/ACCESS.2022.3165193.

[12] K. Ahnaf Alavee *et al.*, "Enhancing Early Detection of Diabetic Retinopathy Through the Integration of Deep Learning Models and Explainable Artificial Intelligence," *IEEE Access*, vol. 12, no. May, pp. 73950–73969, 2024, doi: 10.1109/ACCESS.2024.3405570.

[13] S. Almas *et al.*, "Visual impairment prevention by early detection of diabetic retinopathy based on stacked auto-encoder," *Sci. Rep.*, vol. 15, no. 1, p. 2554, 2025, doi: 10.1038/s41598-025-85752-2.

[14] O. F. Gurcan, O. F. Beyca, and O. Dogan, "A comprehensive study of machine learning methods on diabetic retinopathy classification," *Int. J. Comput. Intell. Syst.*, vol. 14, no. 1, pp. 1132–1141, 2021, doi: 10.2991/IJCIS.D.210316.001.

[15] V. Dong, D. D. Sevgi, S. S. Kar, S. K. Srivastava, J. P. Ehlers, and A. Madabhushi, "Evaluating the utility of deep learning for predicting therapeutic response in diabetic eye disease," *Front. Ophthalmol.*, vol. 2, no. August, pp. 1–13, 2022, doi: 10.3389/fopht.2022.852107.

[16] S. S. Mondal, N. Mandal, K. K. Singh, A. Singh, and I. Izonin, "EDLDR: An Ensemble Deep Learning Technique for Detection and Classification of Diabetic Retinopathy," *Diagnostics*, vol. 13, no. 1, pp. 1–14, 2023, doi: 10.3390/diagnostics13010124.

[17] S. Tummala, V. S. G. Thadikemalla, S. Kadry, M. Sharaf, and H. T. Rauf, "EfficientNetV2 Based Ensemble Model for Quality Estimation of Diabetic Retinopathy Images from DeepDRiD," *Diagnostics*, vol. 13, no. 4, 2023, doi: 10.3390/diagnostics13040622.

[18] A. E. Minarno *et al.*, "Classification of Diabetic Retinopathy Disease Using Convolutional Neural Network," *Int. J. Informatics Vis.*, vol. 6, no. 1, pp. 12–18, 2022.

[19] B. T. Chicho and A. B. Sallow, "A Comprehensive Survey of Deep Learning Models Based on Keras Framework," *J. Soft Comput. Data Min.*, vol. 2, no. 2, pp. 49–62, 2021, doi: 10.30880/jscdm.2021.02.02.005.

[20] B. O. Olorunfemi *et al.*, "Efficient diagnosis of diabetes mellitus using an improved ensemble method," *Sci. Rep.*, vol. 15, no. 1, p. 3235, 2025, doi: 10.1038/s41598-025-87767-1.

[21] I. P. W. Prasetia and I Made Gede Sunarya, "Image Classification of Balinese Seasoning Base Genep Based

on Deep Learning," *J. Nas. Pendidik. Tek. Inform.*, vol. 13, no. 1, 2024, doi: 10.23887/janapati.v13i1.67967.

[22] I. P. B. G. Prasetyo Raharja, I. M. Suwija Putra, and T. Le, "Kekarangan Balinese Carving Classification Using Gabor Convolutional Neural Network," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 13, no. 1, p. 1, 2022, doi: 10.24843/lkjiti.2022.v13.i01.p01.

[23] I. G. Putra, M. Yusadara, N. Made, R. Masita, I. G. Bintang, and A. Budaya, "Performance Comparison of KNN and CNN in Classifying Balinese Gangsa Instrument Tones," vol. 8, no. 4, pp. 2233–2241, 2024.

[24] A. Bilal, G. Sun, Y. Li, S. Mazhar, and A. Q. Khan, "Diabetic Retinopathy Detection and Classification Using Mixed Models for a Disease Grading Database," *IEEE Access*, vol. 9, pp. 23544–23553, 2021, doi: 10.1109/ACCESS.2021.3056186.

[25] P. N. Chen, C. C. Lee, C. M. Liang, S. I. Pao, K. H. Huang, and K. F. Lin, "General deep learning model for detecting diabetic retinopathy," *BMC Bioinformatics*, vol. 22, pp. 1–14, 2021, doi: 10.1186/s12859-021-04005-x.