Identification of Eye Diseases from Fundus Images Using Convolutional Neural Network with ResNet50 Architecture

*Lyra Zulyanda Daulay
Department of Informatics Engineering, , Faculty of Science and Technology, Universitas Islam Negeri Sultan Syarif Kasim, Riau, Indonesia
12250120375@students.uin-suska.ac.id

Muhammad Faruq
Department of Informatics Engineering, , Faculty of Science and Technology, Universitas Islam Negeri Sultan Syarif Kasim, Riau, Indonesia
12250111791@students.uin-suska.ac.id

Muhammad Rafly Wirayudha
Department of Informatics Engineering, Faculty of Science and Technology, Universitas Islam Negeri Sultan Syarif Kasim, Riau, Indonesia
12250111489@students.uin-suska.ac.id

Abstract— This study aims to identify eye diseases from fundus images using ResNet50-based Convolutional Neural Network (CNN) architecture with a transfer learning approach. The dataset used comes from Kaggle with a total of 4217 images, covering four classes of eye diseases: Diabetic Retinopathy, Glaucoma, Cataract, and Normal. The process includes preprocessing with augmentation and normalization, transfer learning using a pre-trained ResNet50 model on imagesNet, and evaluation with a confusion matrix. The results show a testing accuracy of 87.91%, with the best performance in the Diabetic Retinopathy class and challenges in the Glaucoma class. Suggestions include balancing the dataset and further fine-tuning.

Keywords— Convolutional Neural Network, ResNet50, Eye Disease Classification, Deep Learning, Medical Image Processing

I. INTRODUCTION

The eye is one of the most important organs of the human body. In addition to being an organ, the eye also functions as a sense of sight. As one part of the body's organs, of course the eye is not free from disease attacks, whether from inside or outside the eye. The most common attack on the eye is irritation due to the entry of small objects such as dust or insects with very small sizes into the eye. In addition to irritation, there are also other diseases that if left untreated can cause blindness. Several eye diseases such as diabetic retinopathy, glaucoma, cataracts, age-related macular degeneration, ocular hypertension, and myopia are one of the main causes of blindness in humans. Early detection of eye disease is an economical and effective way to prevent blindness caused by diabetic retinopathy, glaucoma, cataracts, age-related macular degeneration, ocular hypertension, and myopia. Detection of retinal disease is done with a complete eye examination.

Computer Vision and Deep Learning can automatically detect eye disease by building a classification model using Convolutional Neural Networks (CNN). CNN is one of the Deep Learning methods for supervised learning so that this method can learn by itself and look for features or characteristics that can help identify images. This has led to CNN being widely developed. In previous research by Qulub and Agustin (2024) which used the VGG-16 architecture for fundus photo image classification with a dataset of 5,245 images, the accuracy achieved was 45% with 30 iterations, indicating challenges in detecting minority classes such as glaucoma [1]. Another study by Cahya et al. (2021) using the AlexNet architecture with a dataset of 610 images from Jr2ngb, achieved an accuracy of 98.37% after 150 epochs for the classification of normal eyes, cataracts, glaucoma, and retinal disease, demonstrating the potential of CNN with more controlled data [2]. From the problem of early detection of eye disease, this study carries out the identification of eye disease from fundus images with four classes, namely Diabetic

Retinopathy, Glaucoma, Cataract, and Normal. Based on previous research, the method that can be implemented to solve this problem is to use a Convolutional Neural Network (CNN) with the ResNet50 architecture, which is known to be able to overcome the vanishing gradient problem through shortcut connections, with preprocessing adjustments and data augmentation to improve accuracy in multi-class classification.

II. LITERATUR REVIEW

A. Fundus Photo Image

Fundus photo image is a procedure to take pictures of the fundus of the eye. Fundus photo eye action can capture images of the area behind the eye including the retina, optic nerve, macula, and retinal blood vessels [3].

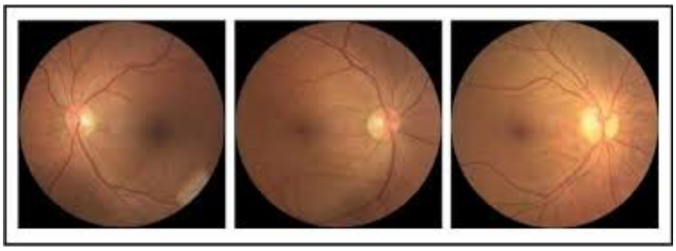

Figure 1. Example of Fundus Photo Image

B. Augmentation

Data augmentation is a technique that can prevent overfitting, which is the variation of data used that is too complex, resulting in high accuracy in training but inversely proportional to accuracy in the testing process or low accuracy in the testing process [1].

C. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is a development of Multilayer Perceptron (MLP) designed to process two-dimensional data. CNN is included in the type of Deep Neural Network because of its high network depth and is widely applied to image data. Convolutional Neural Network (CNN) has a deep feed-forward architecture that offers extraordinary generalization capabilities compared to other artificial neural networks. The complex CNN model consists of a number of special processing layers that are able to extract various features from input data, such as images. CNN has the advantage of learning very specific abstract features, especially in spatial data, allowing it to recognize patterns with high efficiency [1].

Convolutional Neural Network (CNN) is a very popular method in medical image classification, including eye fundus images. Research by Qulub and Agustin [1] used the VGG-16 architecture and obtained low accuracy due to problems with the minority class. In contrast, research by Verdy and Hartati showed the success of the ResNet-50 model in classifying eye diseases with high accuracy [4]. Cahya and Suwanda [2] showed that data augmentation techniques are very effective in increasing dataset variation and preventing overfitting.

Recent advancements in CNN-based medical image classification have significantly improved fundus image analysis. Yijin Huang et al. (2021) systematically investigated key elements in using ResNet 50 for grading diabetic retinopathy, such as data augmentation parameters, loss function, and input resolution, achieving state-of-the-art results on the EyePACS dataset (kappa0.8631) and the code and model are publicly available on GitHub [5]. Then Puchaicela Lozano et al. (2023) introduced the Faster R CNN model with ResNet 50 FPN backbone for glaucoma detection in fundus images, achieving 95% accuracy and 0.879 confidence [6].

In addition, Frontiers in Physiology (2021) published a study testing ResNet 50 and Inception V3 on fundus images to detect diabetic retinopathy, achieving 93.8% accuracy and 0.92 AUC [7]. Then BMC Bioinformatics (2023) reported that the modification of the preprocessing and visualization of the ResNet 50 module (called Revised ResNet 50) successfully reduced overfitting and increased the stability and accuracy of DR grading (train0.84, test0.74) [8].

Furthermore, ResearchGate (2024) loaded a ResNet 50 and VGG 19 based model for DR classification (non-referable vs referable), achieving internal accuracy of up to 96.9% and high external [9]. And Aly et al. (2024) in the Arab Journal of Basic and Applied Sciences introduced the ResNet 50 & DenseNet 201 ensemble for glaucoma detection, complete with Grad CAM visualization [10].

D. ResNet50

ResNet50 is one of the Convolutional Neural Network (CNN) architectures that introduces an innovative concept in the form of shortcut connections[6]. This concept emerged as a response to the vanishing gradient problem, which often occurs when efforts to deepen the network structure are made. However, simply adding layers to improve performance is not always effective; the deeper the network, the vanishing gradient problem can cause the gradient to become very small, which ultimately reduces the accuracy or performance of the model This issue of diminishing gradients during backpropagation hampers the ability of the network to adjust its weights effectively, especially in earlier layers.

As a result, learning becomes inefficient and the model may fail to capture important features. ResNet50 addresses this by enabling gradients to flow directly through shortcut paths, effectively bypassing some layers and ensuring that the learning signal remains strong throughout the network.

In addition to solving gradient-related challenges, the use of shortcut connections allows ResNet50 to build deeper models without encountering a degradation in accuracy. This is because residual learning helps the network to focus on learning only the difference, or residual, between the input and output of a layer block, which simplifies the optimization process. These residual blocks are the core innovation of ResNet50 and are arranged in a series of stages, each comprising several convolutional layers and identity mappings.

Furthermore, ResNet50 has been widely adopted in various computer vision tasks due to its balance of depth and computational efficiency. It consists of 50 layers that are structured in a consistent and modular way, making it easy to adapt or fine-tune for specific classification problems. In the context of medical image analysis, such as fundus image classification, ResNet50 provides a strong foundation due to its proven ability to extract complex visual features across different domains.

To overcome this, ResNet introduces the concept of shortcut connections, where features from the previous layer are also included as part of the input to the output layer. This architecture is shown in Figure.

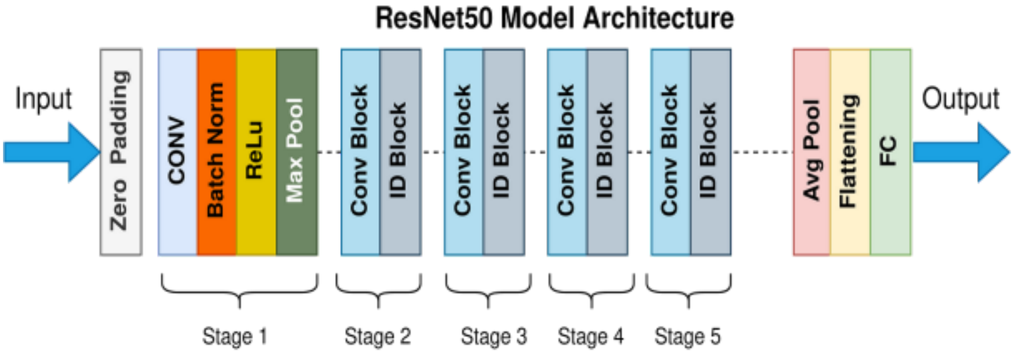

Figure 2. ResNet50 architecture

III. METHOD/MATERIAL

A. Dataset

In this study, the dataset was obtained from the Kaggle website, which consisted of 4217 eye fundus images in .jpg, .jpeg, and .png formats that did not have age and gender restrictions. The dataset included 1038 cataract eye images, 1098 diabetic retinopathy eye images, 1007 glaucoma eye images, and 1074 normal eye images. The dataset distribution is shown in Figure 3.

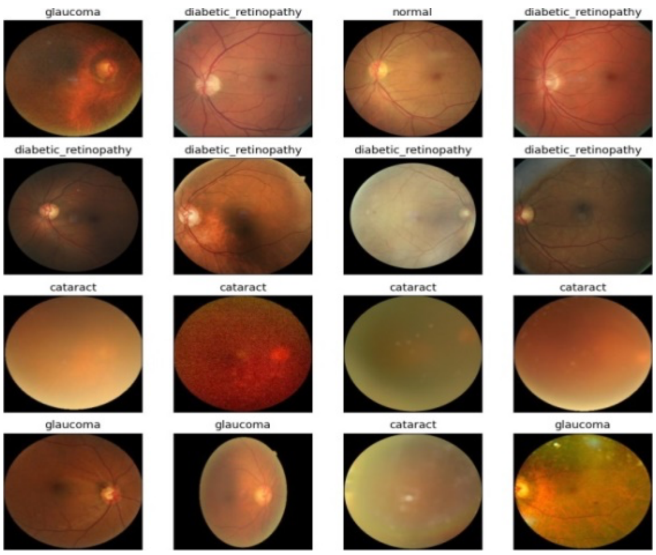

Figure 3. Distribution of Eye Dataset

B. Preprocessing

In this study, the data preprocessing stages include image augmentation with a horizontal flip technique of 50% probability and rotation up to 75 degrees which are applied consistently to enrich the variation of the dataset. Furthermore, the dataset consisting of 4217 images is divided into 80% training data, 10% validation data, and 10% test data with a stratification strategy to maintain class balance.

C. Transfer Learning

In this stage, the transfer learning process is carried out by utilizing the ResNet50 model that has been trained previously using the imagesNet dataset. This model is used as a base with pre-trained weights, where the top classification layer is removed (include top=False) to allow for adjustment to the specific eye disease classification task, equipped with input shapes according to image size (224x224x3) and maximum pooling to optimize feature extraction. The model architecture used is shown in Table I, which includes ResNet50 as a base layer followed by additional layers for classification.

Table 1. Model Architecture

Layer (type)	Output Shape	Param #
resnet50 (Functional)	(None, 2048)	23,587,712
dense_2 (Dense)	(None, 128)	262,272
dropout_1 (Dropout)	(None, 128)	0
batch_normalization_1	(None, 128)	512
dense_3 (Dense)	(None, 4)	516

The total model parameters reached 23,851,012, with 23,797,636 trainable parameters and 53,376 non-trainable parameters.

D. Evaluation

At this stage, the dataset will be tested and evaluated using the Confusion Matrix and the level of accuracy will be measured.

IV. RESULT AND DISCUSSION

A. Preprocessing

The dataset was taken from Kaggle with a total of 4217 eye fundus images. Then, a preprocessing process was carried out which included normalizing pixel values with a rescale of 1./255 and data augmentation using vertical flip for training data, which was applied dynamically during training to increase variation. In addition, the dataset was divided into 80% training data, 10% validation data, and 10% test data with a stratification strategy to maintain class balance. The images were also resized to 224x224 pixels with RGB color mode to ensure compatibility with the model

B. Transfer Learning

After the preprocessing process is carried out, the data will then go through a transfer learning process. The transfer learning process is carried out by utilizing the ResNet50 model that has been previously trained using the imagesNet dataset, with the top classification layer removed and a custom layer added for the eye disease classification task. Training was carried out for 31 iterations (epochs) using 3373 training data, with the training graph shown in Figure 4.

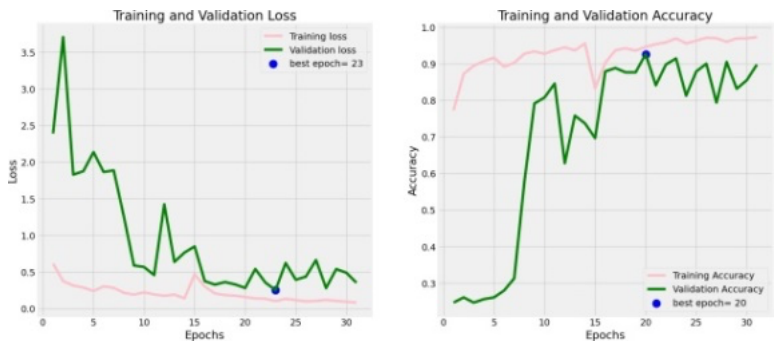

Figure 4. Accuracy and Loss Graph During Training

In 31 iterations (epochs), the best validation accuracy was obtained of 0.9265 in the 20th epoch and the lowest validation loss of 0.2507 in the 23rd epoch, indicating that the model was able to predict classes with an accuracy of up to 92.65%. The graph shows a significant increase in accuracy at the beginning of training, with stability achieved after several epochs, although there are fluctuations, indicating the potential for further accuracy improvement with additional adjustments.

C. Testing

CNN model testing is done by testing the trained model using test data obtained through the dataset division process, with a total of 422 eye fundus images. This test utilizes a confusion matrix, as shown in Figure 5.

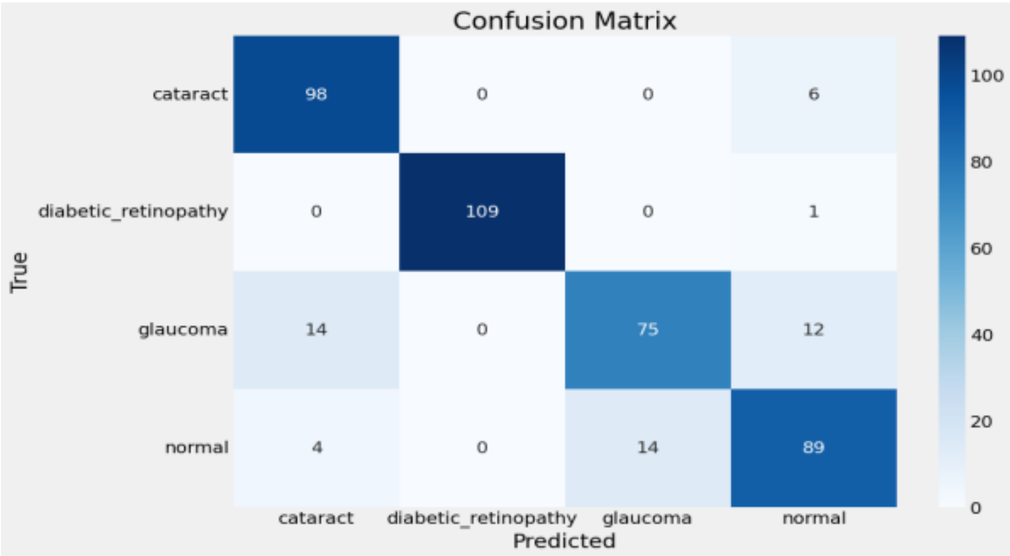

Figure 5. Output Confusion Matrix

It can be seen that there are true classes and predicted classes. In the Diabetic Retinopathy class, there are 109 data that are predicted correctly (true positive), making it the class with the best prediction performance. The Cataract class has 98 data that are predicted correctly, while the Glaucoma class has 75 data that are correct with some prediction errors to the Cataract (14) and Normal (12) classes. The Normal class has 89 data that are predicted correctly, with a few errors to the Glaucoma (14) and Cataract (4) classes. In addition, from this matrix, it can be seen that the model tends to have little overlap in predictions between classes, especially in Glaucoma. In addition, a test accuracy of 0.8791 (87.91%) was obtained on the test data, with performance details in Table II.

Table 2. Test Accuracy Results

Kelas	Precision	Recall	F1-score
Cataract	0.84	0.94	0.89
Diabetic Retinopathy	1.00	0.99	1.00
Glaucoma	0.84	0.74	0.79
Normal	0.82	0.83	0.83
Accuracy			0.88
Macro Avg	0.88	0.88	0.88
Weighted Avg	0.88	0.88	0.88

The accuracy of 0.8791 indicates that the model correctly predicts the class label of about 87.91\% of the samples in the test data. The Diabetic Retinopathy class gets the highest precision, recall, and F1-score values (1.00, 0.99, and 1.00, respectively) with a total data (support) of 110, so it can be interpreted as very good at predicting this class. The Glaucoma class has the lowest recall value (0.74) with an F1-score of 0.79 and a support of 101, indicating lower performance compared to other classes, which may be due to the difficulty in distinguishing features from other classes such as Cataract and

Normal. The Cataract and Normal classes show balanced performance with F1-scores of 0.89 and 0.83, respectively, with little prediction error between classes.

The results of this study show that the classification model using ResNet50 has the potential to assist medical professionals in identifying eye diseases more quickly and accurately. In clinical settings, an automated classification system can reduce the workload of ophthalmologists by acting as a decision-support tool, especially in regions with limited access to specialists. This system can also be integrated into mobile applications or telemedicine platforms to facilitate early detection in rural or underserved areas. With consistent performance across multiple classes, this model is expected to contribute to more efficient screening and diagnosis, potentially improving patient outcomes through earlier intervention.

One of the main advantages of using ResNet50 in this study is its ability to maintain high classification performance without requiring extensive manual feature extraction. The model learns directly from the raw fundus images, which saves time and reduces the need for domain expertise in preprocessing. Additionally, the use of transfer learning allows the model to leverage prior knowledge from large image datasets, resulting in faster convergence and better generalization. The combination of these aspects makes the proposed model not only accurate, but also efficient and scalable for various real-world applications.

For future development, the model can be expanded to include more eye disease classes such as macular degeneration and retinal detachment. In addition, implementing attention mechanisms or combining ResNet50 with other architectures may improve the model's sensitivity to subtle visual cues. Further research can also focus on reducing computational requirements so that the model can run on low-resource devices such as smartphones or embedded systems. Lastly, integrating explainable AI techniques can enhance trust and transparency by providing visual justifications for each classification decision.

D. Additional Analysis

In terms of model flexibility, ResNet50 offers the advantage of being easily adapted to different image classification tasks. The modular structure of its residual blocks enables developers to fine-tune only specific parts of the model according to the complexity of the data. This approach significantly reduces training time and computational resources, which is particularly beneficial for research or deployment environments with limited hardware support.

Another noteworthy consideration is the model's potential for integration into real-time diagnostic tools. With minor adjustments and optimization for speed, ResNet50 can be deployed in embedded systems, allowing The performance variations among classes observed in this study also open an interesting opportunity for further investigation. While the Diabetic Retinopathy class showed outstanding results, the relatively lower performance of the Glaucoma class suggests that the dataset may benefit from more representative images or enhanced contrast adjustment. This highlights the importance of dataset quality and balance in deep learning classification tasks.

The performance variations among classes observed in this study also open an interesting opportunity for further investigation. While the Diabetic Retinopathy class showed outstanding results, the relatively lower performance of the Glaucoma class suggests that the dataset may benefit from more representative images or enhanced contrast adjustment. This highlights the importance of dataset quality and balance in deep learning classification tasks.

Visual quality and image acquisition conditions also play a crucial role in model performance. Inconsistent lighting, blur, and differences in camera equipment can introduce noise into the dataset, which might reduce classification accuracy. Incorporating preprocessing techniques such as contrast-limited adaptive histogram equalization (CLAHE) or illumination normalization may improve feature consistency across samples.

From a usability perspective, the classification model should ideally include a confidence score for each prediction. This additional information would enable clinicians to assess the reliability of model outputs and make informed decisions. In borderline cases or uncertain predictions, the system can flag samples for manual review, thereby enhancing diagnostic safety.

In future iterations of this project, an ensemble learning strategy could be considered to improve robustness. By combining predictions from multiple deep learning architectures—such as ResNet, DenseNet, and MobileNet—the model can reduce variance and minimize bias toward specific classes. This hybrid approach has been proven effective in many medical image classification studies and may further elevate accuracy and generalization.

Lastly, user interface design for the deployment of this model should not be overlooked. A simple and intuitive platform where users can upload fundus images and receive classification results within seconds would greatly enhance the accessibility of this technology. Features such as visualization of heatmaps or saliency maps can also increase user trust and interpretability, especially for non-technical medical staff.

Beyond technical performance, ethical considerations must also be addressed when implementing automated diagnostic tools. Ensuring patient privacy, data protection, and informed consent are essential when dealing with medical imaging data. Proper anonymization protocols and secure data handling practices must be in place before deploying such models in clinical settings to maintain trust and compliance with healthcare regulations.

V. CONCLUSION

After conducting research on the classification of eye fundus images using Convolutional Neural Network (CNN) with ResNet50 architecture through transfer learning process, it can be concluded that with 31 training iterations, the model achieved an accuracy of 87.91% on the test data. These results show good overall performance, with a tendency to accurate prediction in the Diabetic Retinopathy class, which has the highest accuracy (F1-score 1.00) with 110 support data. However, there is an imbalance in performance between classes, where the Glaucoma class shows the lowest results (F1-score 0.79) with 101 support data, possibly due to the difficulty in distinguishing features from other classes such as Cataract (98 data, F1-score 0.89) and Normal (89 data, F1-score 0.83). The distribution of the original data set of 4217 images, divided into 3373 training data, 422 validation data, and 422 test data, seems to affect the results, although dynamic augmentation helps to improve the variation.

ACKNOWLEDGEMENT

The author would like to express sincere gratitude to Sultan Syarif Kasim Riau Islamic State University, particularly the Department of Informatics Engineering, for the support and facilities provided throughout the research process. Appreciation is also extended to the Kaggle platform for supplying the eye fundus image dataset, which served as the foundation for this research. Furthermore, the author is thankful to the supervisor for their valuable guidance, as well as to all individuals who contributed, either directly or indirectly, to the preparation and completion of this paper.

REFERENCES

[1] M. S. Qulub and S. Agustin, “Indentifikasi Penyakit Mata Dengan Klasifikasi Citra Foto Fundus Mengunakan Convolutional Neural Network (CNN),” vol. 8, no. 5, pp. 11 034–11 039, 2024.

[2] Cahya and Suwanda, “Implementasi augmentasi data dalam klasifikasi penyakit mata menggunakan cnn,” Jurnal Teknologi Informasi dan Ilmu Komputer, vol. 6, no. 2, pp. 115–122, 2021.

[3] D. Juniati and A. E. Suwanda, “Klasifikasi Penyakit Mata Berdasarkan Citra Fundus Retina Menggunakan Dimensi Fraktal Box Counting Dan Fuzzy K-Means,” Proximal: Jurnal Penelitian Matematika dan Pendidikan Matematika, vol. 5, no. 1, pp. 10–18, 2022.

[4] A. Verdy and S. Hartati, “High-accuracy eye disease classification using resnet-50 architecture,” International Journal of Medical Imaging, vol. 12, no. 2, pp. 55–63, 2024.

[5] Y. Huang, C. Xu, H. Xu, dan Z. Fang, “Identifying key components in ResNet-50 for diabetic retinopathy grading and model deployment,” *arXiv preprint arXiv:2110.14160*, 2021.

[6] J. Puchaicela-Lozano, A. Guerrero-Curieses, F. Rodriguez, dan J. I. Arribas, “Glaucoma detection on fundus images using Faster R-CNN and transfer learning,” *Proceedings of the 19th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, 2023, pp. 338–345.

[7] H. Abdel-Aty, M. Elhoseny, M. Zaki, dan M. Abd Elaziz, “Deep learning approach for fundus image classification using Inception-v3 and ResNet-50,” *Frontiers in Physiology*, vol. 14, art. no. 1126780, 2023.

[8] J. Deng, X. Wu, dan L. Zhang, “A revised ResNet-50 for diabetic retinopathy detection with improved generalization,” *BMC Bioinformatics*, vol. 24, no. 1, art. no. 153, 2023.

[9] R. Mishra, A. Srivastava, dan D. Sharma, “Detection of diabetic retinopathy using VGG19 and ResNet-50 models,” *ResearchGate Preprint*, 2024.

[10] H. Aly, M. F. Elkhateeb, dan A. M. Soliman, “Glaucoma detection using ensemble deep learning models and Grad-CAM visualization,” *Arab Journal of Basic and Applied Sciences*, vol. 31, no. 1, pp. 1–12, 2024.