Multiclass Eye Disease Classification Using Transfer Learning Approach

Alvin Ming Siang Choo[1], Stephanie Chua[1,*], Lik Thai Lim[2], Dayang Nurfatimah Awang Iskandar[1], Mohd Hanafi Ahmad Hijazi[3], and Puteri Nor Ellyza Nohuddin[4]

[1]Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, Kota Samarahan, Sarawak, Malaysia
[2]Faculty of Medicine and Health Sciences, Universiti Malaysia Sarawak, Kota Samarahan, Sarawak, Malaysia
[3]Faculty of Computing and Informatics, Universiti Malaysia Sabah, Kota Kinabalu, Sabah, Malaysia
[4]Higher Colleges of Technology, Sharjah, United Arab Emirates
Email: 23020166@siswa.unimas.my (A.M.S.C.); chlstephanie@unimas.my (S.C.); ltlim@unimas.my (L.T.L.); dnfaiz@unimas.my (D.N.A.I.); hanafi@ums.edu.my (M.H.A.H.); pnohuddin@hct.ac.ae (P.N.E.N.)
[*]Corresponding author

Manuscript received February 28, 2025; revised April 22, 2025; accepted June 13, 2025; published October 20, 2025

***Abstract*—Ophthalmologists commonly use retinal fundus images for diagnosis. Recently, automation of this process using deep learning has gained significant attention. Multiclass classification, which distinguishes among multiple eye diseases, is more representative of actual clinical settings, however, it presents challenges such as limited availability of annotated datasets, class imbalance, overlapping clinical features across various eye diseases and disease heterogeneity. This study develops deep learning models for multiclass classification of three major eye diseases—cataracts, diabetic retinopathy, and glaucoma—alongside normal cases. A larger and more diverse dataset was obtained by combining multiple publicly available, well-annotated datasets. Four deep learning models: VGG16, Inception-v3, ResNet50 and EfficientNet-B0, were deployed using a transfer learning approach. These models achieved test accuracies ranging from 74.29% to 78.79%, with ResNet50 performing the best, achieving an accuracy of 78.79%, precision of 80.04%, recall of 78.79%, and an F1-score of 78.76%. The results demonstrate the effectiveness of transfer learning for multiclass classification of eye diseases. Notably, the models were trained and evaluated on a heterogeneous dataset that simulates real-world variability in image acquisition, highlighting their generalization capabilities and robustness to inconsistency. The study provides valuable insights about the performance of pre-trained deep learning models under realistic conditions, supporting their potential as assistive diagnostic tools in actual clinical scenarios.**

***Keywords*—multiclass classification, eye diseases, retinal fundus images, transfer learning, Convolutional Neural Networks (CNNs), pre-trained deep learning models**

I. Introduction

Visual impairment occurs when certain eye conditions affect the visual system and its functioning [1]. It covers a range of conditions, from mild visual disturbances to complete blindness. According to Ref. [2], there are at least 2.2 billion people worldwide suffering from some form of vision impairment, and in nearly half of these cases, the problem could have been prevented. The most prevalent eye diseases that cause vision impairment and blindness include age-related macular degeneration, cataracts, diabetic retinopathy, and glaucoma [2]. Age-related macular degeneration results from the deterioration of macula due to aging and is prevalent among individuals aged 50 and older [2]. Cataracts are caused by clouding that accumulates in the lens of the eye and are often linked to the clustering of proteins in the eye [3]. While cataracts can develop at any age, they are more common in the elderly [2]. Diabetic retinopathy arises from chronically high and uncontrolled blood sugar levels, which damage the retinal blood vessels. Glaucoma, on the other hand, is caused by abnormally high intraocular pressure, which progressively damages the optic nerve [1].

Early diagnosis and timely treatment are important in halting disease progression and preventing blindness. However, eye disease detection is a challenging task that typically requires years of clinical experience. Retinal fundus images, captured by fundus cameras or ophthalmoscopes, are one of the key tools that allow the ophthalmologists to detect signs of various eye conditions [3, 4]. These images depict the retina appearance including blood vessels, macula, fovea and optic disc [3]. The manual analysis of these images is time-consuming, laborious, and arbitrary, resulting in low repeatability [5]. Moreover, ophthalmologists sometimes face difficulties in making accurate diagnoses when the quality of retinal fundus images is compromised [5]. Several challenges complicate eye disease detection from retinal fundus images, include the complex and subtle symptom of eye diseases, overlapping clinical features especially in the early stages, and the vast number of images that need to be analyzed [5]. Such factors can lead to misdiagnosis, delays, and inconsistent evaluations.

These challenges have motivated the deployment of deep learning-based approaches for automated eye disease classification. In real-world settings, ophthalmologists often diagnose multiple types of eye diseases; thus, multiclass eye diseases classification offers a more realistic representation of actual clinical practice. A significant limitation in this area is the lack of large, well-annotated dataset [5]. Most publicly available datasets consist of only a few hundred to a few thousand retinal fundus images, which constrain model performance and generalizability, especially when training from scratch.

This study proposes a multiclass classification framework for detecting three major eye diseases: cataracts, diabetic retinopathy, and glaucoma, using retinal fundus images that prioritize both diagnostic performance and real-world applicability. The contributions of this study are as follows:

- We investigate the effectiveness of transfer learning in developing robust and generalizable models for retinal fundus image classification, particularly in the context of multiclass eye disease classification.
- We construct a larger and more diverse dataset by combining multiple publicly available datasets to

address the limitations of small or homogeneous datasets. This approach better reflects real-world variations and inconsistencies in medical imaging.

- We evaluate the diagnostic performance and robustness of four state-of-the-art pre-trained models: VGG16, Inception-v3, ResNet50, EfficientNet-B0, ensuring balanced evaluation on heterogeneous dataset to assess their generalizability.

II. Literature Review

Recent years have witnessed growing interest in employing deep learning, particularly Convolutional Neural Networks (CNNs), for automated classification of eye diseases using retinal fundus image [5]. Much of existing research has focused on major eye diseases, such as age-related macular degeneration, cataracts, diabetic retinopathy, and glaucoma using color retinal fundus images [6–10], Optical Coherence Tomography (OCT) [11] below or multimodal imaging approaches [12, 13]. The primary goal of these studies is to reduce reliance on manual assessment by enabling accurate automated multiclass classification

Several studies have explored various CNN architectures for classifying individual eye diseases. Singh *et al.* [10] below deployed VGG16 models to differentiate between cataracts and normal cases, reporting a remarkable accuracy of 96.10%. Hemelings *et al.* [14] employed a pre-trained ResNet50 model, achieving an accuracy of 94% in classifying glaucoma from retinal fundus images. Shoukat *et al.* [15], on the other hand, applied a transfer learning approach for glaucoma classification using three pre-trained models: VGG19, ResNet50, and EfficientNet-B7. These models were trained and evaluated on three different datasets (RIM-ONE, G1020, REFUGE). Among them, EfficientNet-B7 demonstrated the best overall performance, achieving accuracies of 97% on RIM-ONE, 99.2% on G1020, and 99% on REFUGE.

Thanki [16] introduced a dual learning-based approach that integrates deep neural networks with traditional machine learning classifiers for the classification of glaucomatous retinal fundus images. The proposed system utilizes SqueezeNet to extract deep features from color fundus images, which are then classified using six different machine learning algorithms: k-nearest neighbor, decision tree, Support Vector Machine (SVM), random forest, naive bayes, and logistic regression.

Kallel and Echtioui [17] evaluated four pre-trained convolutional neural network models—VGG16, VGG19, InceptionV3, and DenseNet169—on the APTOS2019 dataset, for classification of diabetic retinopathy severity. Among the models tested, InceptionV3 achieved the highest classification accuracy of 96.88%. Furthermore, Sarki *et al.* [18] proposed a hybrid framework combining traditional image processing methods with both pre-trained deep learning model and custom CNN model. Their experiments, conducted on Messidor, Messidor-2, and DRISHTI-GS datasets, showed that the custom CNN trained on pre-processed images achieved superior performance.

In the context of multiclass classification, Chea and Nam [19] used ResNet (ResNet50, ResNet101, ResNet152) and VGG models (VGG16, VGG19) for classification of three prevalent eye diseases: age-related macular degeneration, diabetic retinopathy, glaucoma. Guergueb and Akhloufi [20] evaluated EfficientNet variants (EfficientNet-B5, EfficientNet-B6, EfficientNet-B7) and DenseNet models (DenseNet121, DenseNet169, DenseNet201) for the classification of eight eye disease categories. Among these, EfficientNet-B7 demonstrated the highest performance, achieving an Area Under the Curve (AUC) of 96.04%.

Babaqi *et al.* [21] demonstrated improved multiclass accuracy by comparing a custom CNN model (84% accuracy) with a pre-trained EfficientNet model (94% accuracy) to distinguish between normal eyes and those affected by cataracts, diabetic retinopathy, or glaucoma. Additionally, Toki *et al.* [22] introduced a CNN-based architecture named RetinalNet-500 and benchmarked its performance against several pre-trained models, including Inception-v3, MobileNetV2, and Xception, with accuracy scores ranging from 95.15% to 97.30%. Cui *et al.* [23] assessed the performance of ResNet50, VGG19, EfficientNet-B0 and DenseNet on a relatively small dataset consisting of normal, cataracts, diabetic retinopathy, and glaucoma fundus images. Among these models, ResNet50 achieved the highest test accuracy, reaching 92.91%.

Tashkandi [24] extended classification to multiple eye diseases, including diabetic retinopathy, glaucoma, cataracts, high myopia, and age-related macular degeneration—using retinal images. Several models were evaluated, including traditional machine learning algorithms (SVM and random forest) and deep learning architectures (VGG16, MobileNetV1, and a hybrid CNN-RNN model). Among these, MobileNetV1 achieved the highest accuracy of 98%.

Muntaqim *et al.* [25] proposed a multi-stage deep learning framework addressing limitations in feature extraction, computational efficiency, and disease coverage. Their model comprises three stages: initial fine-grained features extraction via convolutional layers; features enhancement through two parallel convolutional and identity blocks; and features fusion and classification using Long Short-Term Memory (LSTM) and dense layers. Evaluated on three benchmark datasets (OCT2017, Dataset-101, and Retinal OCT C8), the model achieved high accuracy—97.52%, 92.97%, 94.81% for multiclass classification, and 99.33% for binary classification on OCT2017—surpassing several state-of-the-art methods.

Most recently, multimodal approaches have emerged as a growing trend, offering improved diagnostic performance by integrating complementary information from multiple data sources. Wardhani *et al.* [26] proposed an early fusion of deep learning model combining fundus retinal images, OCT, and electronic health records for diabetic retinopathy detection. Their approach, which employed Local Binary Pattern (LBP) for feature extraction and a LSTM network for classification, demonstrated superior AUC of 0.99, outperforming unimodal approaches.

Kang *et al.* [27] proposed a multimodal deep learning framework for identifying retinal vascular diseases that require treatment, utilizing a combination of retinal fundus images, OCT, and Fluorescein/Indocyanine Green Angiography (FA/ICGA). It achieved impressive diagnostic performance, with AUC values of 0.996 for myopic choroidal neovascularization, 0.995 for diabetic macular edema, 0.990 for neovascular age-related macular degeneration, 0.959 for branch retinal vein occlusion, and 0.988 for central retinal vein occlusion.

El-Ateif and Idri [12] compared seven state-of-the-art deep learning models (VGG19, DenseNet121, InceptionV3, InceptionResNetV2, Xception, ResNet50V2, and MobileNetV2) for diagnosing diabetic eye diseases (diabetic retinopathy, age-related macular degeneration, glaucoma, cardiovascular disease) across mono-modality and late fusion multimodal imaging approaches. While DenseNet121 and ResNet50V2 achieved high accuracy in mono-modality (99.57% and 99.51%, respectively), multimodal late fusion of retinal fundus photography, OCT, and histology significantly improved diagnostic performance, with ResNet50V2 reaching 100% accuracy.

While the current research direction demonstrates promising potential for deep learning in retinal fundus image classification, many existing studies focus on limited-class classification and are based on curated datasets with limited variability [6, 20, 23]. In real-world clinical settings, retinal fundus images may vary in terms of resolution, illumination and quality, especially when captured from diverse imaging devices [4, 5]. This variability can significantly affect model performance, highlighting the need to assess the robustness and generalization capability of deep learning models across heterogeneous dataset. There remains a need for evaluating the robustness and generalization ability of pre-trained models across heterogeneous dataset with varied image conditions. Moreover, while multimodal approaches, such as the integration of fundus images, OCT and electronic health records, have shown notable improvements in diagnostic performance, they often rely on access to advanced imaging technologies. In resource-constrained or underserved regions, access to OCT or comprehensive electronic health records systems may be limited, making the widespread use of such approaches challenging. Therefore, there is a need to develop cost-effective and scalable solutions that can perform reliability using single-modality data, particularly retinal fundus images, to ensure broader clinical applicability in later time.

III. Materials and Methods

This research comprises several sequential stages aimed at developing multiclass classification models for the detection of eye diseases using retinal fundus images. The key stages include data collection, data pre-processing, model training, and model evaluation. A visual summary of the entire research workflow is presented in Fig. 1, which illustrates the step-by-step process adopted in this study. Each stage is described in detail in the following subsections.

A. Data Collection

This study utilizes retinal fundus images representing four classes: cataracts, diabetic retinopathy, glaucoma, and normal eyes. Retinal fundus imaging offers a non-invasive, cost effective, and widely accessible method for capturing detailed views of the retinal, making it suitable for large-scale screening and early detection of eye diseases. Acquiring such medical imaging data can be time-consuming and laborious as it is necessary to rely on domain experts to accurately annotate the retinal fundus images.

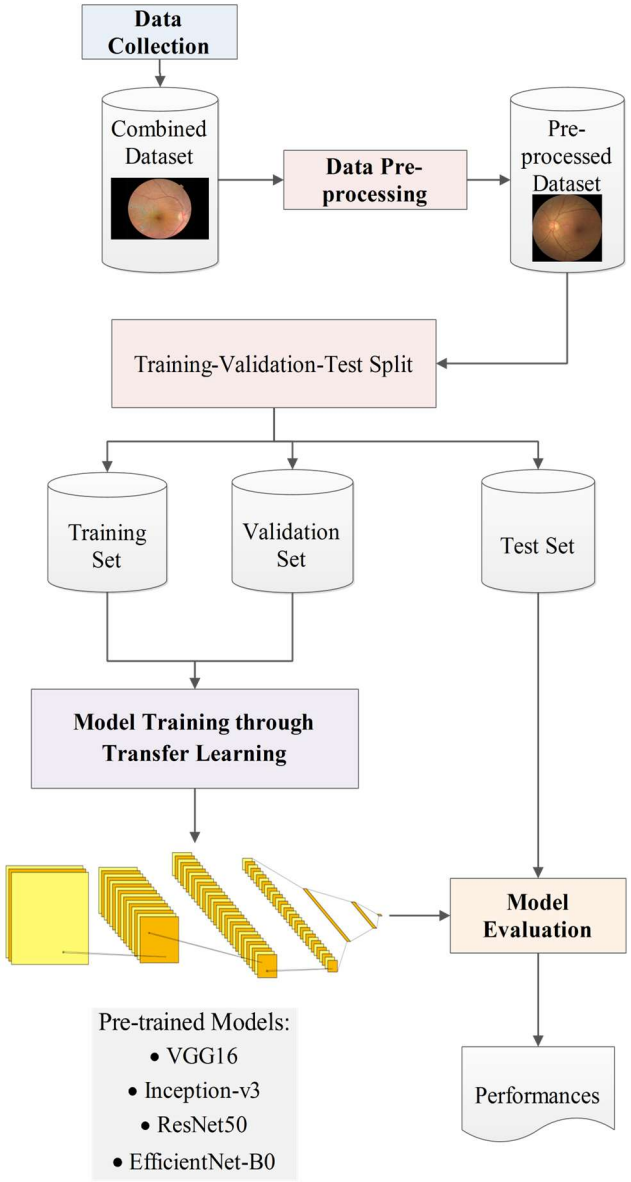


Fig. 1. Research flow diagram for transfer learning-based models' development for multiclass eye diseases classification.

To address these challenges, we used publicly available, well-annotated datasets, which were sourced from online databases and datasets provided by other researchers in previous studies. The publicly available datasets used in this study are as follows:

1) Eye Disease Retinal Images [28]
2) Glaucoma Fundus Imaging Dataset [29]
3) Ocular Disease Intelligence Recognition [30]
4) Retinal Fundus Multi-disease Image Dataset (RFMiD) [31]

All images were reviewed across sources before pre-processing and used for subsequent process. The distribution of the retinal fundus images among four classes in the combined dataset is summarized in Table 1. Fig. 2a–d showcase sample retinal fundus images for each respective class in the combined dataset. By combining multiple publicly available datasets, a larger and comprehensive dataset was created to ensure volume and diversity, better reflecting the variation of images taken in actual clinical settings.

Table 1. Distribution of dataset

Class	Number of Images
Cataracts	1305
Diabetic retinopathy	2875
Glaucoma	1636
Normal	3251
Total	**9067**

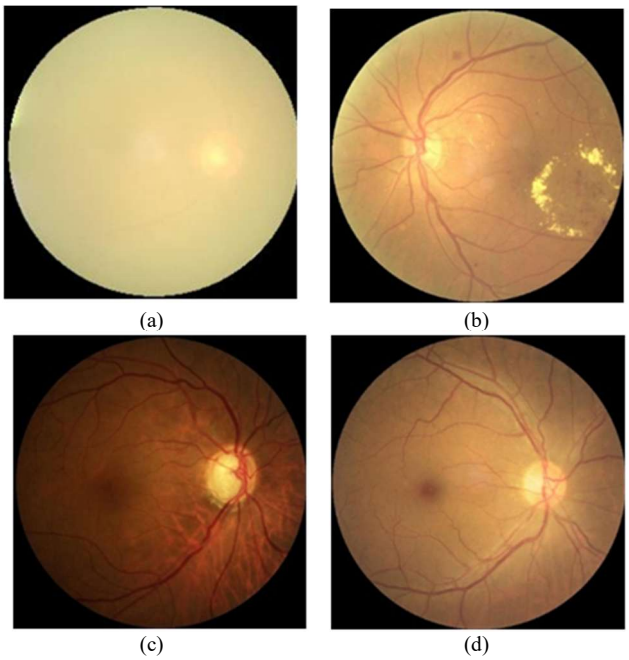


Fig. 2. Sample retinal fundus images representing each class in the combined dataset: (a) Cataract, (b) Diabetic Retinopathy, (c) Glaucoma, and (d) Normal.

B. Data Pre-processing

While the approach of combining several publicly available datasets to obtain a larger dataset effectively increased the dataset size, it also introduced inconsistencies in image quality and resolution due to differences in data sources. Rather than filtering out these inconsistencies, this study aims to evaluate the performance of pre-trained deep learning models in the presence of such variability, reflecting real-world scenarios. This approach enhanced the practical relevance of the findings, particularly for deployment of the models in diverse clinical settings.

To minimize the impact of these variations, all fundus images underwent a series of pre-processing steps. First, we cropped all the retinal fundus images into square shapes, with the purpose of removing extraneous noise, particularly the black background, and preserving only the relevant fundus region. This step helped to improve the focus of the models on meaningful anatomical structures. The cropped images were then resized to a fixed resolution of 224×224 pixels to match the input size requirements of the pre-trained models used in this study.

We randomly split the pre-processed dataset into three subsets: a training set, a validation set, and a test set, following an 8:1:1 split ratio. The random split ensured that data from all classes were proportionally represented in each subset. The training set and validation set were used during model training, whereas the test set was used to evaluate the models' performance.

C. Model Training

This research presented an image classification task involving retinal fundus images, and pre-trained deep learning models based on CNN architectures, known for their high performance in computer vision tasks, were selected. Given the limited size of the dataset, transfer learning was employed to leverage the knowledge gained from large-scale training on large amounts of data such as ImageNet. These models, having learned biases and weights during the initial training process, can be employed to adapt to new classification tasks without the need to train from scratch. In this study, four pre-trained deep learning models: VGG16, Inception-v3, ResNet-50 and EfficientNet-B0 were implemented as they performed fairly well using smaller datasets and binary classification in the context of medical image classification tasks [6, 9, 19, 23]. These models were selected as they are well-established, widely used and offer a good balance between computational efficiency and performance, making them well-suited for real-world clinical applications using diverse data sources. A brief description of each model is provided in the following subsections.

1) VGG16

The Visual Geometry Group (VGG) architecture was developed by [32] and won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2014. The authors gradually increased the depth of the network by adding more convolution layers. This was made feasible due to the use of very small convolution filters (3×3) in every layer, along with adjustments to other parameters of the architecture. Among VGG variants, VGG16, which consists of 16 layers, gave a better performance [32].

2) Inception-v3

The first Inception architecture was introduced by Szegedy *et al.* [33]. GoogLeNet, the initial iteration of Inception architecture, was designed to solve the problem of limited computational resources for deeper neural networks and to increase efficiency by reducing the number of parameters without sacrificing accuracy. Inception-v3 is a variant of Inception-v2, which added batch normalization in the auxiliary classifiers. This version not only normalized the convolutions but also applied normalization to the fully connected layer of the auxiliary classifier. Inception-v3 outperformed prior architectures with lower computational cost [34].

3) ResNet50

The Residual Network (ResNet), introduced by Tan *et al.* [35], was a deep CNN architecture designed to address the problem of vanishing and exploding gradients in deep neural networks. It employs a technique known as skip connections, allowing the network to learn residual functions. The first ResNet architecture, ResNet34, converted a plain network into its residual network counterpart by inserting skip connections. The authors later developed a larger architecture using a stacked 3 layers instead of the 2 layers used in ResNet34. Therefore, each 2-layer block in ResNet34 was replaced by a 3-layer bottleneck block, forming the ResNet50 architecture [35].

4) EfficientNet-B0

EfficientNet was developed by [36] with the introduction of a new scaling method called compound scaling. They identified that better performance could be achieved by

carefully balancing the depth, width, and resolution of the network. EfficientNet-B0 was the first model to implement compound scaling. The EfficientNet family of models outperformed earlier CNNs in terms of accuracy and efficiency [36].

These four pre-trained deep learning models were implemented using Python and the Keras Application Programming Interface (API) with TensorFlow as the backend. All models were initialized with weights pre-trained on the ImageNet dataset and the top layers were fine-tuned on the combined dataset. The models were trained for 10 epochs using Adam optimizer with a learning rate of 0.001 and a batch size of 64. The categorical cross entropy was used as the loss function for this multiclass classification task. The training was conducted under the same configuration to ensure a consistent and fair comparison among four models and facilitated subsequent performance evaluation.

D. Model Evaluation

The performances of the trained models were evaluated using several evaluation metrics, including accuracy, precision, recall, and F1-score. These metrics were derived from the values of True Positives (TP), True Negatives (TN), False Positives (FP), and false Negatives (FN).

1) Accuracy measures the overall correctness of model's predictions, using Eq. (1).

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

2) Precision quantifies the model's ability to identify positive instances out of the total predicted positives correctly, using Eq. (2).

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

3) Recall measures the model's ability to classify positive instances out of the total actual positives correctly, as shown in Eq. (3).

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

4) F1-score calculates the harmonic mean between precision and recall using Eq. (4)

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

Additionally, confusion matrices were generated for each model to visualize class-specific performance and identify misclassification patterns.

IV. Result and Discussion

This section presents and analyzes the results of the four pre-trained deep learning models on multiclass classification of eye diseases. Before training on the combined dataset, the models were first evaluated on the test set to assess their baseline performance. Subsequently, the models were trained using the combined dataset and re-evaluated using the same test set. The performance metrics of the pre-trained models before and after training on combined dataset are presented in Table 2. The models achieved relatively low test accuracy, ranging from 13.41% for ResNet50 to 29.78% for Inception-v3, indicating a significant domain gap between ImageNet and combined dataset. After training on combined dataset, the results demonstrate a clear improvement in test accuracy for all four models, with ResNet50 achieving the highest value at 78.79%, followed by VGG (76.26%), EfficientNet-B0 (75.05%), and Inception-v3 (74.29%). The training and validation accuracy and loss plots are shown in Fig. 3. The red line in each figure represents the training accuracy and loss of each model while green line represents the validation accuracy and loss of the models.

Table 2. Comparison of the pre-trained models before and after training using combined dataset

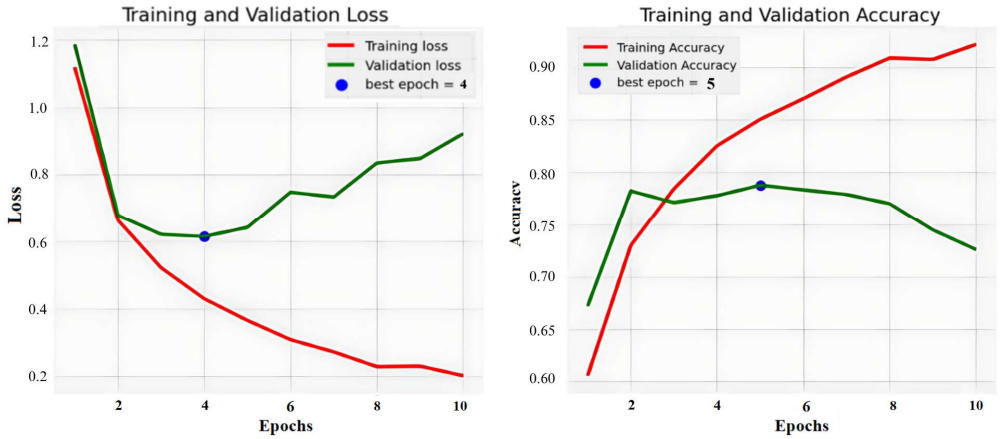
Results / Model	Before Training with Combined Dataset		After Training with Combined Dataset	
	Test Accuracy (%)	Test Loss	Test Accuracy (%)	Test Loss
VGG16	29.56	3.75	76.26	0.80
Inception-v3	29.78	1.41	74.29	0.74
ResNet50	13.41	2.91	78.79	0.76
EfficientNet-B0	23.85	1.48	75.05	0.57

According to the performance of each model on the test set shown in Table 3, all models achieved classification accuracies exceeding 70%, indicating the effectiveness of transfer learning using ImageNet pre-trained weights in the context of eye disease classification. ResNet50 demonstrated the best overall performance among the evaluated models, achieving the highest test accuracy of 78.79%, a precision of 80.04%, recall of 78.79%, and F1-score of 78.76%. These results suggest that ResNet50 was not only accurate in its predictions but also balanced in its sensitivity and specificity across all classes. Following ResNet50, VGG16 achieved a test accuracy of 76.26%, with a precision of 78.67%, recall of 76.26%, and F1-score of 76.44%. EfficientNet-B0 and Inception-v3 also performed competitively, with accuracies of 75.05% and 74.29%, respectively. These results highlight the ability of deep CNN architecture to generalize well to complex medical image classification tasks.

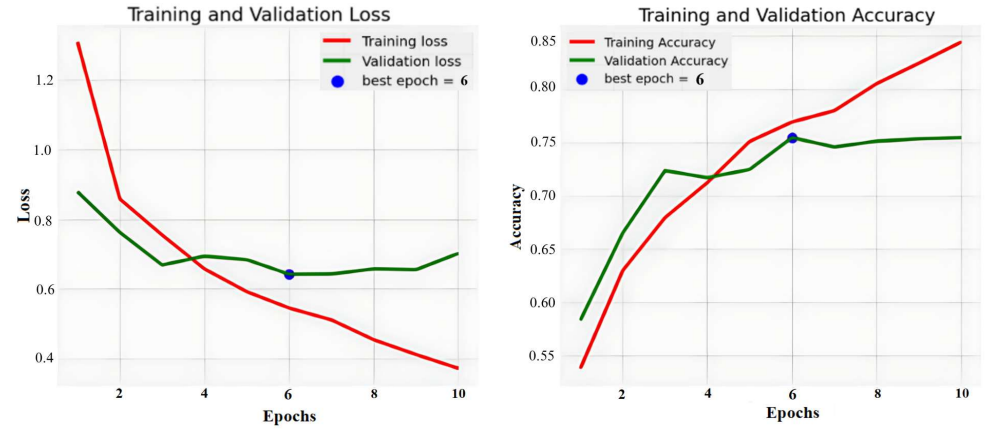
Table 3. Performances of pre-trained deep learning models

Pre-trained Model	Accuracy (%)	Recall (%)	Precision (%)	F1-score (%)
VGG16	76.26	76.26	78.67	76.44
Inception-v3	74.29	74.29	74.48	74.37
ResNet50	**78.79**	**78.79**	**80.04**	**78.76**
EfficientNet-B0	75.05	75.05	78.08	74.83

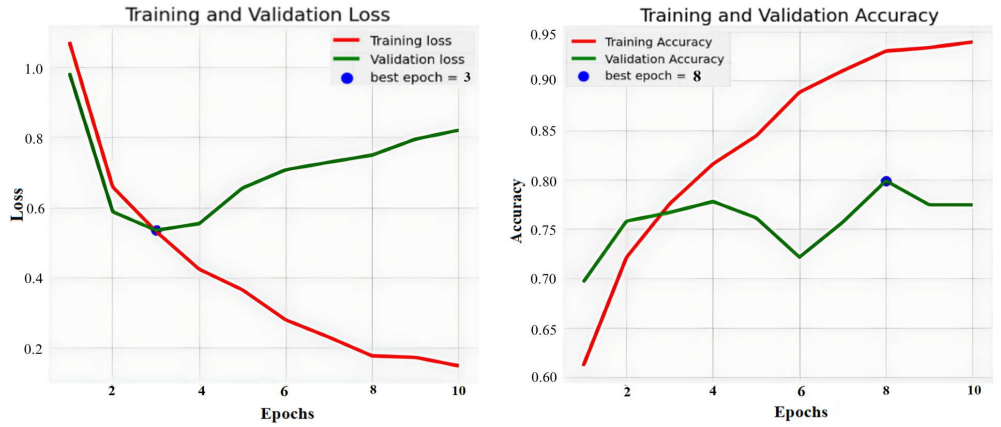
Fig. 4 shows the confusion matrix for each of the pre-trained models in predicting the classes of retinal fundus images in test set. All the four pre-trained models exhibited relatively low misclassification rates for cataracts and glaucoma classes. However, a notable number of misclassifications occurred between diabetic retinopathy and normal retinal fundus images. As observed in the confusion matrices, the models tended to misclassify diabetic retinopathy images as normal, which indicates overlapping features or subtle abnormalities that were not adequately captured during training. Specifically, VGG16, Inception-v3, ResNet50, and EfficientNet-B0 incorrectly classified 30, 62, 81 and 113 out of 288 diabetic retinopathy fundus images as normal, respectively. This trend suggests that more advanced feature augmentation may be required to improve the classification performance for diabetic retinopathy.



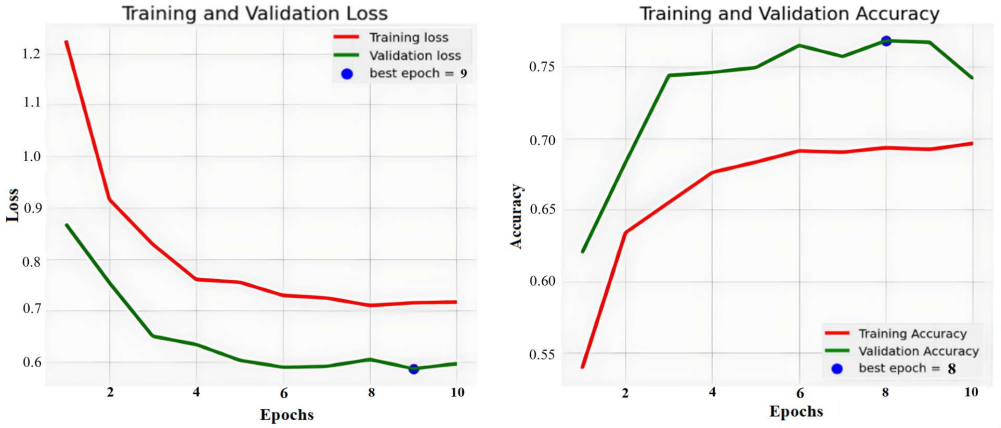
(a)



(b)



(c)



(d)

Fig. 3. Training and validation accuracy and loss plots for four models: (a) VGG16, (b) Inception-v3, (c) ResNet50, and (d) EfficientNet-B0.

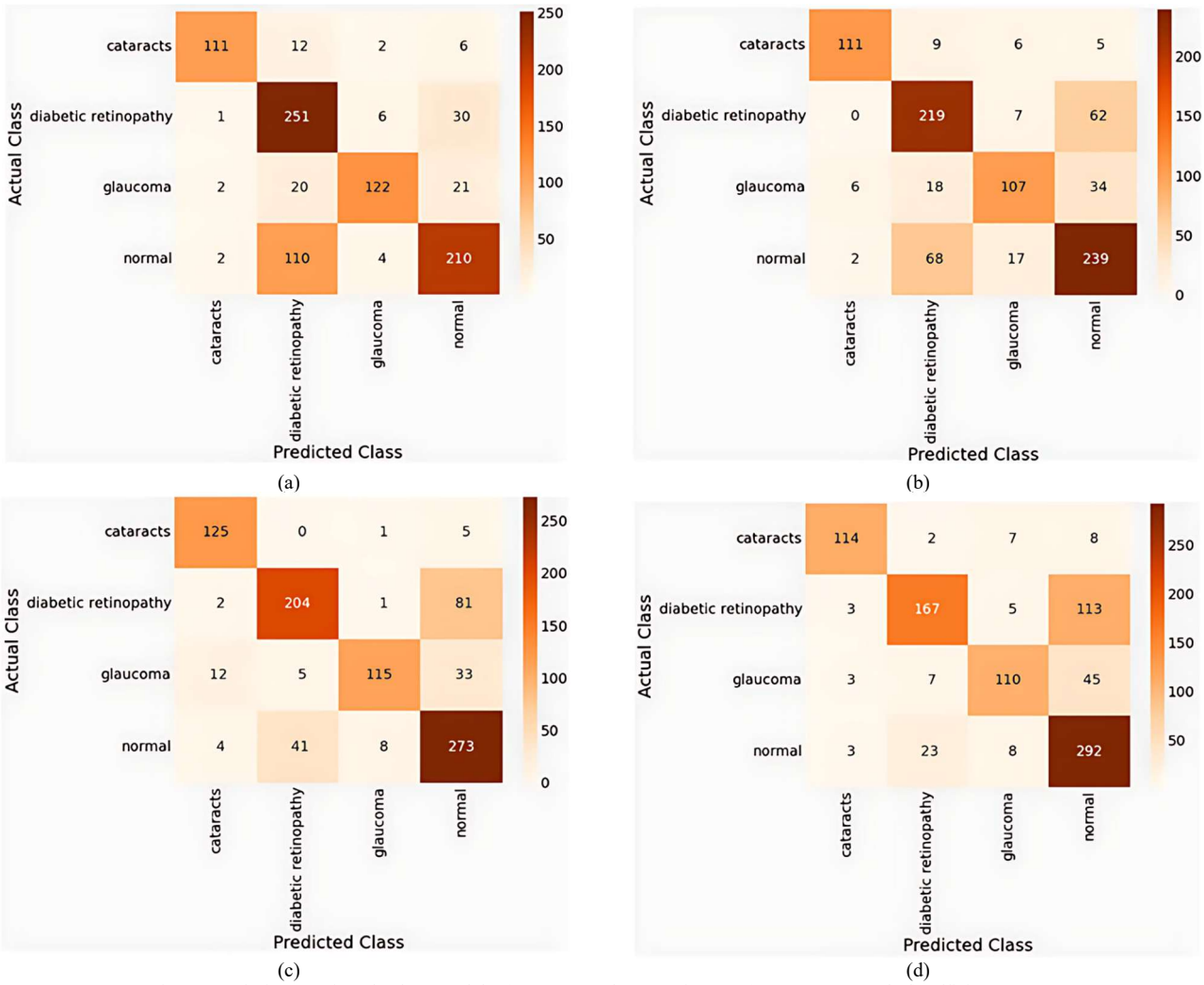


Fig. 4. Confusion matrices for four models: (a) VGG16, (b) Inception-v3, (c) ResNet50, and (d) EfficientNet-B0.

V. Conclusion

This study demonstrated that using transfer learning approach, deep learning models can effectively perform multiclass classification of eye diseases, specifically cataracts, diabetic retinopathy, glaucoma, and normal retinal fundus images. All four pre-trained deep learning models deployed in this study achieved encouraging accuracies above 70%, with ResNet50 performing the best at 78.79%. While there is still room for improvement, these results show the potential of pre-trained models as assistive tools in the early screening of multiple eye diseases.

The use of transfer learning also proved suitable for scenarios with limited computational resources, allowing for efficient model development without the need for extensive training from scratch. A larger and more diverse dataset was obtained by combining several publicly available datasets, contributing to improved generalization of the models, though this also presented challenges in terms of maintaining higher accuracy. Beyond achieving competitive performance, this study highlights the generalization capability of pre-trained deep learning models across heterogeneous retinal fundus image data. By training and evaluating models on a combined dataset with inherent variability and inconsistencies, we closely simulate real-world clinical setting. The models' ability to maintain consistent performance under these conditions reinforces their potential for practical deployments in computer-aided systems. This study offers a practical and scalable solution for automated eye disease screening, particularly in settings with limited access to advanced imaging technologies.

For future work, the performance of the model may be further enhanced through various computer vision techniques, such as improved image pre-processing (e.g., illumination correction, noise reduction), advanced feature enhancement, robust data augmentation strategies to address class imbalance and increase model generalization. Additionally, integrating advanced deep learning techniques such as attention mechanisms, transformer-based architectures, or ensemble learning, could improve feature discrimination and classification accuracy. Future research may also explore domain adaptation techniques to reduce performance drops when deploying models across different populations or imaging devices. Finally, real-world validation through prospective clinical studies will be essential to ensure the reliability and clinical utility of the proposed system.

Conflict of Interest

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

A.M.S. Choo collected and pre-processed the data, conducted the research, performed data analysis and wrote the manuscript; S. Chua contributed to data analysis, reviewed the research methodology, proposed improvement, and co-authored the manuscript; L.T. Lim reviewed the collected data and provided feedback on the documentation; D.N.A. Iskandar, M.H.A. Hijazi, and P.N.E. Nohuddin reviewed the manuscript and contributed to its final version. All authors had approved the final version.

FUNDING

This research was funded by the Ministry of Higher Education, Malaysia under the Fundamental Research Grant Scheme (FRGS) [FRGS/1/2023/ICT02/UNIMAS/02/1].

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Universiti Malaysia Sarawak (UNIMAS), in particular the Faculty of Computer Science and Information Technology (FCSIT) and the Faculty of Medical and Health Sciences (FMHS) for their support and opportunity to conduct this research. We also extend our appreciation to the dataset owners and providers for making their datasets publicly available, which was instrumental in the completion of this research.

REFERENCES

[1] World Health Organization. (2019). World report on vision. [Online]. Available: https://www.who.int/publications/i/item/world-report-on-vision

[2] World Health Organization: WHO. (August 10, 2023). Blindness and vision impairment. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment

[3] A. Raza, M. U. Khan, Z. Saeed, S. Samer, A. Mobeen, and A. Samer, "Classification of eye diseases and detection of cataract using Digital Fundus Imaging (DFI) and inception-v4 deep learning model," in *Proc. 2021 International Conference on Frontiers of Information Technology (FIT)*, 2021, pp. 137–142. https://doi.org/10.1109/fit53504.2021.00034

[4] A. Raj, A. K. Tiwari, and M. G. Martini, "Fundus image quality assessment: Survey, challenges, and future scope," *IET Image Processing*, vol. 13, no. 8, pp. 1211–1224, 2019. https://doi.org/10.1049/iet-ipr.2018.6212

[5] S. Muchuchuti and S. Viriri, "Retinal disease detection using deep learning techniques: A comprehensive review," *Journal of Imaging*, vol. 9, no. 4, 84, 2023. https://doi.org/10.3390/jimaging9040084

[6] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Q. Al-Dujaili, Y. Duan, O. Al-Shamma *et al.*, "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, no. 1, pp. 1–74, 2021. https://doi.org/10.1186/s40537-021-00444-8

[7] Z. Arif, R. Y. N. Fuadah, S. Rizal, and D. Ilhamdi, "Classification of eye diseases in fundus images using Convolutional Neural Network (CNN) method with EfficientNet architecture," *JRTI (Jurnal Riset Tindakan Indonesia)*, vol. 8, no. 1, pp. 125–131, 2023. https://doi.org/10.29210/30032835000

[8] L. Dai, L. Wu, H. Li, C. Cai, Q. Wu, H. Kong *et al.*, "A deep learning system for detecting diabetic retinopathy across the disease spectrum," *Nature Communications*, vol. 12, no. 1, 3242, 2021. https://doi.org/10.1038/s41467-021-23458-5

[9] P. Purwono, A. Ma'arif, W. Rahmaniar, H. I. K. Fathurrahman, A. Z. K. Frisky, and Q. M. U. Haq, "Understanding of convolutional neural network (CNN): A review," *International Journal of Robotics and Control Systems*, vol. 2, no. 4, pp. 739–748, 2023. https://doi.org/10.31763/ijrcs.v2i4.888

[10] G. Singh, K. Guleria, and S. Sharma, "A pre-trained VGG16 model for cataract eye disease prediction," in *Proc. 2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*, 2023. https://doi.org/10.1109/smartgencon60755.2023.10442647

[11] M. Z. Muntaqim, T. A. Smrity, A. S. M. Miah, H. M. Kafi, T. Tamanna, F. A. Farid *et al.*, "Eye disease detection enhancement using a multi-stage deep learning approach," *IEEE Access*, 2024. https://doi.org/10.1109/access.2024.3476412

[12] S. El-Ateif and A. Idri, "Multimodality fusion Strategies in eye disease diagnosis," *Journal of Imaging Informatics in Medicine,* vol. 37, no. 5, pp. 2524–2558, 2024. https://doi.org/10.1007/s10278-024-01105-x

[13] K. D. K. Wardhani, S. Kasim, A. Erianda, and R. Hassan, "Deep learning-based method in multimodal data for diabetic retinopathy detection," *International Journal on Advanced Science Engineering and Information Technology,* vol. 14, no. 5, pp. 1602–1608, 2024. https://doi.org/10.18517/ijaseit.14.5.11677

[14] R. Hemelings, B. Elen, J. B. Breda, M. B. Blaschko, P. De Boever, and I. Stalmans, "Deep learning on fundus images detects glaucoma beyond the optic disc," *Scientific Reports*, vol. 11, no. 1, 20313, 2021. https://doi.org/10.1038/s41598-021-99605-1

[15] A. Shoukat, S. Akbar, S. A. E. Hassan, A. Rehman, and N. Ayesha, "An automated deep learning approach to diagnose glaucoma using retinal fundus images," in *Proc. 2021 International Conference on Frontiers of Information Technology (FIT)*, 2021, pp. 120–125. https://doi.org/10.1109/fit53504.2021.00031

[16] R. Thanki, "A deep neural network and machine learning approach for retinal fundus image classification," *Healthcare Analytics*, vol. 3, 100140, 2023. https://doi.org/10.1016/j.health.2023.100140

[17] F. Kallel and A. Echtioui, "Retinal fundus image classification for diabetic retinopathy using transfer learning technique," *Signal Image and Video Processing,* vol. 18, no. 2, pp. 1143–1153, 2024. https://doi.org/10.1007/s11760-023-02820-8

[18] R. Sarki, K. Ahmed, H. Wang, Y. Zhang, J. Ma, and K. Wang, "Image preprocessing in classification and identification of diabetic eye diseases," *Data Science and Engineering*, vol. 6, no. 4, pp. 455–471, 2021. https://doi.org/10.1007/s41019-021-00167-z

[19] N. Chea and Y. Nam, "Classification of fundus images based on deep learning for detecting eye diseases," *Computers, Materials & Continua*, vol. 67, no. 1, pp. 411–426, 2021. https://doi.org/10.32604/cmc.2021.013390

[20] T. Guergueb and M. A. Akhloufi, "Ocular diseases detection using recent deep learning techniques," in *Proc. 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2021, pp. 3336–3339. https://doi.org/10.1109/embc46164.2021.9629763

[21] T. Babaqi, M. Jaradat, A. E. Yıldırım, S. H. Al-Nimer, and D. Won, "Eye disease classification using deep learning techniques," arXiv Preprint, arXiv: 2307.10501, 2023. https://doi.org/10.48550/arxiv.2307.10501

[22] S. A. Toki, S. Rahman, S. M. B. Fahim, A. A. Mostakim, and M. K. Rhaman, "RetinalNet-500: A newly developed CNN model for eye disease detection," in *Proc. 2022 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, 2022, pp. 459–463. https://doi.org/10.1109/miucc55081.2022.9781785

[23] Y. Cui, R. Sun, B. Liu, Z. Liu, and T. T. Toe, "Eye diseases classification using transfer learning of residual neural network," in *Proc. 2023 3rd International Symposium on Computer Technology and Information Science (ISCTIS)*, 2023, pp. 244–248. https://doi.org/10.1109/isctis58954.2023.10213138

[24] A. Tashkandi, "Eye Care: Predicting eye diseases using deep learning based on retinal images," *Computation,* vol. 13, no. 4, 91, 2025. https://doi.org/10.3390/computation13040091

[25] M. Z. Muntaqim, T. A. Smrity, A. S. M. Miah, H. M. Kafi, T. Tamanna, F. A. Farid *et al.*, "Eye disease detection enhancement using a multi-stage deep learning approach," *IEEE Access*, 2024. https://doi.org/10.1109/access.2024.3476412

[26] K. D. K. Wardhani, S. Kasim, A. Erianda, and R. Hassan, "Deep learning-based method in multimodal data for diabetic retinopathy detection," *International Journal on Advanced Science Engineering and Information Technology,* vol. 14, no. 5, pp. 1602–1608, 2024. https://doi.org/10.18517/ijaseit.14.5.11677

[27] E. Y. Kang, L. Yeung, Y. Lee, C. Wu, S. Peng, Y. Chen *et al.*, "A multimodal imaging–based deep learning model for detecting treatment-requiring retinal vascular diseases: Model development and validation study," *JMIR Medical Informatics,* vol. 9, no. 5, e28868, 2021. https://doi.org/10.2196/28868

[28] G. V. Doddi. (2020). Eye Disease Retinal Images [Dataset; Kaggle]. *Kaggle*. [Online]. Available: https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification/data

[29] A. Jain. (2022). Glaucoma fundus imaging datasets [Dataset]. *Kaggle*. [Online]. Available: https://www.kaggle.com/datasets/arnavjain1/glaucoma-datasets?select=G1020

[30] Larxel. (2020). Ocular disease recognition [Dataset; Kaggle]. *Kaggle*. [Online]. Available: https://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k
[31] S. Pachade, P. Porwal, D. Thulkar, M. Kokare, G. Deshmukh, V. Sahasrabuddhe *et al.*, "Retinal fundus multi-disease image dataset (RFMID) [Dataset]," *IEEE DataPort*, 2021. https://doi.org/10.21227/s3g7-st65
[32] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv Preprint, arXiv:1409.1556, 2014. https://doi.org/10.48550/arXiv.1409.1556
[33] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov *et al.,* "Going deeper with convolutions," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9. https://doi.org/10.1109/CVPR.2015.7298594
[34] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition,* 2016, pp. 2818–2826. doi: 10.1109/CVPR.2016.308
[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778. https://doi.org/10.1109/cvpr.2016.90
[36] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. International Conference on Machine Learning,* 2019, pp. 6105–6114.