ILKOM Jurnal Ilmiah Vol. 18, No. 1, April 2026, pp. 195-206
Accredited 2nd by KEMDIKTISAINTEK No. 10/C/C3/DT.05.00/2025; E-ISSN 2548-7779 | P-ISSN 2087-1716

Research Article

Enhancing Eye Disease Classification Accuracy Using Convolutional Neural Networks with Transfer Learning

Nazlina Izmi Addyna [a,1,*]; **Rahmat Widia Sembiring** [b,2]; **Agus Perdana Windarto** [a,3]
[a] *STIKOM Tunas Bangsa, Jl. Jenderal Sudirman, Pematang Siantar, 2112, Indonesia*
[b] *Politeknik Negeri Medan, Jl. Almamater No.1, Padang Bulan, Kec. Medan Baru, Kota Medan, Sumatera Utara 20155, Indonesia*
[1] *nazlina@amiktunasbangsa.ac.id*; [2] *rahmatws@polmed.ac.id*; [3] *agus.perdana@amiktunasbangsa.ac.id*
* *Corresponding author*

Article history: Received August 16, 2025; Revised August 22, 2025; Accepted April 20, 2026; Available online April 27, 2026

Abstract

Eye diseases serve as a primary contributor to global blindness, making early detection a critical determinant in effective treatment outcomes. While retinal fundus image analysis is the diagnostic standard, conventional manual methods are often hindered by observer subjectivity and time inefficiencies. This study aims to optimize eye disease classification using a Convolutional Neural Network (CNN) approach empowered by transfer learning techniques. Utilizing a dataset of 1,200 retinal fundus images sourced from Kaggle, this research classifies four categories: normal, glaucoma, cataract, and diabetic retinopathy. To mitigate the challenge of limited labeled medical datasets, specific data augmentation strategies—including random flip, zoom, and contrast adjustments—were applied. The study conducts a comparative evaluation of three architectures: standard VGG16, baseline MobileNet, and a proposed optimized MobileNet. The proposed method utilizes Random Search to systematically optimize hyperparameters such as learning rates, dense layer units, and dropout rates. Experimental results demonstrate that the optimized MobileNet achieved superior performance with 89.17% accuracy, significantly outperforming the VGG16 baseline 82,00% and baseline MobileNet 85,00%. Notably, the model achieved perfect recall for diabetic retinopathy, although glaucoma remained the most challenging class due to subtle morphological similarities with normal eyes. These findings confirm that integrating lightweight CNNs with appropriate transfer learning yields a diagnostic system that is not only accurate but also efficient for deployment in resource-constrained environments.

Keywords: CNN; Eye Disease Classification; MobileNet; VGG16; Transfer Learning.

Introduction

Eye diseases represent a significant global health concern, profoundly affecting human quality of life. These visual impairments may impact various anatomical structures of the eye, such as the retina, cornea, lens, and the optic nerve structure, and may arise from a range of causes such as infections, inflammation, congenital abnormalities, degenerative processes, and trauma [1], [2], [3]. According to the World Health Organization (WHO), over 2.2 billion people worldwide suffer from visual impairments, with at least 1 billion of these cases being preventable or treatable if diagnosed at an early stage [4], [5]. In Indonesia, the prevalence of eye diseases has increased, particularly due to an aging population and the high incidence of systemic diseases such as diabetes and hypertension [6], [7]. This trend poses considerable challenges to the national healthcare system, especially in remote areas where access to ophthalmologists and advanced diagnostic tools is limited [8], [9]. Currently, the diagnostic process for eye diseases largely relies on manual examination by medical specialists, which demands significant time and expertise, making the outcomes highly dependent on the examiner's skill. Consequently, there is a pressing need for an automated classification system powered by advanced technologies that can deliver fast, accurate, and consistent diagnoses, particularly in resource-limited settings [10].

Recent developments in Artificial Intelligence (AI), particularly deep learning methods like Convolutional Neural Networks (CNNs), has played a transformative role in medical image classification, including fundus imaging of the retina [11]. The ability of CNNs to detect subtle visual cues and extract relevant features from medical images such as those obtained from retinal imaging and Optical Coherence Tomography—has made them widely adopted in automated diagnostic systems [12]. One increasingly adopted technique is transfer learning, which allows the utilization of pre-trained CNN models on large datasets like ImageNet and their adaptation to specific classification

http://dx.doi.org/10.33096/ilkom.v18i1.2886.195-206

tasks, such as eye disease detection. This method offers benefits in terms of training efficiency and improved accuracy, especially when labeled data is limited [13].

Several CNN architectures like InceptionV3 [14], ResNet [15], and VGG [16] have been successfully implemented for retinal fundus image classification to detect diseases such as glaucoma, cataracts, and diabetic retinopathy. However, standard CNNs without further optimization often face limitations in generalization, data requirements, and computational efficiency. For instance, Putri & Rakasiwi (2025) reported that a baseline CNN model achieved only 75% accuracy, which improved significantly to 88% after applying transfer learning on the VGG-16 architecture. This indicates that conventional CNN models may not be optimal without further adjustments. Similarly, An et al. (2021) found that direct classification using CNNs was ineffective with limited data, prompting them to use a two-stage hierarchical classification strategy. Moreover, Sarki et al. (2022) highlighted that while CNNs could classify diabetic eye diseases with 81.33% accuracy, they struggled to distinguish early-stage conditions due to the complexity of inter-class visual features. These findings underscore the necessity of architectural refinements and optimization strategies such as hyperparameter tuning for CNNs to achieve robust performance in automatic eye disease diagnosis. In response to these challenges, this study proposes the optimization of the MobileNet architecture using hyperparameter tuning with a Random Search algorithm for retinal fundus image classification [17]. Parameter such as the number of neurons in dense layers, dropout rates, and learning rates were systematically adjusted to identify the best-performing model [18]. Furthermore, fine-tuning of the pre-trained MobileNet weights was performed to improve adaptation to the specific characteristics of the retinal dataset. Model performance was evaluated through training visualization and comparative analysis against two benchmark models: unoptimized MobileNet and VGG16 [19], [20].

Beyond algorithmic performance, the transition of deep learning models into actual clinical implementation presents multidimensional challenges [21], [22]. Primary hurdles include the complexity of integrating software with diverse medical hardware infrastructures and the urgency of validating models against patient demographic heterogeneity to prevent diagnostic bias [23]. However, these challenges coexist with significant opportunities in digital health development. The utilization of efficient and lightweight CNN architectures paves the way for mobile based screening systems [24]. This approach enables periodic eye health monitoring via smartphones, potentially democratizing access to ophthalmology services, particularly for populations in remote regions with limited access to conventional medical facilities [25].

The objective of this study is to design a CNN-based classification system that incorporates transfer learning and model optimization to enhance diagnostic accuracy. By comparing the performance of standard VGG16, baseline MobileNet, and the optimized MobileNet, this research seeks to determine the effectiveness of hyperparameter tuning in improving model performance and its feasibility for implementation in real-world medical settings, especially those with limited resources and access to specialists.

Method

A. Related Work

Several studies have been conducted to develop automated classification systems based on Convolutional Neural Networks (CNNs) for detecting eye diseases using retinal fundus images [21]. While the results obtained so far are promising, there remain notable challenges related to accuracy, the need for large datasets, and computational efficiency that must be addressed. Therefore, **Table 1** presents a review of relevant previous studies as a foundation for comparison and to strengthen the contribution and direction of this research.

Table 1. Related Work on Eye Disease Classification Using CNN and Transfer Learning

Study	Research Focus	Method / Model	Limitations Identified
Putri and Rakasiwi. (2025)	Retinal Fundus Image Classification	Standard VGG-16	Lack of architectural modifications and limited number of classified disease categories.
An et al. (2021)	Glaucoma Eye Disease Classification	CNN + Hierarchy Transfer Learning	Approach is restricted to a single disease type and limited dataset size.
Sarki et al (2022)	Multi-Class Eye Disease Classification Caused by Diabetes (DED)	Standard CNN + Proposed Model	Absence of transfer learning limits model optimization and performance.

Study	Research Focus	Method / Model	Limitations Identified
This Study (Proposed)	Improving CNN Model Accuracy Through Transfer Learning for Eye Disease Classification	Standard CNN (VGG-16, MobileNet) and Optimized MobileNet	Optimization performed using Random Search with limited trials and fine-tuning enabled.

Table 1 presents a summary of recent studies that apply convolutional neural networks (CNNs) to eye disease image classification. Putri and Rakasiwi (2025) implemented the VGG-16 architecture by leveraging transfer learning techniques to perform classification of ocular diseases using retinal fundus imagery. Although the study reported improved accuracy, it was constrained by a small dataset size, potential data bias, and the model's difficulty in distinguishing between visually similar classes, such as glaucoma and normal [26]. An et al. (2021) proposed a two-stage approach based on hierarchical transfer learning that mimics the clinical diagnostic process.

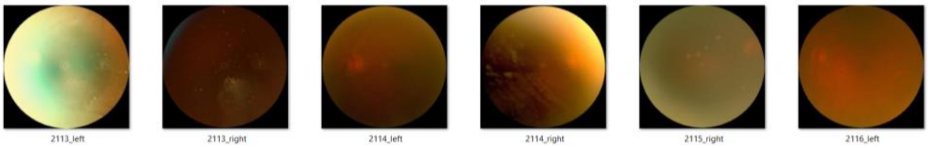
However, their model relied heavily on physician-structured diagnosis, was limited to a single type of data, and introduced high system complexity due to the use of ensemble learning making its application in real-world clinical settings less practical [27]. Meanwhile, Sarki et al. (2022) developed a convolutional neural network approach designed for the classification, of multiple categories of eye diseases, yet the model struggled to detect subtle visual features, lacked generalization due to limited data, and required high computational power. Additionally, the absence of transfer learning strategies further limited the model's diagnostic effectiveness [28].

In contrast, the present study proposes an optimized MobileNet approach enhanced with transfer learning, hyperparameter tuning, and fine-tuning using the Random Search algorithm specifically tailored to the characteristics of retinal fundus images. Unlike previous studies, the developed model was not only tested against untuned baseline architectures but also comparatively validated against VGG-16 as a benchmark. This innovation aims to address the common limitations identified in earlier literature, particularly in terms of training efficiency, accuracy improvement, performance benchmarking, and adaptability for deployment in resource-constrained healthcare environments.

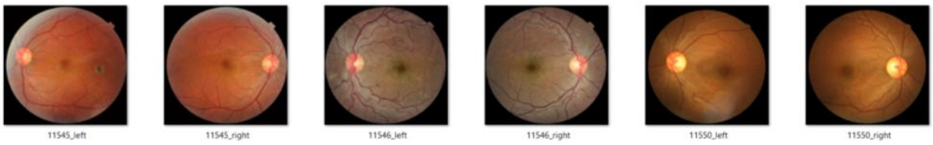
B. Dataset

The dataset employed in this study consists of labeled retinal fundus images sourced from the open-access platform Kaggle, specifically the Eye Diseases Classification Dataset. This dataset comprises a collection of retinal images that have been manually categorized by domain experts into four primary classes of ocular diseases: cataract, glaucoma, diabetic retinopathy, and normal eye condition. Each class contains an equal number of samples 300 images resulting in a total of 1,200 images used in this study. All images have a fixed resolution of 512 × 512 pixels and are stored in high-quality JPG format, which facilitates the precise extraction of retinal morphological features. The dataset includes both left and right eye images, enriching the spatial variability during model training. This bilateral representation broadens the model's understanding of the asymmetric characteristics of retinal structures. Additionally, variations in lighting conditions, image orientation, and contrast levels enhance the overall data distribution and improve the model's robustness against noise and real-world clinical conditions that are often unstandardized.

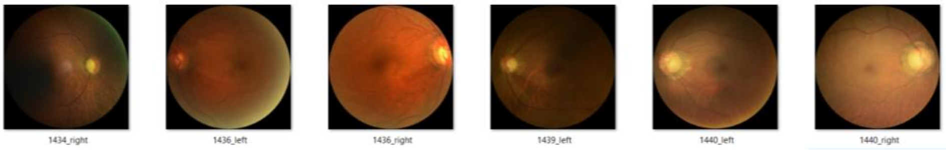
Beyond spatial and illumination diversity, all images have been manually verified and labeled by certified ophthalmologists, thus enhancing the credibility and labeling accuracy of the dataset as a reference for developing CNN-based classification systems. The high intra-class variation resulting from diverse clinical imaging settings significantly contributes to the model's discriminative ability in recognizing subtle features commonly found among individuals within the same disease category. Given these characteristics, the dataset is considered highly representative and suitable as a foundational benchmark for the development, training, and validation of deep learning-based medical image classification models aimed at the automatic and accurate detection of eye diseases.



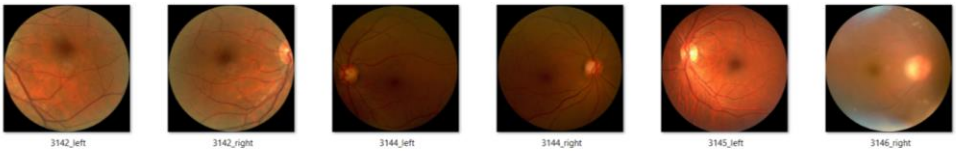
(a). Cataract Eye Disease Images



(b). Diabetic Retinopathy Eye Disease Images



(c). Glaucoma Eye Disease Images



(d). Normal Eye Disease Images

Figure 1. The Example of retinal fundus images representing four different eye disease classes: normal, cataract, glaucoma, and diabetic retinopathy.

C. The Research Framework

The research framework is designed to represent the systematic relationship between problem formulation, modeling approach, implementation strategy, and evaluation process. This framework provides a comprehensive overview of the research stages, starting from problem identification and objective formulation, followed by key methodological steps such as medical image data collection and preprocessing, application of CNN architecture and transfer learning, and culminating in model testing using measurable evaluation parameters.

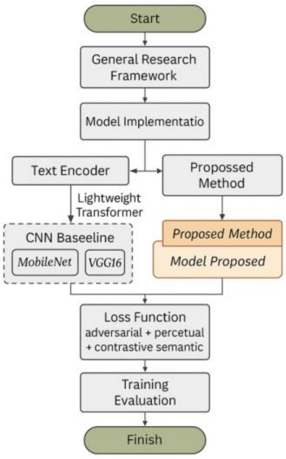


Figure 2. The Research Framework

The research framework illustrated in Figure 2 begins with the initial stage outlining the overall research design, followed by the collection of retinal fundus image data and a data preprocessing phase. This preprocessing includes resizing, normalization, image augmentation, and labeling aimed at improving input quality and preventing overfitting during model training. The core of this study lies in the implementation of the classification model through two main approaches. First, the baseline CNN architectures are applied using two widely adopted models in medical image classification: MobileNet and VGG16. Second, a proposed method is developed by modifying the baseline architecture with specific adjustments to enhance accuracy and efficiency in multiclass eye disease classification. Both the baseline and the proposed models are then evaluated in a comparative manner to assess their relative performance. The evaluation process employs quantitative metrics such as accuracy, precision, recall, and F1-score. In addition, a confusion matrix analysis is conducted to identify misclassifications between classes. This research framework therefore emphasizes not only model development but also a systematic and objective comparison between different classification approaches.

D. Proposed Model

In the proposed model of this study, the MobileNet architecture was selected as the base model due to its efficiency and capability to effectively extract visual features from high-resolution images. To improve classification performance particularly the accuracy in classifying retinal fundus images a series of optimizations were performed on the model structure and training parameters using a hyperparameter tuning approach based on Random Search, implemented via the Keras Tuner library in **Table 2**.

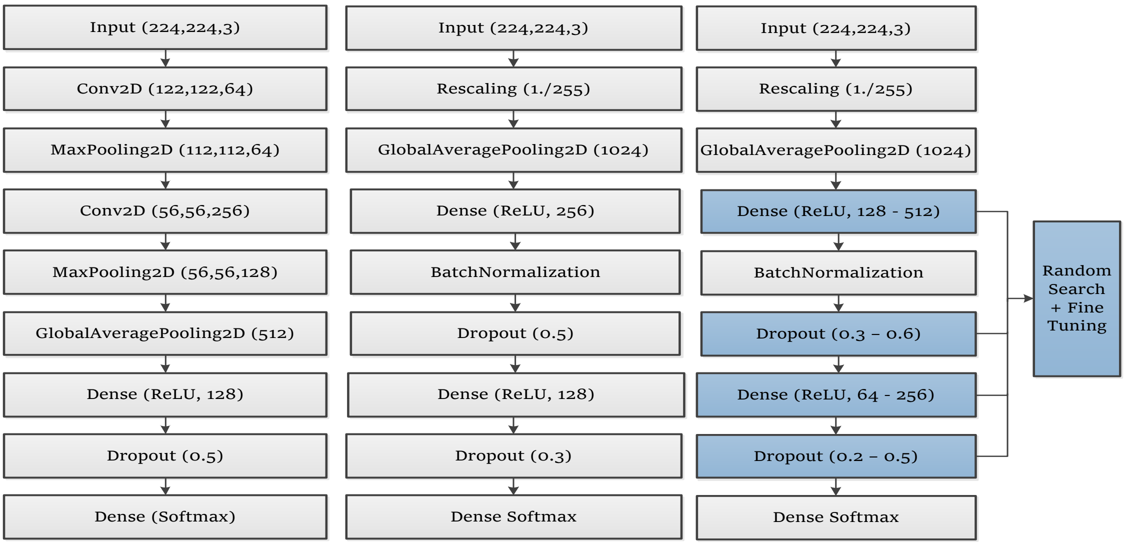
Table 2. Proposed Architecture For The Mobilenet-Based Model

Layer	Output Size	Details
Input	224×224×3	RGB image resized to 224×224 pixels
Squential Layer	224×224×3	Includes RandomFlip, RandomZoom, Contrast, and Brightness adjustments
Rescaling(1./255)	224×224×3	Normalizes pixel values from 0–255 to 0–1
GlobalAveragePooling2D()	1024	Flattens feature maps before Dense layers
Dense(units1, activation='relu')	128-512	ReLU activation; number of units optimized
BatchNormalization()	224×224×3	Stabilizes and accelerates training
Dropout(dropout1)	0.3–0.6	Randomly drops neurons for regularization
Dense(units2, activation='relu')	64-256	ReLU activation; number of units optimized
Dropout(dropout2)	0.2-0.5	Additional dropout for regularization
Dense(4, activation='softmax')	jumlah_kelas=4	Classifies into four eye disease categories
Optimizer: Adam (learning rate tuned) + Random Search	1e-4 / 5e-5 / 1e-5	Learning rate optimized via Random Search

The proposed model is constructed by sequentially stacking several layers, beginning with a data augmentation phase that applies random flip, zoom, contrast adjustment, and brightness enhancement. This step enriches the diversity of the training data to reduce the risk of overfitting. Following normalization through the Rescaling(1./255) layer, MobileNet is used as a feature extractor with include_top=False and pretrained weights from ImageNet. Unlike the baseline models, which freeze the MobileNet weights, the proposed model enables fine-tuning by setting trainable=True, allowing the feature representation to be more specific to the characteristics of the dataset used. The model is then enhanced with two Dense layers, whose neuron counts are determined through tuning—ranging from 128 to 512 in the first Dense layer and 64 to 256 in the second. Dropout layers with probabilities between 0.3–0.6 and 0.2–0.5 are inserted to prevent overfitting. Additionally, the learning rate for the Adam optimizer is automatically selected from 1e-4, 5e-5, or 1e-5 during the tuning process, ensuring optimal and stable training speed.

To optimize model performance without incurring excessive computational costs, this study employed a Random Search based hyperparameter tuning technique [29]. Unlike Grid Search, which tests every parameter combination systematically but rigidly, Random Search explores the search space by sampling random combinations of parameters over a fixed number of iterations (50 iterations). Random Search has proven to be more effective because, in many instances, only a small subset of hyperparameters (such as the learning rate) significantly impacts model accuracy [30], [31]. By scattering search points stochastically, the model has a higher probability of converging on optimal values for these critical parameters in significantly less time compared to an exhaustive search.

This approach allows the model to discover the best combination of network structure and training parameters based on validation performance. The application of fine-tuning to MobileNet, optimal Dense layer configurations, and adaptive learning rate tuning significantly improved model accuracy compared to baseline models. These results demonstrate that a well-targeted and comprehensive hyperparameter tuning strategy can enhance the generalization capability of CNN-based models for retinal fundus image classification. A performance comparison between the proposed MobileNet model (with hyperparameter tuning) and standard models (MobileNet and VGG-16) is illustrated in **Figure 3** below.



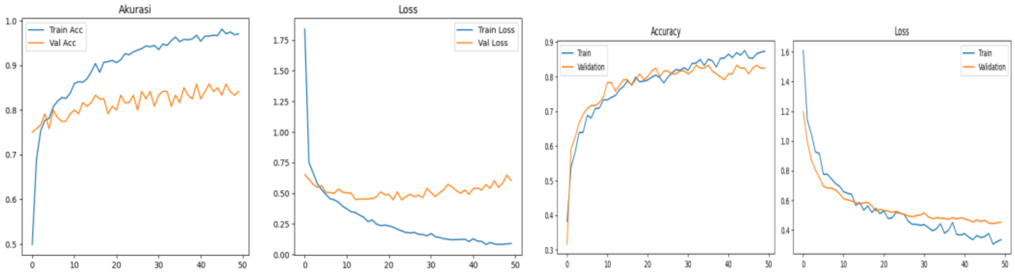
(a). VGG-16 Architecture (b). MobileNet Architecture (c). MobileNet Proposed Architecture

Figure 3. Comparison of Model Architectures

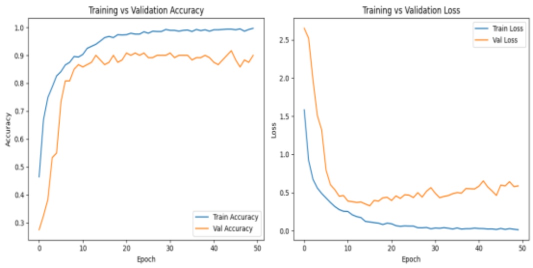
Results and Discussion

A. Training Performance Analysis

Training performance analysis was conducted to evaluate how well the model learned patterns from retinal fundus image data throughout the training process. This evaluation includes monitoring the progression of both loss and accuracy values on training and validation datasets at each epoch. The observation of these two metrics aims to assess the stability of the learning process, the model's generalization capability on unseen data, and to detect signs of overfitting or underfitting.This analysis provides a critical foundation for determining the effectiveness of the architectural configuration and the optimization strategies applied to the model. To offer a clearer visualization of each model's performance during training, the loss and accuracy curves are presented in **Figure 4**.



(a). VGG-16 (b). MobileNet



(c). MobileNet Hyperparameter Tuning (Model Proposed)

Figure 4. Comparison of training results for all models (accuracy and loss).

In this study, **Figure 4(a)** illustrates the training performance of the VGG16 model. The training accuracy curve shows a steady increase, approaching nearly 100%. However, the validation accuracy curve exhibits noticeable fluctuations and stagnates below 85%, accompanied by a significant gap between training and validation loss values indicating a potential overfitting issue. This suggests that while VGG16 effectively fits the training data, its ability to generalize to unseen data remains limited.

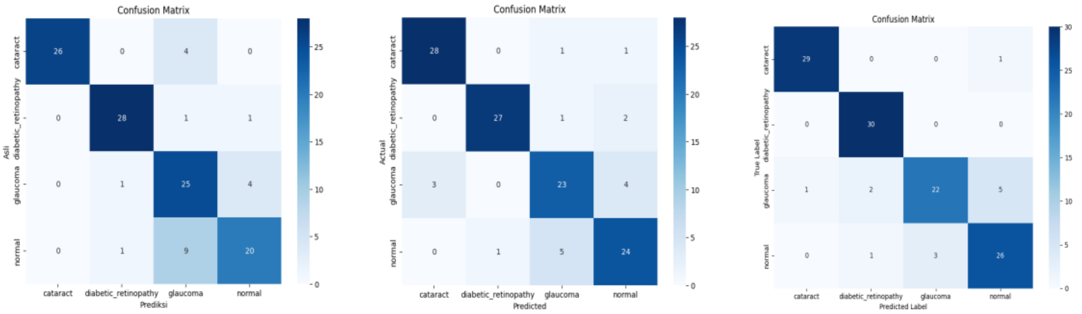
Figure 4(b) presents the results of the baseline MobileNet model, which demonstrates more consistent and gradually improving training and validation accuracy curves over time. The loss curves for both training and validation converge toward similar points, indicating that the model generalizes better and achieves higher validation accuracy than VGG16. This pattern implies that MobileNet, with its lower architectural complexity, is capable of producing a more stable model without experiencing significant overfitting. In contrast, **Figure 4(c)** shows the training outcome of the MobileNet model optimized through hyperparameter tuning using Random Search. This model displays excellent training stability, with training and validation accuracy curves closely aligned and validation accuracy consistently increasing. The loss values continue to decrease until they reach a low and stable point. Notably, the validation accuracy of the optimized model outperforms both baseline models, indicating that the tuning process successfully identified an optimal parameter configuration.

Overall, the visual analysis of the training curves confirms that the proposed model MobileNet with hyperparameter tuning via Random Search not only achieves higher accuracy but also maintains stability and training efficiency. These advantages make it a strong candidate for deployment in automated eye disease classification systems, particularly in medical environments with limited resources. These findings suggest that strategic hyperparameter optimization can significantly enhance CNN model performance in medical image classification tasks. Nevertheless, it should be noted that although the optimized model demonstrated superior visual and quantitative performance compared to the baselines, the study still presents opportunities for further improvement. One potential direction is to expand the diversity of training data, which could enhance the model's generalization capability to more varied and real-world data distributions.

B. Evaluation Metrics and Comparative Analysis

To comprehensively assess the performance of the models, it is essential to employ relevant and robust evaluation metrics—particularly within the context of medical image classification. The evaluation process extends beyond measuring accuracy alone and includes additional metrics such as precision, recall, and F1-score. These metrics provide a more detailed understanding of how well each model distinguishes between different eye disease classes. Furthermore, the analysis is complemented with detailed visual outputs, including the confusion matrix (**Figure 5**) and classification report (**Figure 6**), to illustrate the model's performance in identifying and separating each class correctly.

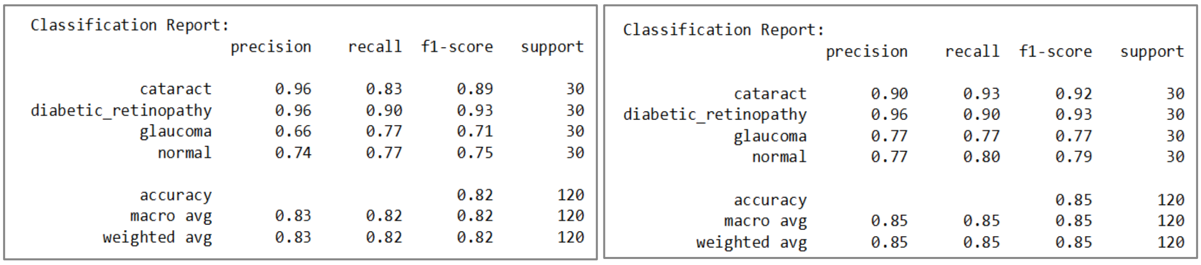
This analysis is crucial for understanding the generalization capability of the models and for identifying potential misclassifications that could impact diagnostic reliability. Accordingly, this section presents a comparative performance evaluation of the three models VGG-16 baseline, MobileNet baseline, and the optimized MobileNet based on the aforementioned metrics. The aim is to assess the effectiveness of the hyperparameter tuning approach via Random Search in improving the overall accuracy, reliability, and practicality of automated eye disease classification systems.



(a). VGG-16 (b). MobileNet (c). MobileNet (Model Proposed)

Figure 5. Confusion Matrix

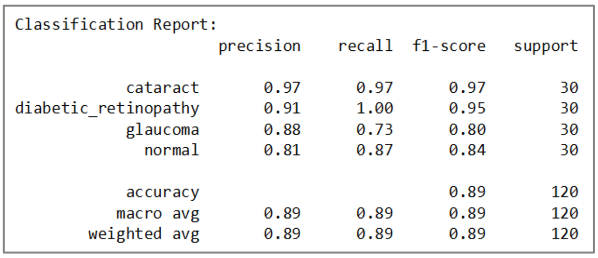
Figure 5(a) presents the confusion matrix of the VGG16 baseline model, which demonstrates a reasonably good performance in classifying cataract and diabetic retinopathy classes. However, it still exhibits significant misclassification in the glaucoma and normal classes. **Figure 5(b)** shows the results of the baseline MobileNet model, indicating more stable performance with improved accuracy in the normal class and reduced prediction errors in the glaucoma class. In contrast, **Figure 5(c)** illustrates the performance of the optimized MobileNet model, which achieves near-perfect classification across all classes with noticeably fewer errors than the two baseline models. These results suggest that hyperparameter tuning using Random Search substantially improves the model's accuracy and generalization capability in eye disease classification. To provide a more comprehensive comparison of model performance, evaluation was also conducted using four key classification metrics: accuracy, precision, recall, and F1-score. The three models VGG16 baseline, MobileNet baseline, and the optimized MobileNet were compared based on their average scores along with 95% confidence intervals, providing insight into the consistency of predictions across the validation dataset. This approach aims to assess the degree of stability and reliability of each model in accurately classifying retinal fundus images. The evaluation results were measured based on precision, recall, and F1-score metrics, as presented in the following report.



```
Classification Report:
                      precision    recall  f1-score   support

            cataract       0.96      0.83      0.89        30
diabetic_retinopathy       0.96      0.90      0.93        30
            glaucoma       0.66      0.77      0.71        30
              normal       0.74      0.77      0.75        30

            accuracy                           0.82       120
           macro avg       0.83      0.82      0.82       120
        weighted avg       0.83      0.82      0.82       120
```

```
Classification Report:
                      precision    recall  f1-score   support

            cataract       0.90      0.93      0.92        30
diabetic_retinopathy       0.96      0.90      0.93        30
            glaucoma       0.77      0.77      0.77        30
              normal       0.77      0.80      0.79        30

            accuracy                           0.85       120
           macro avg       0.85      0.85      0.85       120
        weighted avg       0.85      0.85      0.85       120
```

(a). VGG-16 (b). MobileNet



```
Classification Report:
                      precision    recall  f1-score   support

            cataract       0.97      0.97      0.97        30
diabetic_retinopathy       0.91      1.00      0.95        30
            glaucoma       0.88      0.73      0.80        30
              normal       0.81      0.87      0.84        30

            accuracy                           0.89       120
           macro avg       0.89      0.89      0.89       120
        weighted avg       0.89      0.89      0.89       120
```

(c). MobileNet (Model Proposed)

Figure 6. Classification Report

Based on the classification report, the MobileNet Proposed architecture demonstrated the best overall classification performance compared to the other two models, namely VGG-16 and baseline MobileNet. Overall, MobileNet Proposed achieved an accuracy of 89%, outperforming MobileNet (85%) and VGG-16 (82%), reflecting superior generalization capability on the test data. In terms of precision, recall, and F1-score, MobileNet Proposed showed more balanced results across all eye disease classes, including *cataract*, *diabetic retinopathy*, *glaucoma*, and *normal*. Notably, the model achieved perfect recall (1.00) and an F1-score of 0.95 for the *diabetic retinopathy* class, indicating that all instances of this disease were correctly identified without misclassification.

Furthermore, significant performance improvements were observed in the *glaucoma* class, which had previously been a challenge for VGG-16 (F1-score: 0.71) and baseline MobileNet (F1-score: 0.77). MobileNet Proposed increased the F1-score for this class to 0.80, although it remained the lowest among the four classes. A deeper analysis of the misclassification patterns reveals that the model's primary challenge lies in the accurate identification of the glaucoma category. The majority of errors in this class manifested as false positives, where glaucoma images were predicted as normal eyes. This phenomenon can be attributed to the significant morphological similarities between normal eyes and early-stage glaucoma. Unlike cataract, which presents distinct visual features (lens opacity), or diabetic retinopathy characterized by hemorrhages, the primary marker for glaucoma relies on the Cup-to-Disc Ratio (CDR) of the optic nerve head. This feature often possesses subtle visual variances that are challenging to extract perfectly through global feature extraction performed by standard CNNs without prior optic disc segmentation [32]. This indicates that while MobileNet is efficient, it may require specific preprocessing techniques, such as Region of

Interest (ROI) cropping on the optic disc area, to enhance sensitivity toward glaucoma cases. In addition, the model achieved both macro average and weighted average F1-scores of 0.89, indicating consistent performance across all classes regardless of data distribution. These findings confirm that the architectural enhancements made to MobileNet not only improved overall accuracy but also significantly enhanced class-wise balance. Overall, the results suggest that MobileNet Proposed offers a more effective solution for detecting various types of eye diseases from fundus images and holds strong potential for implementation in accurate and efficient medical diagnostic support systems.

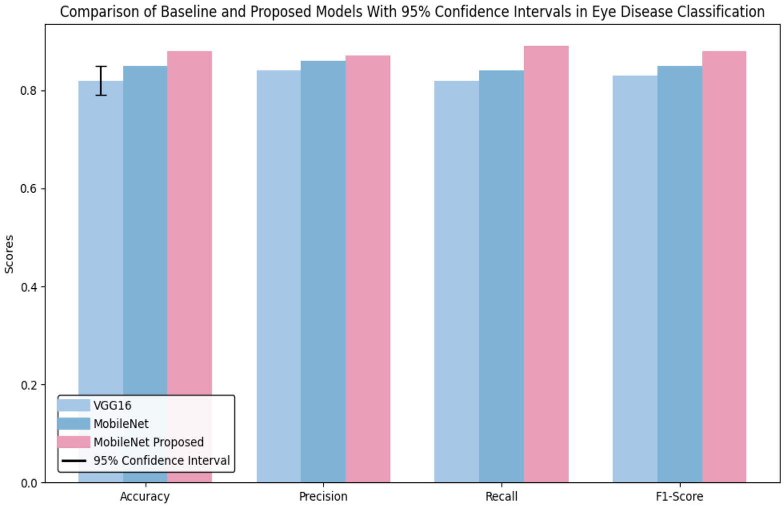


Figure 7. Confidence Intervals

The visualization results indicate that the optimized MobileNet model consistently achieved the highest scores across all evaluation metrics when compared to the two baseline models. Although the performance differences between MobileNet and VGG16 appear modest in certain metrics, the inclusion of confidence intervals in the accuracy graph underscores the proposed model's superior stability and reliability. These findings reinforce the conclusion that hyperparameter tuning via Random Search not only enhances accuracy but also reduces the variability of the model's outcomes, making it more suitable for real-world implementation in automated eye disease classification systems within clinical settings.

C. Discussion of Testing Outcomes

The model testing results provide valuable insights through a comparative summary of evaluation metrics across all models, as presented in **Table 3**.

Table 3. Performance Comparison of Models Based on Accuracy, Precision, Recall, and F1-Score

Models	Accuracy	Precision	Recall	F1-Score
VGG-16	0,82	0,84	0,82	0,83
MobileNet	0,85	0,85	0,85	0,85
Model Proposed	0,89	0,89	0,89	0,89

Based on the results presented in **Table 3**, the optimized MobileNet model with hyperparameter tuning using Random Search consistently demonstrated the highest performance across all evaluation metrics, achieving 89% in accuracy, precision, recall, and F1-score. This performance outperformed both the baseline MobileNet and VGG-16 models, which scored within the 82% to 85% range. These findings confirm that the hyperparameter tuning process using the Random Search algorithm effectively enhanced the model's generalization ability while improving classification stability across classes. Moreover, the increase in the F1-score highlights the proposed model's ability to maintain a balanced trade-off between precision and recall—an essential aspect in medical diagnostics to minimize both false positives and false negatives in eye disease classification.

Conclusion

This study successfully demonstrates the effectiveness of integrating transfer learning techniques with Random Search-based hyperparameter optimization in CNN architectures. The primary novelty of this work lies in providing empirical evidence that lightweight models like MobileNet, when rigorously optimized, can rival the performance of deep architectures such as VGG16 while maintaining significantly superior computational efficiency. The proposed MobileNet achieved a testing accuracy of 89.17% and an average F1-score of 89.00%, positioning it as an ideal candidate for computer ided diagnostic systems on resource-constrained devices. However, this study acknowledges certain limitations. Error analysis indicates that the model continues to face challenges in distinguishing early-stage glaucoma from normal eyes. This limitation is attributed to the reliance on global feature extraction, which lacks sensitivity to subtle morphological variations in the optic disc. Consequently, future research is recommended to integrate preprocessing techniques based on Region of Interest (ROI) segmentation to isolate critical feature areas, and to extend dataset testing within mobile-based application deployment scenarios for more comprehensive clinical validation.

References

[1] T. Bonsaksen, A. Brunes, and T. Heir, “Quality of life in people with visual impairment compared with the general population,” *J. Public Heal.*, vol. 33, no. 1, pp. 23–31, 2025, doi: 10.1007/s10389-023-01995-1.

[2] X. Han, W. Cheng, and L. Liang, “· Editorial Global eye health : challenges , innovations , and future prospects,” pp. 97–98, 2025, doi: 10.12419/es2025010101.

[3] A. Almazroa *et al.*, “The Patients’ Perspective for the Impact of Late Detection of Ocular Diseases on Quality of Life: A Cross-Sectional Study,” *Clin. Optom.*, vol. 15, pp. 191–204, 2023, doi: 10.2147/OPTO.S422451.

[4] S. F. Ahmed *et al.*, *Deep learning modelling techniques: current progress, applications, advantages, and challenges*, vol. 56, no. 11. Springer Netherlands, 2023. doi: 10.1007/s10462-023-10466-8.

[5] O. Goktas, “Factors associated with eye disorders and diseases : A retrospective study,” *Pakistan J. Med. Sci.*, vol. 41, no. 1, pp. 176–181, 2025.

[6] T. Y. Wong, S. C. Loon, and S. M. Saw, “The epidemiology of age related eye diseases in Asia,” *Br. J. Ophthalmol.*, vol. 90, no. 4, pp. 506–511, 2006, doi: 10.1136/bjo.2005.083733.

[7] R. Kesuma and A. Prasetyo, *Literature Review on the Prevalence of Vision Impairment and Age-Related Eye Diseases*, no. Bishss 2023. Atlantis Press SARL, 2024. doi: 10.2991/978-2-38476-273-6_111.

[8] I. Kulkov *et al.*, “Technology entrepreneurship in healthcare: Challenges and opportunities for value creation,” *J. Innov. Knowl.*, vol. 8, no. 2, p. 100365, 2023, doi: 10.1016/j.jik.2023.100365.

[9] G. Więckiewicz, S. Weber, I. Florczyk, and P. Gorczyca, “Socioeconomic Burden of Psychiatric Cancer Patients: A Narrative Review,” *Cancers (Basel).*, vol. 16, no. 6, 2024, doi: 10.3390/cancers16061108.

[10] J. Sheppard, B. S. Lee, L. M. Periman, J. Sheppard, B. S. Lee, and L. M. Periman, “Dry eye disease : identification and therapeutic strategies for primary care clinicians and clinical specialists,” *Ann. Med.*, vol. 55, no. 1, pp. 241–252, 2023, doi: 10.1080/07853890.2022.2157477.

[11] S. Iqbal, A. N. Qureshi, J. Li, and T. Mahmood, “On the analyses of medical images using traditional machine learning techniques and convolutional neural networks,” 2023, *Springer*. doi: 10.1007/s11831-023-09899-9.

[12] E. Hassan, S. Elmougy, M. R. Ibraheem, M. S. Hossain, and ..., “Enhanced deep learning model for classification of retinal optical coherence tomography images,” 2023, *mdpi.com*.

[13] U. P. S. Parmar *et al.*, “Artificial Intelligence (AI) for Early Diagnosis of Retinal Diseases,” *Med.*, vol. 60, no. 4, pp. 1–15, 2024, doi: 10.3390/medicina60040527.

[14] T. Ahad, Y. Li, B. Song, and T. Bhuiyan, “Artificial Intelligence in Agriculture Comparison of CNN-based deep learning architectures for rice diseases classification,” *Artif. Intell. Agric.*, vol. 9, pp. 22–35, 2023, doi: 10.1016/j.aiia.2023.07.001.

[15] L. Alzubaidi *et al.*, *Review of deep learning: concepts, CNN architectures, challenges, applications, future*

directions, vol. 8, no. 1. Springer International Publishing, 2021. doi: 10.1186/s40537-021-00444-8.

[16] G. S. Nugraha, M. I. Darmawan, and ..., "Comparison of CNN's architecture GoogleNet, AlexNet, VGG-16, Lenet-5, Resnet-50 in Arabic handwriting pattern recognition," ... *Electron. Control*, 2023.

[17] S. Jamshidiha, A. Rezaee, F. Hajati, M. Golzan, and R. Chiong, "An Explainable Transformer Model for Alzheimer's Disease Detection Using Retinal Imaging," no. Ml, pp. 1–20, 2025.

[18] M. A. K. Raiaan *et al.*, "A systematic review of hyperparameter optimization techniques in Convolutional Neural Networks," *Decis. Anal. J.*, vol. 11, no. September 2023, p. 100470, 2024, doi: 10.1016/j.dajour.2024.100470.

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

[20] J. Debnath *et al.*, "Informatics in Medicine Unlocked LMVT : A hybrid vision transformer with attention mechanisms for efficient and explainable lung cancer diagnosis," *Informatics Med. Unlocked*, vol. 57, no. June, p. 101669, 2025, doi: 10.1016/j.imu.2025.101669.

[21] D. B. Olawade, S. Chinaza, S. Marinze, E. Egbon, A. Osunmakinde, and A. Osborne, "International Journal of Medical Informatics Artificial intelligence in clinical trials : A comprehensive review of opportunities , challenges , and future directions," *Int. J. Med. Inform.*, vol. 206, no. October 2025, p. 106141, 2026, doi: 10.1016/j.ijmedinf.2025.106141.

[22] A. Choi *et al.*, "OPEN A novel deep learning algorithm for real-time prediction of clinical deterioration in the emergency department for a multimodal clinical decision support system," pp. 1–15, 2024.

[23] S. Morelli and D. Giansanti, "Recent Advances in AI-Driven Mobile Health Enhancing Healthcare — Narrative Insights into Latest Progress," pp. 1–80, 2026.

[24] M. P. Singh *et al.*, "OPEN ACCESS A Healthcare System Employing Lightweight CNN for Disease Prediction with Artificial Intelligence," pp. 1–13, 2024, doi: 10.2174/0118749445302023240520111802.

[25] A. A. Abdulamiery, L. M. Shaker, and A. Alamiery, "Synergizing technology and optics : Exploring the wonders of bioelectronics contact lenses," *Next Mater.*, vol. 10, no. December 2024, p. 101399, 2026, doi: 10.1016/j.nxmate.2025.101399.

[26] C. A. Putri and S. Rakasiwi, "Edumatic : Jurnal Pendidikan Informatika Diagnosis Dini Penyakit Mata : Klasifikasi Citra Fundus Retina dengan Convolutional Neural Network VGG-16," *J. Pendidik. Inform.*, vol. 9, no. 1, pp. 208–216, 2025, doi: 10.29408/edumatic.v9i1.29571.

[27] G. An, M. Akiba, K. Omodaka, T. Nakazawa, and H. Yokota, "Hierarchical deep learning models using transfer learning for disease detection and classification based on small number of medical images," *Sci. Rep.*, vol. 11, no. 1, pp. 1–9, 2021, doi: 10.1038/s41598-021-83503-7.

[28] R. Sarki, K. Ahmed, H. Wang, Y. Zhang, and K. Wang, "Convolutional Neural Network for Multi-class Classification of Diabetic Eye Disease," *EAI Endorsed Trans. Scalable Inf. Syst.*, vol. 9, no. 4, 2022, doi: 10.4108/eai.16-12-2021.172436.

[29] T. Kee and W. K. O. Ho, "Optimizing Machine Learning Models for Urban Sciences : A Comparative Analysis of Hyperparameter Tuning Methods," 2025.

[30] M. Sholeh, U. Lestari, and D. Andayati, "Hyperparameter Optimization Using Grid Search and Random Search to Improve the Performance of Prediction Models with Decision Trees," vol. 3, no. 03, pp. 453–464, 2025.

[31] S. K. Parinduri, P. Alkhairi, and H. Qurniawan, "Classification Model Optimization using Grid Search and Random Search in Machine Learning Algorithms," vol. 4, no. 2, pp. 71–78, 2025, doi: 10.61944/bids.v4i2.136.

[32] R. Pannu *et al.*, "Computerized Medical Imaging and Graphics Enhanced glaucoma classification through

advanced segmentation by integrating cup-to-disc ratio and neuro-retinal rim features," *Comput. Med. Imaging Graph.*, vol. 123, no. January, p. 102559, 2025, doi: 10.1016/j.compmedimag.2025.102559.